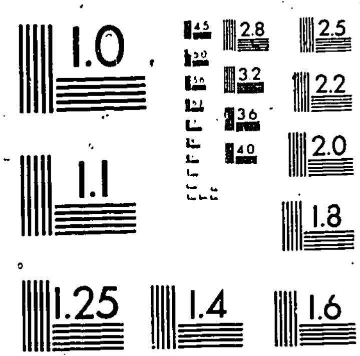


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

DOCUMENT RESUME

D 197 980 SE 034 152

ITLE The Five-Year Outlook: Problems, Opportunities and Constraints in Science and Technology. Volume II.
NSTITUTION National Science Foundation, Washington, D.C.
EPORT NO NSF-80-30
UB DATE 80
OTE 649p.: For related document, see SE 034 151.
VAILABLE FROM Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402 (no price quoted).

DRS PRICE MF03/PC26 Plus Postage.
ESCRIPTORS Agriculture: College Science: Energy: *Government Role: International Relations: *Policy: Public Administration: Science Education: *Sciences: Scientific Enterprise: *Technology

BSTRACT
This is the second volume of a two-volume series roduced in response to the Science and Technology Policy rganization and Priorities Act of 1976. This volume consists of ource materials in three sections: (1) a report prepared by the ational Academy of Sciences which focuses on current and emerging pportunities and problems in selected areas of science and echnology: (2) reports prepared by 21 selected U.S. Government gencies which discuss near-term problems and opportunities related o science, technology, and public policy from the perspective of heir individual missions: and (3) papers prepared by individual pecialists that provide their views on some near-term relationships etween science and technology, public administration, individual elfare, and service activities. (Author/PB)

ED197980

the Five-Year Outlook:

Problems, Opportunities and Constraints in Science and Technology

volume II

National Science Foundation

NSF 80-30

SE 034 152

Preface

This is the second of two volumes of the first Five-Year Outlook on Science and Technology, prepared by the National Science Foundation (NSF) in response to the requirement of the National Science and Technology Policy, Organization, and Priorities Act of 1976 (Public Law 94-282). Volume II consists of source materials in three sections: (1) a report prepared by the National Academy of Sciences which focuses on current and emerging opportunities and problems in selected areas of science and technology; (2) reports prepared by 21 selected U.S. Government agencies which discuss near-term problems and opportunities related to science, technology and public policy from the perspective of their individual missions; and (3) papers prepared by individual specialists that provide their views on some near-term relationships between science and technology, and public administration, individual welfare, and service activities. While the views expressed by the contributors to the first and third components of this volume are their own and do not necessarily reflect the views of the National Science Foundation or the U.S. Government, they do provide an interesting though obviously incomplete range of perspectives on near-term opportunities and problems in the relationships of science, technology and other areas of public concern.

Volume I of the Five-Year Outlook consists of an analysis, summary and synopsis of the materials in Volume II, and an analytical index to both volumes. All components of Volume I were prepared by NSF.

Contents

Volume II: Source Materials

SCIENCE AND TECHNOLOGY: A FIVE-YEAR OUTLOOK
A Report from the National Academy of Sciences

II TECHNOLOGY

III SCIENCE AND THE UNITED STATES

IV INSTITUTIONS

THE GOVERNMENT VIEW

Statements by Selected U.S. Government Agencies

SCIENCE, TECHNOLOGY AND PUBLIC ISSUES:
Papers on Selected Topics Commissioned by The National Science Foundation

I SCIENCE, TECHNOLOGY, AND THE USES OF INFORMATION

II SCIENCE, TECHNOLOGY, AND SOCIOECONOMIC GOALS

III SCIENCE AND TECHNOLOGY FOR THE STATE AND CITY

SCIENCE AND TECHNOLOGY:
A Five-Year Outlook:
A Report from the National Academy of Sciences

Observations

Science and technology continue to transform our world. Our view of the universe about us now ranges outward to the galaxies, back in time to the origin of life and of the universe, and inward into the nucleus of the atom and the molecular basis of heredity.

Science today is ripe with discoveries and new explorations. On the grandest scale, that of the cosmos, it is revealing an array of new objects; X-ray bursters, for example, flashing like a million suns for a second and quiet 10 seconds later. The confusions and blind alleys that marked the search for the very heart of matter now seem to be giving way to clarity. Permutations of only a very few particles, the quarks, rather than of hundreds may account for the structure of particles composing the nucleus of the atom. And through the prism of plate tectonics, we now see a planet whose oceans are growing larger or smaller, whose continents shift, and whose seafloors are younger than its continental masses.

The enormous transformation of biology triggered by the discovery of the nature of the genetic material DNA continues. In neurobiology, powerful instruments and techniques now make it possible to examine the electrical activity of single cells in the brain and to understand in molecular detail how impulses are transmitted among nerve cells, by contact and by newly discovered neurohormones.

The new capabilities and understanding from science and technology continue to mesh with the goals and needs of our society. The prevalence of chronic illnesses such as cancer and cardiovascular disease calls for better understanding of their nature and causes. Pressures on energy supplies are driving investigation of the chemistry of different coals, of the relative hazards of different fuels, and of the possible shapes of a more electricity-dependent society. Government programs—from allocation of funds for wastewater treatment plants to estimating the future number of annuitants of the Social Security System—demand continual refinement of demographic techniques and analyses.

Some of the links between science and societal concerns are more subtle. The carbon dioxide issue is instructive. Its science is concerned with long-term measurements at remote stations—at Point Barrow, Alaska; at Mauna Loa, Hawaii; at the South Pole; at American Samoa—with analysis of factors affecting the rate at which the oceans and the biosphere produce, absorb, and release carbon dioxide; and with complex mathematical models demanding the most powerful computers available. The societal issue is whether the increased release of carbon dioxide from burning fossil fuels, which may lead to higher average climatic temperatures throughout the globe, is an acceptable risk.

Science and technology are, of course, done within institutions—universities, industrial and governmental laboratories, and various organizations for international cooperation. The United States does a large part of its basic research in the universities, whose faculties are both researchers and teachers.

This productive intertwining of research and education has given U.S. science its vigor and also some of its problems. For example, with the present system, the continued high quality of U.S. science depends on the ability of the universities to sustain excellence in basic research while accommodating to several changes. These include the sharp decline in the research growth rate of the 1960's to the more gradual level of the 1970's, the decrease in the potential college population as a result of the decline in birth rates from the early 1960's on, and a diminution in available faculty positions.

Developing new technology into commercial products or processes—innovation—is an even more complex matter than assuring the health of basic research in the United States. One reason is the interaction of economic and technical concerns, invalidating simple analyses. The belief that scientific preeminence goes hand in hand with technological leadership is clearly false. While Japan's scientific stature was until recently not comparable to that of the United States, that country is nevertheless challenging us in many areas of high technology. Scientific excellence is important, but not sufficient for the effective support and introduction of new technology.*

There is now more emphasis on international collaboration in science, occasioned sometimes by the scale of particular research projects, sometimes by the desire to share the costs of major new instruments, by the need for global observations in some areas of science, and to some extent by the recognition of the scientific maturity of other nations.

Space technology and computers have made possible great multinational programs of observation and research in oceanography, meteorology, and geophysics. A vivid example is the Global Weather Experiment, an effort by 140 countries, the World Meteorological Organization of the United Nations, and the International Council of Scientific Unions to observe the world's weather at one time in unprecedented detail and to analyze the data with unprecedented speed. Its goals include reliable weather forecasts over longer periods and better understanding of the dynamics of seasonal climatic change.

Scientific work on an international scale extends beyond basic research. Efforts to improve the economic status and quality of life in developing nations and to strengthen their scientific and technological bases occupy the time and concern of many scientists. Such efforts include large international centers of agricultural research, employing 8,000 people and having an annual budget of $100 million. High-yielding wheat and rice created at these international centers now comprise a third of the worldwide acreage sown with these grains.

This country's involvement in international science and technology is inevitably affected by the changing position of U.S. science. The United States for three decades clearly has been a world leader in almost all areas of science. This is changing as other countries, building on strong economic bases, mount research efforts that in some areas equal and in some surpass our own.

The changing status of American science vis-à-vis that of other countries will continue to prompt examinations of how the United States can best assure the continued vitality of its basic research—whether its investments are properly allocated and of sufficient amount, whether its system for conducting its research principally through the universities is the optimal one, and so forth. Some of that questioning is evident in this report.

Instrumentation is a recurring theme. Electron beam microscopes open a different world than do optical ones. New accelerators are new windows into the ultimate structure of the nucleus of the atom. Electromagnetic sensors, operating at wavelengths we cannot see, reveal new events in the cosmos. And always more is demanded of these tools as we explore ever more distant celestial objects, chemical reactions on the time scale of their actual mechanisms, more subtle and complex genetic controls, and new clues to the forces transforming the face of the earth.

As the contents of this report have been shaped largely by the instruments available, so will the subject matter of future Outlook reports depend on new tools for scientific research and technological development. Computations that once required hours and now take minutes may be reducible to seconds. The large space telescope, the multiple mirror telescope, and the Very Large Array radio observatory will further probe cosmic mysteries. Tools for reading the properties of surfaces will continue to develop, enabling further understanding of corrosion, catalysis, the growth of crystals, and other surface phenomena. The already exquisite methods of studying very fast, minute events among nerve cells will become even more sophisticated.

As we move farther away from the direct perception of our senses in most areas of science, the tools become

*While the complex issue of the relationship of science and technology to industrial innovation is not treated in this report, a very broad analysis by the Department of Commerce of federal policy and industrial innovation is presented in its 1979 report entitled *Domestic Policy Review on Industrial Innovation*.

more complex, often going beyond what can reasonably be built or used by an individual or the typical small university research team. Of course, the very large, extremely expensive instruments such as particle accelerators are already shared internationally. Arrangements for sharing more modest instruments, growing beyond the reach of individual investigators because of their rising costs and complexity, are also needed to sustain first-class research. Some of these arrangements are already in place, and it is likely that new modes of sharing costly instruments regionally and nationally will evolve as universities seek to adjust from the exuberant growth of the 1960's to the more gradual growth of the 1970's.

Regulation is also an issue pertinent to the nation's scientific effort, regulation both of hazardous actions and of possibly hazardous knowledge. The issue appears in different ways, depending on whether it involves automobile emissions, possible intrusions into privacy and confidentiality through the ubiquitous spread of new computer and communications technology, or the problems of nuclear power. The regulation of actions that may be harmful is a proper and necessary function of society. Quite a different issue is raised, however, by the assertion that the creation of knowledge should also be controlled because knowledge in itself is societally destructive or subject to abuse. Basic knowledge can be used for both good and ill. The possible applications of new knowledge are so varied and unpredictable as to preclude the capacity to foresee net harm. Our historical experience tends to support the view that knowledge is better than ignorance and that it is better to regulate applications than knowledge.

A final observation relates to ignorance and to the limits of science. Scientific knowledge is systematic, enormous in its extent, powerful; but it is slight compared to what is not known. Thus, science has contributed precise knowledge on such seemingly esoteric matters as the electronic structure of atoms, knowledge that has been used to create new technologies and indeed new industries. However, we remain uncertain about seemingly common-sense questions, such as the effects of different air pollutants on human health. These uncertainties simply indicate questions whose answers are not yet part of the core of agreed-on science. That core will expand, but it will always be smaller than needed to answer unambiguously all questions asked by society.

I SCIENCE

1 Planet Earth

INTRODUCTION

We live on a restless earth. The evidence is everywhere. Changing ocean tides and the passage of storms confront our senses as part of our daily existence. We are unaware of the imperceptible movements within the earth except when they are manifested by earthquakes and volcanoes. Movements on and within our dynamic planet take place on many scales of time and space, from circulations within the earth's interior over millions of years, to oceanic motions over centuries, to atmospheric storms over hours or minutes.

Our understanding of the planet has changed in recent years. We now believe that the outer shell of the earth is composed of large plates that move like rafts at sea, but only a few centimeters per year. The cumulative effects of these imperceptible motions over millions of years have shaped our planet's surface. The recognition of these plates and the reconstruction of the history of their motions is one of the profound achievements of modern earth science.

We also better understand the complexity of oceanic motions, the fine structure of the great global currents such as the Gulf Stream, and the importance of the slowly moving deep waters of the world's oceans for marine ecosystems and the chemistry of our environment. We are beginning to comprehend the basis of our changing climate, forces that bring on ice ages, and the droughts and floods that plague us.

Our new understanding results from the application of a rapidly changing technology to scientific inquiry. Space technology has given us a finer view of the planet's surface, its oceans, its enveloping atmosphere, and the atmospheres of other planets. Submarine sediments have been sampled at great depths with modern drilling techniques and the earth's interior has been explored remotely with new arrays of seismic instruments.

Finally, we are learning to interpret the new data with the help of modern communication, display, and computer devices. With these, we can compare observations with hypotheses. We can, for example, now electronically simulate probable conditions within the earth, sea, and air; and we can test concepts in a manner beyond the reach of an army of earlier experimenters. We can simulate the intricate, competing chemical reactions occurring within the upper atmosphere, estimate the effect on climate of more CO_2, assess the effects of the oceans on the weather, or test ideas about motions within the solid earth.

We are witnessing a new age of discovery about the earth, and conceptions of the world about us are being altered.

THE CRUST IN MOTION

THEORY OF PLATE TECTONICS

Not since Copernicus displaced the earth from the center of the universe has there been such a revolution in scientists' concept of the planet as plate tectonics. Plate tectonics touches many of the most critical problems of our time—finding new mineral resources and new sites to search for offshore oil and gas deposits, predicting earthquakes, understanding climate changes, and selecting areas for disposing of nuclear wastes.

The concept of plate tectonics is essentially that the rigid surface of the earth is actually divided into separate plates some 100 km in thickness that, in some places, move relative to one another at a rate of several centimeters a year. A plastic semisolid layer beneath the rigid surface lubricates the movements of the overlying plates and may transport them. Some plates, such as that which forms the northern part of the Pacific Ocean, are vast. Others, like those within the Mediterranean region, are smaller fragments wedged between larger plates such as the gargantuan Eurasian and African plates (Figure 1).

We are directly affected. As the Pacific plate slowly moves northwest, it carries with it the rim of North America from Baja California in Mexico to the seaward edge of San Francisco, causing earthquakes along the San Andreas Fault in California.

The Pacific is thought to be a shrinking ocean, and the Atlantic an expanding one. Much of the rim of the Pacific floor dives beneath chains of islands—the Aleutians, Japan, the Philippines, etc.—causing earthquakes. Part of the descending seafloor melts a few hundred kilometers down and resurfaces through the overlying crust to feed volcanoes.

The American and European–African sides of the Atlantic are parted by movements beneath the plates that include the ocean's eastern and western halves. New seafloor is constantly created within a rift valley down the ocean's center, a boundary of the two plates. This rift is in the Mid-Atlantic Ridge, which is part of a global ridge system that extends some 80,000 kilometers over much of the earth.

The ocean basins are geologically young. New ocean crust is continually forming along diverging boundaries between plates. Some continental rocks are more than 3 billion years old; but to date, all rocks recovered from the ocean crust by deep-sea drilling

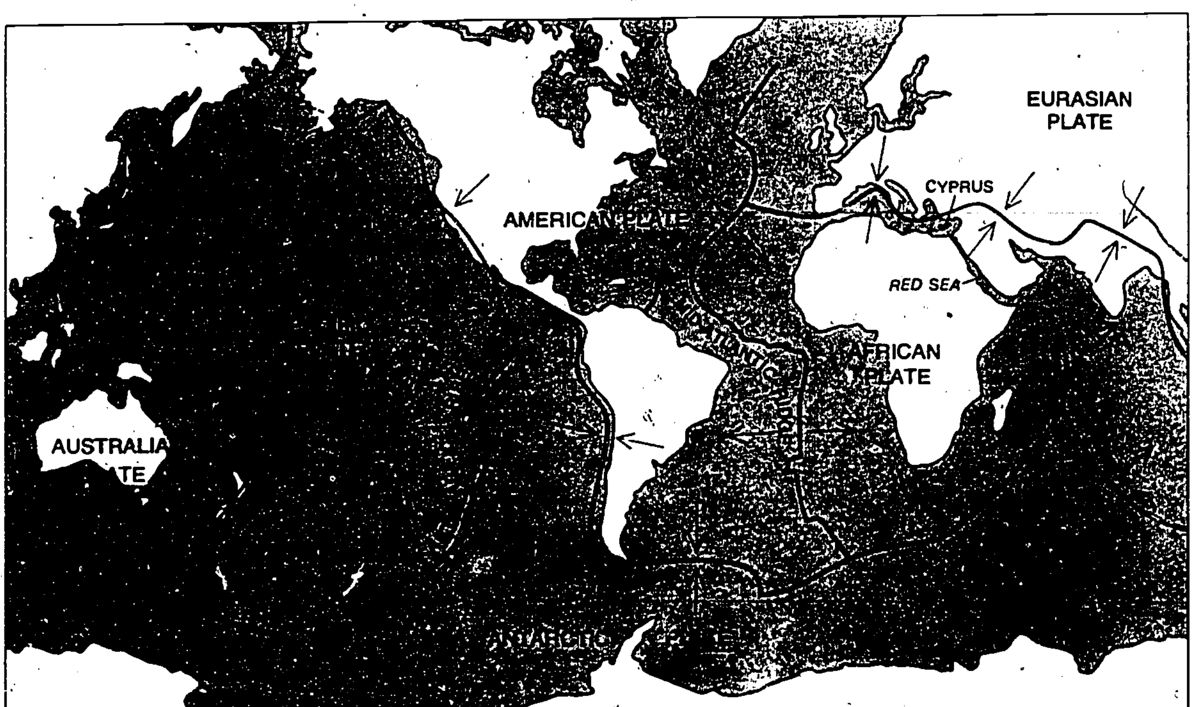


FIGURE 1 The principal plates of the lithosphere, the rigid outer shell of the earth. The paired arrows indicate whether a plate boundary is convergent or divergent. (From "Plate Tectonics and Mineral Resources," by Peter A. Rona, © July 1973, Scientific American, Inc.)

indicate an age of no more than about 200 million years—barely older than the earliest mammals.

A record of plate motion for the last 180 million years is preserved in the magnetized rocks of the ocean crust. Plate motions can be tracked back 600 million years, into the dawn of the Paleozoic era, by using older plate materials that are part of today's continental interior and mountain systems. What happened before 600 million years is uncertain. Our best opportunity to draw inferences about the early history of the earth may well come from comparative planetology—studies of the "terrestrial" planets such as Mars, Venus, and the moon, which have had evolutions significantly different from the earth's.

A UNIFYING CONCEPT

The plate tectonics concept, in providing a logical framework for understanding and linking other earth processes, has helped to unify the earth sciences. For example, the widespread flooding of continents during the Cretaceous period, ending about 100 million years ago, may have been part of the breakup of the supercontinent, Pangea, the primordial landmass from which all present continents were formed. This association is of interest to stratigraphers and petroleum geologists concerned with the distribution or character of marine sediments through time, to paleontologists studying the evolution and distribution of marine organisms, to geomorphologists involved with the effects of sea-level change, and to geodesists, who are interested in sea level itself.

Many ore deposits are related to volcanic activity. The plate tectonics model has provided new clues on the role of subduction, melting and assimilation in concentrating metals, giving economic geologists a new tool to use in identifying promising resource areas. The opening of an ocean, from the first rifting to a fully mature ocean basin, produces a predictable sequence of sediments—important to petroleum geologists investigating offshore areas. Changes in the juxtaposition of the continents have major consequences for both oceanic and atmospheric circulation and are important to paleoclimatologists. Also, the global distribution of vertebrate animals is a puzzle until viewed in terms of moving continents. For example, fossil remains of the ancestral horse, Eohippus, are found in North America, but the modern horse evolved outside of North America and did not appear again on this continent until the age of Spanish exploration.

NEW INSIGHTS, NEW PROBLEMS

The plate tectonics model gives us important clues about what is happening, where it may happen, and why it should happen. But its clues are only suggestive. It tells us where earthquakes and volcanic activity are likely, but not when. It indicates several processes that should concentrate mineral deposits, but doesn't tell exactly where. It accounts for the history of a continental margin where hydrocarbons may be concentrated, but doesn't guarantee that they are actually there.

The ability of the model to reveal the nature of the ocean crust, its age, the nature and thickness of ocean sediments, and the relationship of the depth of the ocean to its age is remarkable. In contrast, it says little about the continents; for example, their history or structure; information important for the safe disposal of radioactive and other wastes on land, or for predicting earthquakes far from the plate boundaries. The model does provide a geological structure into which the continents fit, and, as such, may be looked upon as a starting point for learning more about the continental crust, the lithosphere, and how the surface of the earth forms and is changed.

But puzzles remain. Some postulates of plate tectonics theory remain to be tested, especially the rigidity of plates and their continuous motion. We are still unclear about what moves or deforms the plates. We have only poor explanations of the origin, evolution, structure, and dynamics of the continents and their oceanic margins. The study of the continents, and particularly the deeper, older rocks of the continents, will be of critical importance in understanding the dynamics of the earth. We lack adequate understanding of the early history of the earth; there is no record of the first 800 million years. We need to know more about how chemical elements migrate through the crust and how they become concentrated into economically useful mineral deposits. Information is needed about climates of tens of thousands to millions of years ago, information that may be culled and pieced together from the geological record.

IMPLICATIONS FOR THE FUTURE

The development of the theory of plate tectonics is an excellent example of consolidating information from separate scientific endeavors into a new concept of our planet. Observations over the past decade have yielded a new body of intellectual capital that is now available for a wide variety of applications. Our knowledge about the processes of the solid earth, the formation of the continents, and structures of basin sediments and the deep seabed provides a framework for exploring resources and obtaining data about past environments.

The earth does not easily reveal its secrets. Acquiring our present knowledge has required significant national investments in efforts such as deep-sea geological and geophysical research and drilling programs. What we know now is the product of a long, sustained investment of the nation's funds and scientists. We face decisions about the degree and nature of similar commitments in the future.

We probe the earth for many purposes—some unrelated to science—at very great expense. Maximizing the scientific utility of these efforts will be a challenge to our organizational abilities. Drilling programs illustrate the issue. Drilling is one of the important technologies used to study the solid earth. While industry has drilled on continents for many years for mineral and energy resources, its drill-holes are concentrated in areas where resources are likely to be found. More recently, federal agencies started a range of drilling programs costing over a half billion dollars a year. These include holes drilled for scientific and resource assessment purposes by the U.S. Geological Survey; for geothermal prospecting and development of exploitation technology and for the National Uranium Inventory Project by the Department of Energy; and for military purposes by the Department of Defense. Such extensive drilling efforts afford many scientific opportunities and, conversely, can benefit greatly from participation of the scientific community.

Drilling is one approach to increasing our understanding of the solid earth, but one that must be used in conjunction with others. A broad range of surface and subsurface geophysical research on the structure of the earth's crust and the layer beneath it will depend on the availability of new types of seismic and other geophysical data and new forms of remote sensing from space.

Nor is the understanding we seek about the solid earth a task for the United States alone, but rather, a cooperative task for all nations. This exploration has led to remarkable forms of international collaboration: the Deep Sea Drilling Project, for example, which used the drillship *Glomar Challenger* to verify the plate tectonics concept, and the International Geodynamics Project, which is aimed at understanding the dynamics of the earth's interior. The research opportunities of the next decade should foster even greater international collaboration. Over the next few years, we will have to decide how this collaboration is to be accomplished.

EARTHQUAKES

Our relatively recent understanding of plate tectonics explains much that was previously obscure about earthquakes.

The cost of earthquakes can be enormous. In just the past decade, for example, China lost approximately 700,000 people to earthquakes; Guatemala, 25,000; the Philippines, 10,000; Peru, 70,000; and Iran, 20,000. The United States has been fortunate so far; a densely populated, modern city such as Los Angeles has been spared a great earthquake. Although the death rate in such a modern city should be much lower, costs of damage would be astronomical. In 1971, the San Fernando earthquake, releasing only 1/1,000 the energy of the 1964 Alaskan earthquake, caused nearly three quarters of a billion dollars in damage.

However, new results indicate that we may no longer be so helpless. At least three major earthquakes in various countries have been predicted, saving thousands of lives. At the same time, we are gradually learning how to reduce future losses by building earthquake-resistant structures.

CAUSES OF EARTHQUAKES

Most earthquakes occur as plates of the earth's crust slide past or under one another. Built-up stresses are released suddenly. Plates may jump 10 meters, accomplishing in a few minutes what would ordinarily take decades of steady motion. Earthquakes along plate boundaries include those in San Francisco in 1906; Tokyo in 1923; Chile in 1960; Alaska in 1964; Peru in 1966, 1968, and 1970; and Guatemala in 1976. A rupture as long as 1,000 kilometers may occur along opposite sides of a segment of a boundary.

Enormous quantities of mechanical energy are released when plates lurch past one another. Some of that energy is converted to local heating along the fault, and some is propagated as elastic waves that produce intense shaking of the ground for as much as a minute. Motions may be violent enough near the fault to throw cars and other objects into the air. Structures may be heavily damaged. Farther away, the shaking may be less severe but its duration and periods may be longer. The consequence is a potential for more severe damage to tall buildings when they are farther away. The longer the periods of wave motion, the taller the building that may be set into motion by being shaken.

Large earthquakes occur within plates as well as along their boundaries. In 1811 and 1812, earthquakes in southeastern Missouri changed the course of the Mississippi River and were felt on the eastern seaboard. Boston, Massachusetts, and Charleston, South Carolina, have suffered damaging earthquakes; and many of the devastating earthquakes of China are not along recognized plate boundaries. The tentative conclusion is that stresses within the plates are normally high and that large quakes occur because of a weak spot, a fault, in the plate. No evidence suggests that such earthquakes within plates recur on a regular cycle. Our understanding of these kinds of earthquakes needs to be improved.

EARTHQUAKE PREDICTION

A massive failure of the earth's crust that generates great earthquakes is such a momentous event that it seems unlikely that it should occur without warning signals. Yet, often, as in the disastrous earthquake at Tangshan, China, in 1976, no obvious warnings such as foreshocks are observed. Fortunately, observations over the past decade indicate that warning signals sometimes do precede large earthquakes, even though some may be so weak that they can be detected only by instruments. Scientists have begun to systematically gather data to define classes of premonitory events in order to draw a picture of what happens just prior to an earthquake. Research is underway in China, Japan, the Soviet Union, and the United States.

The Chinese Experience

The earliest indications that earthquakes might be predictable came from ancient China, where farmers noted unusual changes in the level or quality of water in their wells prior to large earthquakes. The Chinese began a systematic program of research in 1968, combining observations by peasant volunteers with measurements of more subtle changes by scientists.

Their approach has been largely empirical, involving extensive observational networks that measure many phenomena. Using simple instruments, tens of thousands of volunteers observe water wells, animal behavior, and electrical currents in the ground. The scientists measure deformation of the land surface, detect and analyze microearthquake patterns, and measure the earth's gravitational, electrical, and magnetic fields. The chief success of the Chinese has been in the public prediction of the Haicheng earthquake of magnitude 7.3 about five hours before it occurred in February 1975. The subsequent evacuation of the populace probably saved tens of thousands of lives.

Many early warnings, precursory phenomena, accompany earthquakes of magnitude 5 or greater. The best documented of these involve distortions of the land surface, foreshock activity, and local magnetic and electrical field variations. Variations in the rate of emanation of radon gas in water wells and a slowing of seismic-wave velocities that typically return to normal just before the quake have also been identified. Geodetic measurements of crustal movements are the most widely used techniques for earthquake prediction in China, the Soviet Union, and Japan.

Occasionally, faults slip and surfaces tilt at the same time. Indeed, strains recorded before modern earthquakes are consistent with a fault having begun to slip at great depths days or weeks before the event. The data are still insufficient for scientists to confirm this deep slippage prior to an earthquake as a general phenomenon. Laboratory experiments, field sampling, and on-site measurements of major fault zones are under way to help develop a better physical understanding of the events that cause earthquakes. Without this knowledge, science will progress only slowly beyond the present empirical stage and prediction will remain an uncertain art.

Monitoring Earthquakes

Although all earthquakes do not give the same warnings, there is general agreement on the data needed to predict them. To monitor the stresses and strains along the San Andreas Fault in California, for example, arrays of telemetering sensors measuring strain, seismicity, magnetic field, and perhaps electric field should be installed along active faults. Instrument spacing would depend on the signal-to-noise ratio and on the scale of the precursory phenomena being monitored. Analysis of the collected data would be performed automatically by digital computers. Geodetic measurements made over distances of hundreds of meters, which can determine rate, orientation, and spatial character of deformation fields, are well suited to these purposes. Prototype systems are in place in California. How soon an operational system can be designed and emplaced depends on acquiring more observations on precursors of moderate-to-large earthquakes.

Moderate-to-large earthquakes now are too infrequent in well-instrumented areas of California for sufficient observations. Only rarely have more than one or two monitoring devices been located near the region of recent large earthquakes, even in China and Japan. Until many more such observations are made, possibly by concentrating effort in seismically active parts of the world, the prediction of moderate-to-large earthquakes will be unsystematic and probably unreliable.

EARTHQUAKE HAZARDS

Damages caused by earthquakes are of several types. Ground shaking may break up buildings, bridges, highways, and other structures—the damage depending on the structure, its interaction with the soil beneath it, and the intensity and duration of shaking. Modern design, particularly in California, has minimized the potential for complete structural collapse and consequent loss of life. However, even when obvious damage seems slight and injuries few, the structure's integrity may be vitiated and replacement therefore necessary. In the 1971 San Fernando earthquake, for example, property losses could have been considerably reduced had better structural design practices been applied uniformly. Masonry structures in California and the rest of the United States may be vulnerable because of poor reinforcement, weak mortar, or inadequate rigidity. Possibly as many as 40,000 buildings would collapse or be damaged so badly as to be unsafe in any future magnitude-8 earthquakes striking near Los Angeles.[1]

Also, subsurface materials can act like liquids and lose their frictional bearing strength because of seismic shaking. An artificial fill near San Francisco Bay in the 1906 earthquake caused severe damage because of this phenomenon. It is confined to areas where subsurface layers are poorly packed and water-saturated. Such areas are widespread in the low-lying lands near San Francisco Bay, and planning for their best use requires detailed study of construction sites.

Earthquake shaking has triggered landslides, especially, as in 1906, when winter rains saturated the soil. Of the 70,000 lives lost in the Peruvian earthquake in 1970, nearly half were people inundated by gigantic ice and rock avalanches.

Fault displacements in great earthquakes may shift the earth's surface. The San Andreas Fault laterally displaced roads, fences, and culverts by as much as seven meters in the 1906 earthquake. Active faults in some geologic provinces have well-defined surface manifestations and can be avoided as building sites, provided they are adequately mapped.

Great offshore earthquakes can produce tsunamis, the seismic sea waves that caused such devastation on the Hawaiian Islands and in Alaska. The arrival of the wave, which is generated by a sudden vertical displacement of the seafloor near the fault, can be timed with moderate accuracy, but a tsunami's height as it hits shore can be predicted only poorly.

Fires or flooding may be the most costly effects of earthquakes. The potential for serious fire damage is perhaps less now than in 1906, when fire accounted for 80 percent of the losses, but it is still substantial. In some places, for example, fire trucks are kept in garages prone to collapse in great earthquakes. Water supplies may be disrupted by pipeline failure. Loss of electrical power, nearly certain in a severe earthquake, would shut down 40 percent of Los Angeles' water services and all of its gasoline pumps. Los Angeles' wide streets would probably confine fires, but a general conflagration could occur under certain weather conditions.

Response

Large cities require pipelines for fuel and water, dams for water storage, and, possibly, nuclear reactors for power. The failure of such critical facilities after an earthquake can be devastating. Schools, hospitals, or power reactors, however, can be sited and constructed to minimize the chance of damage. Little flexibility is available for the siting of dams and pipelines. Pipeline failures should be expected where they cross the rupture of a fault.

A special problem is that of older, earthen dams constructed by hydraulic filling. Because of entrapped water, the core of a dam may liquefy and slump during seismic shaking. Had the water level been 4 ft higher when the Van Norman Dam failed in this manner in the 1971 San Fernando earthquake, the overtopping of the dam and subsequent flooding of the valley would have caused large losses of life and property. Since that earthquake, the state of California has accelerated its program of rehabilitating or removing its 25 remaining earthen dams. Modern dams are much less likely to fail because of seismic shaking. Nevertheless, the risk is still high in densely populated areas within a potential floodplain. Federal dam builders are reviewing seismic safety criteria because of this risk.

REDUCING THE RISK

We now know how to construct or rehabilitate structures to make them earthquake resistant, but who will pay the cost? The added cost for new structures is 5–10 percent, which could add about $2 billion per year to construction costs in the more seismic areas of the United States. To rehabilitate or replace possibly unsafe buildings would cost several billions of dollars in Los Angeles alone. We could assess the value if we knew with some certainty that a great earthquake would occur within the next 10 years. In 20 or 30 years, however, attrition would take care of at least half of the problem; within 50 years, very few of the suspect structures would still stand.

Recurrence intervals for damaging earthquakes are now poorly known; for example, the intervals between San Andreas earthquakes in Southern California now range between 275 and 105 years. Since the last great earthquake on the San Andreas Fault near Los Angeles occurred in 1857, we can only say that a major quake could occur there any time during the next century. However, for Charleston, Memphis, Boston, Salt Lake City, and Seattle, where damaging earthquakes have also occurred, there is yet no basis for guessing when, or even if, they will occur again. Faults and fault motions are not observable in some of these areas by any present technique, and an understanding of these earthquakes is needed.

A timely forecast of a large earthquake could save many lives. If clear warning signals were observed a few days in advance, emergency preparations could be made. Unsafe buildings, if identified, could be evacuated. Fire fighting equipment and standby power sources could be augmented; water levels in reservoirs reduced; and nuclear power reactors shut down to protect against fire, floods, and nuclear contamination. People could move to safe places and take precautions against fires with even a few hours' warning.

The effects of earthquake prediction on local economies and social structures are being assessed.[2] It appears that under certain social conditions, predictions issued one or more years prior to an expected earthquake could hurt the local economy, principally by reducing new construction, but a short-term prediction is not likely to provoke panic.

The uncertainty about when an earthquake may strike contributes to lack of preparation for it. Because the cost of preparedness is high, communities may understandably plead that other matters take higher priority. Since the federal government will be expected to compensate for much of any earthquake damage, federal encouragement of adequate preparation to mitigate the effects of a disaster is appropriate.

The amount of preparation, however, must be based on the severity of the risk. Without knowing when the next earthquake will occur or even approximately the interval between earthquakes, the uncertainty in the degree of risk is so large that only limited measures for preparedness can be taken. Studies of earthquake recurrence are, therefore, critical. The Earthquake Hazards Reduction Act of 1977 (PL 95-124) requested that the President prepare a plan for reducing earthquake hazards by community preparedness, land use planning, insurance, building design, and prediction. To do that, we must develop the techniques and knowledge needed to predict the occurrence of earthquakes. National plans to accomplish this objective are being implemented.

GOVERNMENT POLICY ISSUES

Policy issues need to be addressed at the same time that research on prediction proceeds. Improved contingency planning to enable effective community response to earthquakes is needed. Long-range plans to lessen the social and economic effects of an earthquake prediction need to be formulated. Policies for reducing hazards and planning reconstruction immediately after an earthquake should be adopted. In short, we must seek a better understanding of what communities will do if we can predict with precision that an earthquake will occur.

OUR MINERAL FUTURE

In time, the demand for some mineral resources in addition to fossil fuels will far exceed supply, creating shortages throughout the world.[3] Forecasts of supply and demand are uncertain because accurate estimates of potential reserves do not exist. Also, the mining of lower-grade ores will be greatly dependent on economic conditions that are as yet unknown. Nor is it certain that energy will be available at costs needed to make some deposits economically exploitable.

There are several reasons for potential shortages in the United States: Consumption is increasing; there is a geochemical limit to economically accessible and extractable world mineral resources; current markets do not encourage more active exploration and development of domestic sources; and U.S. access to world supplies may be curtailed. (See Table 4 for U.S. imports.)

UNDERSTANDING MINERAL FORMATION

The principles of mineral concentration are still poorly understood. Events that concentrate elements into deposits are rare: They result from infrequent and fortuitous combinations of many different chemical, physical, and biological processes that, for the most part, are also poorly understood (Figure 2).

Important concepts of element concentration remain untested in either laboratory or field. For example, one concept is that many ore deposits are associated with ancient molten rock intrusions, yet rarely are they identified with modern volcanoes. Do modern volcanoes have ore deposits at depths equivalent to eroded levels of old magma intrusions? Or do ore deposits develop with time as magma cools, and hot, circulating groundwaters extract and concentrate elements from surrounding rock? Or are additional as yet unrecognized factors responsible for the fact that only a small fraction of all intrusions are mineralized, that is, harbor concentrations of various elements?

Ore deposits appear to be related to hydrothermal waters; yet active hot springs and brines—the Salton Sea and Red Sea are examples—only rarely have commercially exploitable concentrations of metals. Are hot springs the end result of an extensive hot-water plumbing system depleted of its minerals by mineral deposition at greater depths? Is there a predictable sequence of elements with increasing temperature and depth? Such concepts of ore zoning have been suggested, but they have not as yet been proven.

Ancient sediments are the most frequent sources of ores. Yet, the source regions and the factors that lead to the concentration of elements are imprecisely defined for these sediments and even for modern sediments. Some investigators believe that the metals dispersed in sediments may be the next major area for resource development.

Biological agents may be responsible for the selectivity and specificity of some concentrations in sediments, yet the relationship of organisms to inorganic mineral deposition is known for only a few living organisms. Can selected organisms be used on a large scale to concentrate specific elements under controlled conditions?

Hypotheses of transport that concentrate elements during rock metamorphism, weathering, groundwater circulations, and evaporation have been formulated. Few have been verified or demonstrated either in nature or in the laboratory.

In short, the major concepts on which efficient exploration must be based have not yet been demonstrated.

THE STATE OF RESEARCH

The federal government, the mining industry, and universities have only limited programs of basic research for understanding how ores form.[4]

The mining industry has neither the capacity nor the tradition for basic research and there is no cooperative research program for the entire industry similar to those supported in the past by, for example, the steel and cement industries. There may be insufficient proprietary advantage for a single corporation to do the basic research necessary for exploration. Mining companies tend to resolve local problems only within those properties in which they are directly interested.

The U.S. Geological Survey performs regional studies, defines potential resource regions, and makes resource assessments. It also undertakes some limited research on the origins of ore deposits and on exploration methods. The Bureau of Mines is concerned mainly with methods of ore extraction and beneficiation essential for economic evaluation. Substantial improvements in these processes will be made when results are forthcoming from the recently established mining and mineral research institutes that now exist at universities in each of 22 states, as mandated under Title III of the Surface Mining Control and Reclamation Act of 1977.

University research programs are limited in number and scope. A typical effort by a university consists of one or two staff members and their students. The reason is a weak demand for highly trained economic geologists, with universities apparently supplying them at a rate just adequate for present industrial needs.

SOME NEW DEPARTURES

There is a need to determine whether the scope of this research—in universities, industry, and government—

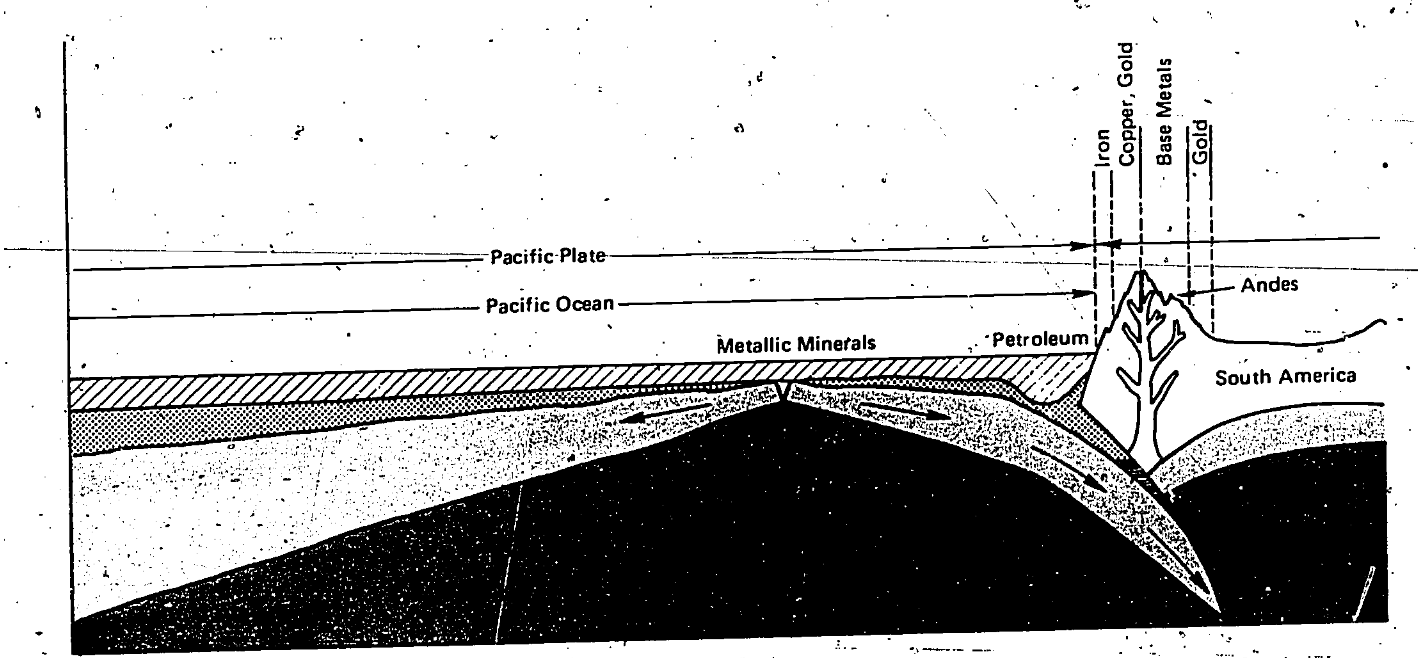


FIGURE 2 Tectonics and mineral formation. The diagram provides one hypothesis of the relationship of plate motion to the formation of mineral deposits. In this view, the spreading of the seafloor outward from the Mid-Atlantic Ridge widens the Atlantic and rafts South America over a trench, or convergent plate boundary. The resulting deformation forms the Andean mountains. Metals also accu-

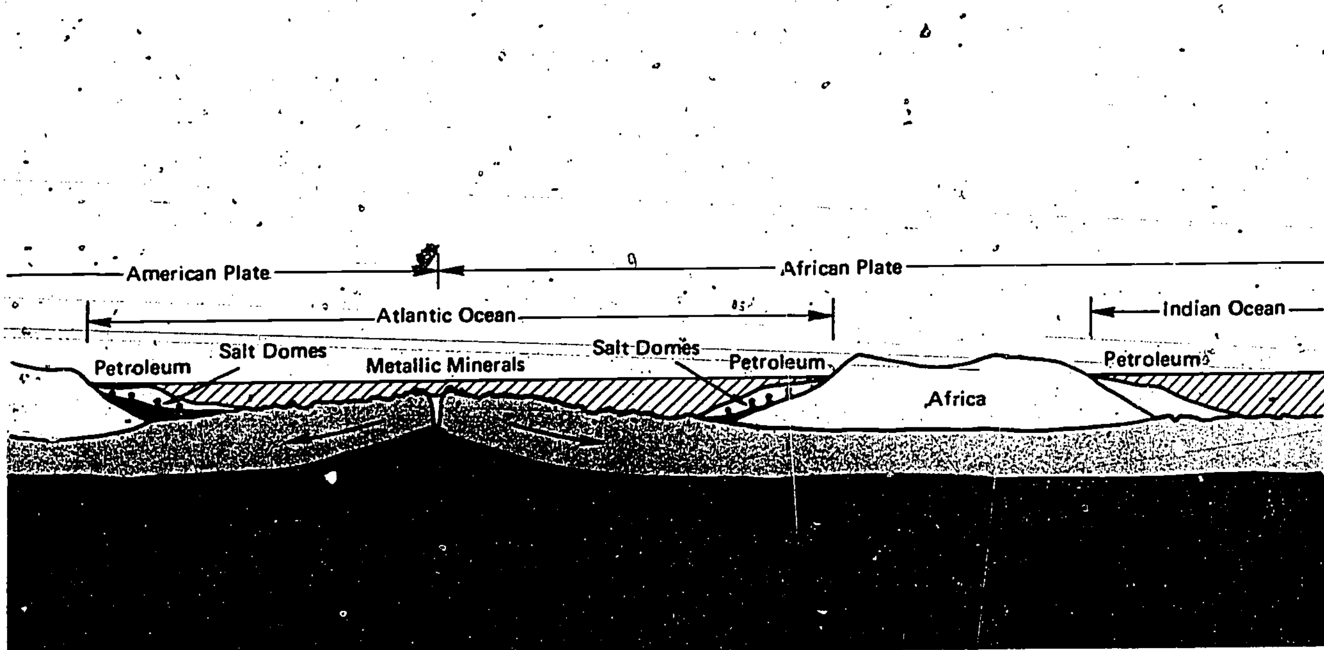
is commensurate with the need for new approaches to mineral resource development. For example, many fundamental properties of rocks should be studied in order to understand ore deposition. Such properties include mineral solubilities, permeability of rocks, flow rates of fluids under various conditions, electrical conductivity, and geochemical changes required for selective mineral deposition. Because these properties are closely related, it is essential that they be measured on the same materials. Mathematical modeling of the many factors affecting mineral economies may be a useful approach to evaluating feasibility of mining, yet little is being done.

Research programs are essential to developing unconventional sources of critical elements. For example, aluminum ore, which is almost entirely imported, could be obtained from the large deposits of dawsonite, alunite, kaolinite, or feldspar in the United States, none of which are now mined for their aluminum content. Other examples include gallium from coal and underclays, nickel from residual laterites, rhenium from molybdenite-bearing deposits, iron from nonmagnetic taconites, manganese from silicates and deep-sea nodules, metals from seawater trapped in sedimentary formations, and copper and uranium from black shales. Often, it is not economically attractive for industry to do research on unconventional sources of critical elements, since economic evaluation for such sources requires building pilot and demonstration plants.

THE STATE OF MINERAL EXPLORATION

Of greatest immediate concern is the present state of exploration techniques, many of which are still primitive and developing only slowly. Exploration seems to be going in two directions. First, because the historic economic practice of basing a profitable mine on a single element may no longer be practical, multiple-element yields are sought in relatively shallow but large-volume deposits, even if the element concentrations are low. Some relatively metal-rich but otherwise ordinary rocks, if widely distributed in an area, can be mined for several elements simultaneously on a large scale; but the environmental impacts are potentially large. Second, some exploration is emphasizing higher-grade deposits at great depths under considerable overburden. Some theories of ore zoning predict higher-grade ores with increasing depth. Geophysical techniques will aim at detecting these buried deposits, which are now prohibitively expensive to drill and excavate.

New geochemical techniques for exploration are appearing slowly. Geophysical techniques are advancing, but only indirectly as the result of other concerns. Thus, the study of rock mechanics has identified properties of rocks that bear on ore deposition, and structural studies outline potential fluid pathways and traps for ore deposition. Radiometric surveys serve to correlate rock units with ore deposits.



mulate about the Mid-Atlantic Ridge. Salt originating in the sediments of the continental margins rises to form large, dome-shaped masses that then trap the oil and gas generated from organic matter. (From "Plate Tectonics and Mineral Resources," by Peter A. Rona, © July 1973, Scientific American, Inc.)

Finally, geological mapping is one of the primary methods for finding ore deposits. Only 40–45 percent of the United States has been geologically mapped on a reconnaissance scale (one inch = four miles). Further, increasing portions of the nation's land are not accessible to private exploration. Geological mapping of the United States and its adjacent ocean areas is necessary to discover major ore-producing regions at a rate adequate to meet anticipated demand and to gain a comprehensive inventory of mineral resources. In addition, there is a continuing need for comprehensive regional studies of mining districts and their surroundings.

ECONOMIC AND LEGAL IMPEDIMENTS TO MINERAL DEVELOPMENT

There are many nontechnical restraints on exploration by private enterprise. For example, taxation of proven reserves constrains owners of a mine to verify only those ores that can be mined within a given year. Another constraint is the withdrawal of federal lands, or the reluctance of companies to develop acquired properties for economic reasons. Lack of milling and processing capacity impedes exploration. The potential for adverse environmental impact dampens enthusiasm to explore many potentially economically attractive areas. Finally, capital requirements are high and economic risks are great, tending to exceed the limits of private enterprises involved in developing major new ore deposits.

The extensive lead time required for developing an operating mine is a significant contributing factor to the potential mineral shortages. After discovery through geological exploration and mapping, and before extensive drilling, sampling, and laboratory study can begin, business must obtain land rights to secure tenure. Only if the environmental impact assessment is favorable can access-road and housing construction begin and development work proceed. Only then can mining begin. The total elapsed time for these steps is approximately 15–20 years for an underground mine and 6–10 years for a surface mine.

ENSURING OUR MINERAL FUTURE

A prudent nation should develop its domestic supplies and make provisions for a flexible program of imports from alternative sources and substitutes. Although the question of imports is a matter of how this nation will interact with other nations in assuring sources of essential minerals, the first step is a matter for national action. A systematic program in which both conventional and nonconventional resources in the United States are explored would seem a wise long-range national investment. There is a clear need for fundamental knowledge about ore deposits and their distribution, as well as for the development of technologies necessary to locate and identify them. These investments, however, will pay off only if the policies and regulatory and legal structures within which mining enterprises must operate encourage vigorous exploratory efforts.

PROMISE OF THE OCEANS

The arrangements among nations governing the use of the oceans are changing rapidly. Through the U.N. Law of the Sea Conference, nations are seeking to define their rights to the ocean's use and their obligations for its protection. A new body of domestic law enacted during the past decade frames our national usage. The era of freedom of the seas is becoming an era of the managed ocean. Stakes are high as nations evolve ocean policies for developing ocean minerals, managing fisheries, protecting the oceans against pollution, performing ocean science research, and protecting coastal areas against conflicting economic interests of other nations. Ocean science has much to contribute in providing the scientific basis for evolving national policies.

THE POTENTIAL FOR MINERAL RESOURCES

The oceans are a source of a few important mineral resources, especially petroleum and gas, sulfur, magnesium, and sand and gravel. For example, the United States obtains 20 percent of its total oil and gas supplies from offshore sources. But it is the potential of mineral resources from the oceans that is intriguing. Cores drilled through deep accumulations of seafloor sediments and into underlying basement rocks over the past 10 years by the drillship *Glomar Challenger*, combined with other geophysical information, have yielded new data about the processes of oceanic mineral concentration. Signs of minerals of many different kinds have been found in various parts of the oceans. For example, a few hundred miles off New York City, beneath 240 m of sediment, cores have revealed layers containing copper, iron, and zinc laid down a short distance above basement rock and volcanic debris formed millions of years ago.

Deep dives by U.S. and French research submersibles have collected rocks that are newly formed by volcanic activity in the valleys of the Mid-Atlantic Ridge and that are often heavily coated with manganese-bearing minerals. Interesting prospects for future exploration for petroleum resources may occur in the deep-ocean sediments off the continental shelf of the United States. Such petroleum and other resources are beyond current commercial development, but they signal that on and beneath certain parts of the seafloor may lie deposits of minerals that might one day be accessible.

Exploration and test-mining of the potato-shaped manganese nodules scattered across vast areas in low latitudes of the Pacific Ocean have been under way for many years. Many of these nodules are rich in nickel, copper, manganese, and cobalt. Mining equipment is being tested aboard the *Glomar Explorer* and other vessels and metallurgical problems of nodules are being investigated. How these mineral resources will be developed and under whose aegis is being intensely debated by the U.S. Congress and the U.N. Law of the Sea Conference.

New technological capabilities and a better understanding of basic oceanic and geological processes are needed to explore and develop nonliving ocean resources. Improved knowledge of the forces affecting the ocean's crust and in active and passive continental margins is needed. The technology for doing this is at hand in the form of improved geophysical techniques such as modern multichannel seismic reflection systems, submersibles, and drilling technology that can safely penetrate to greater depths in deeper waters.

MANAGING LIVING RESOURCES

Realizing the promise of the ocean's living resources means solving different kinds of problems. These, unlike nonliving resources, are renewable; they are also fragile. Fishery and marine mammal stocks have been destroyed by overfishing and changes in environments and habitats. The collapse of the New England haddock fishery is associated with massive foreign and domestic overfishing in the 1960's. The collapse of the sardine fishery off California and the anchovy fishery off Peru have been caused partly by overfishing and partly by environmental changes. Uncontrolled commercial hunting of whales has endangered five of the eight species of great whales. The destruction of coastal fishery habitats, as in Chesapeake Bay, has compounded these effects, since a significant fraction of the ocean fish stocks of the United States spend at least part of their life cycles in the waters and marshlands of coastal areas. Many of these habitats have been destroyed by construction and polluted by human activities.

But we have learned. Our improved understanding of ocean ecosystems has provided a basis for establishing new management policies to maintain the health of our fisheries. The passage of the Fishery Management and Conservation Act of 1976, extending the fisheries jurisdiction of the United States to 200 miles, marked a major turning point. Criteria that provide for a sustained high level of fisheries yield were written into law as a basis for fisheries management, and the law further requires that the best-available scientific information be considered in formulating fisheries management policies and plans. Segments of the fishing industry, in economic straits just a few years ago, are making a comeback. The restoration of depleted marine mammal populations and their protection have been advanced through the passage of the Marine Mammal and Endangered Species Acts, and the International Whaling Commission has introduced new management procedures to protect the world's whales.

Still, wise management of the living resources of the sea has only begun. Many species of fish and crustaceans that abound in the oceans are hardly used to meet human needs; for example, the shrimplike krill of the Antarctic could sustain a harvest of perhaps as much as 50 million tons per year. However, before the development of such resources is begun and other underused fishery resources exploited, it is essential to understand the marine ecosystems that sustain them. The study of marine ecosystems, and biological oceanography in general, should yield practical management tools that will enable us to conserve and maintain our fisheries and other living resources.

UNDERSTANDING OCEANIC PHENOMENA

The influence of oceans extends far beyond their resources. Oceans determine weather and climate. They are the final sink for many wastes. All uses of the seas and their protection against pollution and other insults are advanced by an improved understanding of oceanic phenomena.

TREND TO BIGGER SCIENCE

While field oceanographic research has traditionally required the deployment of ships and other expensive facilities, during the past decade it has become evident that many of the vital problems requiring investigation are so large and require such extensive facilities that they can be carried out only as multi-institutional efforts. To acquire observations of processes involved in ocean–atmosphere interactions, changes in major ocean current systems, circulation of the southern oceans, and coastal upwelling, it has become necessary to mount comprehensively planned, long-term investigations using the ships and facilities of many institutions.

For example, in the North Pacific Experiment (NORPAX), scientists have attempted to examine the nature of the seasonal changes in average temperatures at the sea surface. Large patches of warm surface water covering several hundred kilometers and persisting for many months, and sometimes for years, have been investigated. Scientists believe that these patches, penetrating about 100 meters below the surface, are related to seasonal fluctuations in global weather patterns. Other projects, such as the Mid-Ocean Dynamics Experiment (MODE) in the Atlantic and its U.S./U.S.S.R. extension known as POLYMODE, are examining the nature of the Gulf Stream. These investigations have revealed a complex structure composed of continuously forming and dissipating smaller oceanic eddies. These eddies have been found to resemble in some ways atmospheric disturbances that form along the jet streams of the prevailing westerlies.

Oceanic Research and New Technology

New technology may provide new kinds of data on the oceans. Global synoptic observations of surface ocean conditions are now possible by remote sensing from satellites. These observations will also be useful in providing global information on surface currents and temperatures, wave conditions, surface pollutants, and biological productivity. Another new source of data will be satellite altimeters, which may measure changes in sea level with time and, in association with gravimetric surveys, will indicate changes in the mean sea level.

Although the study of what occurs beneath the surface of the oceans is naturally more difficult, it is critical to understanding three-dimensional movements and the nature of vertical mixing among ocean layers—information important to assessing ocean pollution and ocean-climate relationships. With ocean buoys that sink to set depths, we can obtain information about three-dimensional flows within the oceans. One of the promising approaches is the wide deployment of instrumented buoys that can be interrogated via earth-orbiting satellites.

We need to proceed in new ways, however, and, possibly, developments in theoretical oceanography will eventually help us interpret the internal structure of the oceans in terms of what we see from the surface. Thus, internal waves at moderate depths in the oceans are visible in satellite optical photography. Backscatter radar can detect surface shears. It also should be possible to estimate internal heat flow from direct measurements of temperature gradients near the surface of the oceans.

An additional remote-sensing technology is available using low-frequency, long-distance acoustic ranging. Considerable information about the propagation of sound waves in the ocean—and the technology to generate, detect, and interpret them—has come from the nation's need for more information about subocean acoustics for national defense purposes. By using our knowledge of acoustic propagation in the oceans, it might be possible, with appropriate installations, to map remotely the variability of ocean properties. Such measurements might immediately provide information about the ocean thermocline, of probable value to ocean and weather forecasting. These acoustic techniques have major potential.

OCEANS AND CLIMATE

The oceans affect climate in many different ways. A basic uncertainty in estimating the effects of increased CO_2 in the atmosphere is the fraction of CO_2 that enters the oceans in various forms. This problem will be an important concern of ocean science in the years ahead.

To make reasonable estimates of the amount of CO_2 that can and actually would be absorbed by the oceans, we need better knowledge of the extent of oceanic carbonates and bicarbonates, oceanic alkalinity, and related physical and chemical factors. The extent to which the oceans are a reservoir for carbon depends on the rate at which surface and deeper waters mix. The mixing rate is not well known, but efforts such as the Geochemical Sections Program (GEOSECS) should determine that. The mixing rate may be determined by using radioactive tracers such as tritium from nuclear-weapons tests.

Carbon levels cannot now be measured in the oceans with the precision achieved in the atmosphere, and this technology must be improved. Particularly serious is the lack of information concerning the nature and extent of organic compounds in the oceans. The total oceanic living matter is only about 0.2 percent of the living matter on land, and, logically, has no significant role in the carbon cycle. However, the amount of dissolved organic carbon is comparable to the carbon existing on land. Therefore, dissolved carbon—its origin, turnover rate, and fate—becomes a major factor in estimating the ocean's role as a reservoir for CO_2.

Carbon is but one substance whose fate in the oceans is of concern. Growing pollution of coastal areas, including oil spills, has spurred new studies of the impacts of pollution on marine ecosystems—in the New York Bight and Puget Sound, for example. Legislation has authorized major research efforts on ocean pollution. Protection of the seas includes the wise management of coastal zones and adjacent waters; and knowledge of the fate and effects of many different substances in the oceans becomes essential.

IMPROVING OCEAN MANAGEMENT

Problems of conflicting use are inherent to the oceans. What happens in one part of the ocean affects conditions in another. Some fish migrate from one part of the ocean to another, and habitat destruction in one place, therefore, can damage fisheries in another. Development of oil and gas resources may conflict with access to important fishing grounds. Refinery construction or routes of oil tankers may threaten the oceanic environment.

The United States has established a set of ocean-management policies over the past several years. Laws have been enacted for managing fisheries, developing oil and gas resources at sea, siting deep-water ports, managing coastal zones, and conserving marine mammals. If such policies are to work, they will require an improved base of scientific knowledge about the oceans to illuminate the consequences of various management options.

Finally, a major scientific, political, and international issue of great concern to ocean scientists is that access to ocean areas for research is being increasingly restricted as nations begin to exercise jurisdiction over the oceans within 200 miles from their shores.[5] This issue, being debated in the U.N. Law of the Sea Conference, poses major questions for the Congress of the United States and international bodies. For the continued vigor of the national ocean research effort, we need to ensure maximum freedom for ocean science research.

THE CHANGING CLIMATE

Like the oceans, climate has its own rhythms—rhythms that influence the ebb and flow of civilization. During the warm climates of the Middle Ages, between 900 and 1350, Arctic sea ice diminished, the Norse colonized Greenland and cultivated grains to 60°N latitude in Norway. During the little ice age, from about 1450 to 1750, the climate became colder and more severe, at least in Northern Europe: glaciers expanded, Arctic sea ice returned, and human settlements in Greenland collapsed.

A fluctuating climate has been evident in our own century as a general warming early in the century seems to have shifted toward cooling from the 1940's to the late 1960's. From a geological perspective, however, the climate has been unusually warm for the past 10,000 years. The peak of the last major ice age occurred about 18,000 years ago, when average global temperatures were about 6°C colder and the oceans 100 meters lower.

Evidence about ancient climate comes from deep-sea sediments, fossil records, and other sources. From time to time over geological history, our planet has been subject to major shifts from ice age to warm, largely ice-free periods, such as the present, and back to ice age. Over the last million years, these shifts have taken place at intervals of tens of thousands of years.

CLIMATE AND THE ENERGY CYCLE

Climate is the cumulative product of more familiar weather phenomena—patterns of rainfall, cloud cover, daily and weekly temperature changes, and so forth. Scientists generally consider climate to be the aggregate of weather phenomena for time periods in excess of those for which daily weather prediction is possible—between one and three weeks. As a practical matter, the statistics of weather conditions for periods of a month or greater can be considered as climate. Its character is determined by the interplay between parts of the planet, including the atmosphere, oceans, lakes and rivers, snow and ice, the solid crust of the earth, and plant and animal life.

A global energy cycle links these elements to other parts of the planet and to the sun. Through this cycle, energy can be delivered, extracted, and transformed in the atmosphere. The cycle starts with the sun's energy. This radiated energy creates ozone from molecular oxygen in the high stratosphere and the ozone then absorbs almost all the ultraviolet radiation. Most of the remaining solar energy passes through the atmosphere where it can be absorbed by water vapor or scattered or absorbed by clouds and dust. About 40 percent of the sun's energy is absorbed at the earth's surface, mainly in the oceans. About 30 percent is reflected back into the atmosphere and space.

The energy cycle has only begun. The oceans and the atmosphere become working elements of a global energy transfer system. Ocean currents move energy to other parts of the planet. Energy is injected into the atmosphere when water vapor is evaporated from oceans, lakes, and rivers. The vapor is transported by winds and released in other parts of the atmosphere when it condenses as clouds and rain. Storms in the atmosphere transform potential energy from heating and cooling to the kinetic energy of the winds, which in turn transfers energy from the tropics to polar latitudes.

Anything interfering with or modifying the channels by which energy is moved and changed may in turn change the climate. There are many possibilities. Dust from plowed land, cleared forests, or volcanic explosions can significantly affect the transmission of solar energy. Gases produced by industry or by natural biological and physical processes selectively absorb transmitted solar or earth radiation. Changes in ocean temperatures or varying snow and ice cover affect the global distribution of energy.

We need to find out how these and other perturbations interact and what their effects are on the global circulations that determine climate. Fortunately, science and technology are developing ways to measure and understand the consequences of these interactions.

GROWING VULNERABILITY TO CLIMATE CHANGES

Natural fluctuations of climate frequently and, often, severely disrupt society. Moreover, pressures of growing population, demand for greater food supplies, and deterioration of land intensify the impact of climatic fluctuations. Recent examples include the drought in the Sahelian countries of Africa in the late 1960's and early 1970's, which caused famine and death on a continental scale; the 1972 and 1975 droughts in the Soviet Union, which affected the world grain markets; the failure of the Indian monsoons in 1974, which caused a food crisis in that subcontinent; and the abnormally cold winters of 1977 and 1978 in the eastern and midwestern United States, which closed schools and industries and caused widespread temporary unemployment. Of course, changing climates can benefit society: Bountiful grain harvests during the past two years have produced the grain surplus the world now enjoys.

Human Impact on Climate

We have growing evidence that not only nature but also humans may alter the climate; that industrial and agricultural practices may be causing climatic fluctuations not only on local but also on global scales.[6] We have inadvertently learned to alter the basic energy processes that control the climate in ways that are subtle but may have large effects. Humanity is not only vulnerable to climate, but climate now appears to be vulnerable to humanity. We must confront the fundamental question: Are we changing the climate of the earth?

How can we change it? A straightforward way is to dump great amounts of heat into the atmosphere. We already have done this on an urban scale. Cities are concentrated heat islands; in some cases, the amount of heat energy released in the city is equal to or greater than the average amount of energy absorbed from the sun. But since only within cities does the heat released approach a level comparable to that received from the sun, we can discount direct heat addition as a global concern, at least for the next several centuries.

There are, however, more subtle influences. Gaseous wastes from industry and agriculture selectively absorb different wavelengths of radiated energy. In some important instances, these gases are transparent to incoming shortwave solar energy but opaque to longwave infrared radiation from the earth's surface. The wastes thus trap energy, much like greenhouses do, hence the "greenhouse effect." The net result, in the absence of other processes, is a warming of the lower atmosphere.

The CO_2 Problem

Carbon dioxide, which absorbs infrared radiation, is released to the atmosphere when we burn fossil fuels. More is added when we clear forests and either burn the debris or allow it or the soil humus to decay. The CO_2 content of the atmosphere has been recorded at a network of observatories extending from Point Barrow, Alaska, through Mauna Loa, Hawaii, and American Samoa to the South Pole. Everywhere, the record is similar: The CO_2 content in the atmosphere indicates a systematic increase over the globe.[7]

The amount of CO_2 in the atmosphere seems to have increased about 10 percent in the past 50 years, or from about 300–335 parts per million. Projections of fossil fuel use suggest that the amount of CO_2 will reach about 380–390 parts per million by the end of the century. By the middle of the next century, the present level of CO_2 would be double, and, by 2100, quadruple current levels.

Mathematical models of climate indicate that global surface temperatures will increase as CO_2 levels increase. For each doubling of CO_2, the models project a 2–3°C average global temperature rise. The inference is that a global surface temperature increase of 4–6°C would occur by the end of the next century. Such temperature differences are characteristic of major climatic shifts to and from ice ages.

It is important to distinguish clearly between observed facts and projections based upon calculations. That CO_2 is increasing is a fact established by measurements. All else is calculation and projection. There are major uncertainties. For example, the observed increase in CO_2 is only about 50 percent of the amount estimated to have been introduced into the atmosphere during the first half of this century by burning fossil fuels. While it is likely that much of the remaining expected CO_2 has entered the world's oceans, we have no verification; and we are uncertain about the role of plant and animal life in the CO_2 balance.

Other Problems

Although CO_2 levels have been increasing since the beginning of the industrial age, there are more recent insults. The chlorofluoromethanes (CFM's) used as propellants in spray cans and as refrigerants absorb long-wave infrared radiation from the earth, as does CO_2, and are transparent to shortwave solar radiation. Over the next several decades, a continuation of past production rates of CFM's could result in a global surface temperature rise of up to 1°C as early as the year 2000. This projected temperature increase would be in addition to that caused by CO_2. There is also nitrous oxide, another infrared-absorbing gas released when green plants convert nitrogen to their use. The atmosphere has long since adjusted to natural releases of this gas; but increasing use of commercial nitrogen fertilizers may significantly increase the amount of nitrous oxide released into the atmosphere. A doubling of the use of nitrogen fertilizer would significantly add to the warming caused by other gases.

Other human practices may pose problems for climate, though none so far reaching as those that add infrared-absorbing gases to the atmosphere. Anyone who has observed smog in Los Angeles, the dust raised by a plow, or the smoke plumes of factories can appreciate that particles can obscure sunlight. Particles both scatter and absorb sunlight; and scientists are still not sure whether the net effect of aerosols is to increase the temperature of the lower atmosphere because they absorb sunlight or decrease it because they scatter sunlight back to space.

Changing patterns of human habitation, and more significantly, agriculture, may also affect climate. Generally, cleared areas reflect more sunlight; and when trees are replaced by crops and grassland, the amount of sunlight reflected from that land increases, especially when covered with snow. The consequences of this for the energy balance of the atmosphere are not well known.

Acid rains are of increasing concern. While not directly a climate problem, the location of acid rains and their duration is controlled by the climatically determined wind and rain patterns in relation to the source regions of the industrial effluents that cause them. These rains, extremely dilute acids, form from the entry into solution of sulfur oxides and other industrial pollutants. They are deleterious to crops, forests, and natural ecosystems. Fish populations in lakes in the eastern United States and Canada and in Scandinavia have been affected.

The Stratosphere Problem

There recently has been increasing concern about the effects of contaminants within the stratosphere.

Temperatures tend to drop with increasing elevation in the lowest layer of the atmosphere, the troposphere, leading to vertical air movements that mix air masses. And tropospheric precipitation tends to wash out pollutants. In the stratosphere, however, temperature increases with height, suppressing vertical mixing. There is no precipitation. Substances injected into the stratosphere remain there for a long time. Therefore, that region is particularly sensitive to the cumulative effects of small amounts of contaminants.

The stratosphere contains small concentrations of the gas ozone, maintained by photochemical processes. Ozone, while a minor constituent of the stratosphere, nevertheless is important to life on earth, for it absorbs those wavelengths of solar ultraviolet radiation that are destructive to life and that can, for example, induce skin cancer. Any long-term change in average abundance of ozone will also affect the temperature structure of the stratosphere. Because of the interaction between the stratosphere and troposphere, such a temperature change could affect the circulation of the lower atmosphere and, therefore, the climate.

The balance between ozone formation and destruction can be shifted by the addition of chemicals such as oxides of chlorine and nitrogen. The oxides of chlorine come mainly from chlorofluoromethanes and the oxides of nitrogen mainly from nitrous oxide flowing from metabolism in plants and from the effluents of stratospheric aircraft or rockets.

These chemicals can catalytically reduce the amount of ozone in the stratosphere. If one assumes 1973 rates of production of CFM's, calculations indicate a possible future reduction of from 7–15 percent in total ozone. Observations do not yet show such a reduction since the normal variability of ozone is large enough to obscure it. The effects of an ozone reduction upon global surface temperature would be slight, but those on the stratosphere, human health, and global ecology would be significant.

The history of predicting the effects of nitrogen oxides on the ozone layer illustrates the need for caution in advising the government. It was suggested in 1970 and 1971 that nitrogen oxide might reduce ozone.[8,9] Calculations indicated that reductions up to 23 percent might result from operating a fleet of supersonic transports. The Department of Transportation, at the request of Congress in 1971, initiated its Climatic Impact Assessment Program[10] to determine the effect of SST operations on ozone. In 1975, after several years of investigation, that study and a parallel one by the National Academy of Sciences, supported such projections and indicated that ozone reductions of about 10 percent were likely from a large fleet of supersonic transports.[11]

Deep concern about these effects led to expanded investigations of the critical photochemical reaction rates in the stratosphere. The original assumptions about chemical rate constants in the stratosphere were modified by improved laboratory and field measurements. The new evidence, although not yet definitive, now suggests that nitrous oxide additions to the lower stratosphere by SST's might increase the amount of ozone, reversing earlier projections. But in the case of chlorofluoromethanes, the effects have not only been confirmed but also found to be twice as destructive of ozone.

If, as it now appears, we will have to address the climatic impact of CO_2, the oxides of nitrogen, CFM's, and as yet unknown substances in the future, we will need a more systematic approach to understanding the processes that control their atmospheric concentrations.[12] This means understanding how they are cycled among the ocean, atmosphere, biosphere, and solid earth. The emphasis now is on the biogeochemical cycles of carbon, nitrogen, and chlorine. We are most likely to influence climatic processes through these cycles.

IMPACT OF NATURAL CLIMATE FLUCTUATIONS

Our vulnerability to natural climatic changes will continue and intensify. Natural seasonal and interannual fluctuations of climate will continue both to plague and to benefit us. Because of increasing world demand for food and energy in the face of limited supplies, the impact of natural climatic changes will become of greater concern to all nations, including our own.

Since the 1930's, we have been fortunate to have experienced generally good growing weather in the major grain-producing regions of the United States with only occasional, short-lived dry periods. In some years, as in 1972, the world depended on U.S. grain to make up shortages. At other times, 1977 and 1978, for example, the world as a whole experienced good growing conditions. Fluctuations between world shortages and surpluses probably will be more frequent as world food supplies tighten in the future.

Improved warning systems and pre- and postdisaster planning and assistance can help us deal with the impacts of climatic change. Food reserves, water storage facilities, and postdisaster financial aid are prudent options for governments until predictions are improved.

The very long-term cycles of climatic change marked by the beginning and end of ice ages are now widely thought to result from changes in solar energy reaching the earth due to changing orbital relationships between the sun and earth. There is evidence that the long-term (thousands of years) trend is a return to a colder climate. Some scientists have suggested that the projected warming of the global climate from increasing CO_2 is a short-term counter to this trend. The long-term changes in climate due to the variations in the amount of solar energy reaching the earth will probably be gradual, although there is much uncertainty about the suddenness of the termination of an interglacial period and the beginning of a new ice age.

IMPACT OF CLIMATIC CHANGES CAUSED BY HUMAN ACTIVITIES

The most difficult policy choices arise from climatic fluctuations that occur over decades to centuries. Evidence already recounted on the effects of infrared-absorbing gases suggests a warming of the earth's atmosphere over the next several centuries. But, to repeat, this is a very uncertain projection.

What would a warmer or cooler earth be like? What changes in storm tracks may occur? How will patterns of rainfall and the variability of temperature be affected? What changes might occur in the planetary circulations that determine climatic belts, or the location of deserts and rain forests? How will such changes affect polar ice caps and sea ice? We can now only speculate about such questions; mathematical models have given us some insight. But if we were wise enough to read nature's record more exactly, we might be able to say more, because nature has provided past examples of climates warmer or cooler than the present one. Global temperature increases of nearly the same magnitude now projected, due to more CO_2, actually occurred 4,000–8,000 years ago. Paleoclimatologists and geologists have attempted to portray past climates from ocean sediments, pollen and lake sediments, histories of lake conditions, records of mountain glaciers, and other sources. Our knowledge is sketchy; but evidence is strong that several thousand years ago present subtropical deserts were wetter than they are today. North Africa was probably more favorable for agriculture; parts of Europe and western North America were wetter, and eastern North America drier.

The lesson of the past and our knowledge of climatic processes suggest that the largest effects on agriculture will be felt in the arid and semiarid regions, where rainfall is impeded by descending air motions that are associated with semipermanent high-pressure circulations. Such belts of descending air motion would likely move poleward and perhaps widen. On the other hand, with more CO_2 in the air, there is experimental evidence that crops, grasslands, and forests may grow at a faster rate.

The effects of a warmer atmosphere might be most noticeable in the polar regions of the world, possibly through a reduction in the amount of sea ice. The projected warming may eventually cause sea ice to melt, leaving an ice-free Arctic Ocean. Many scientists consider the West Antarctic ice sheet to be unstable to modest changes in temperature; there are some signs that this ice sheet is already retreating. During a significant warming, it might retreat much faster, leading to global rises in sea level of several meters.[13] In general, however, the effects on sea level are not expected to be large over the next century.

TASK FOR SCIENCE AND TECHNOLOGY

Given the unprecedented stakes, the task for science and technology during the next several decades is to reduce the uncertainties in our knowledge of climatic change. National and international scientific and technological efforts are under way. The first task is to monitor climatic variations and their causes. The technology for many aspects of this monitoring is available: for example, the global monitoring platforms provided by space technology. Earth-monitoring satellites—the weather satellites, Landsats, and Seasat—let us monitor the earth's radiation budget, trace gases, changes in the biosphere, the extent of snow, ice, and cloud cover, sea-surface conditions, and other important processes and features. More conventional land- and sea-based observations will also be needed, such as those already made at stations that monitor climatic change and CO_2 levels. There is no longer a technological bar to making many necessary observations.

Mathematical Models and Climate Predictions

The central problem is to understand the dynamics of climate: the forces that shape it. The mathematical model, which, with computers, can simulate the climate system, is one key to understanding these dynamics. With such models, we can study the probable causes of climate variations and the sensitivity of climate to human activities. Mathematical models have already been used successfully to simulate the first-order effects of CO_2, volcanic eruptions, and stratospheric changes. However, many potentially interactive processes were neglected in those simulations, and there is currently no model that fully represents the interactions among the atmosphere, oceans, and cryosphere. Only with such complex models can we understand and simulate important feedback mechanisms among the various parts of the climate system.

There is a more fundamental question. Is climate predictable? We know that our ability to forecast the daily progression of storms and other weather events decays with time to a point where forecasts are no better than guesses. The reason is that we can neither precisely specify the initial conditions of the earth–atmosphere system as a whole nor precisely describe all of the physical processes involved in the day-to-day evolution of weather. Small errors due to our lack of knowledge of the state of the atmosphere or in approximating physical processes inevitably become magnified until they limit our ability to forecast weather. The theoretical limitation to our ability to predict daily weather events is somewhere between one and three weeks. We need to know the extent to which climate is predictable. This question remains unresolved.

SOME SOCIOECONOMIC ISSUES

Societal issues raised by the prospects of human influences on climate could not be more pervasive. Energy policy is an example. Best estimates today indicate that demand for oil will outrun supply, and we will turn to coal as one of our basic fuels until new and renewable sources of energy can take over. The primary limiting factor in energy production from fossil fuels in the next few centuries may turn out to be the climatic effects of the release of CO_2. How will we balance our energy needs, usage, and allocations against possible impacts of a changed climate?

Food and agriculture policy issues are equally grave. The specific effects of climate variability or changes on agricultural productivity and a knowledge of where specific shifts of climate will take place are uncertain; but it is not too early to apply existing knowledge about variability and to begin thinking about adjustments that might be necessary in food supply systems. What changes will be needed in water conservation, storage, and irrigation methods to accommodate a different climate? Should we step up development of drought-resistant crop strains? The time when the amount of CO_2 will have doubled, about a century from now, is also the time when a tripling of world food production will be needed to meet the demand of increased population. Since the world's best agricultural land is already under cultivation, special attention will have to be given to raising productivity on existing cultivated lands and reducing the vulnerability of more marginal lands to climate. How will the United States assist in doing this?

Many aspects of the problem have been recognized. The U.S. Climate Program Act of 1978 provides for the necessary study of climate, its changes, and its impacts. The World Meteorological Organization (WMO), other U.N. agencies, and the International Council of Scientific Unions have joined to mount a World Climate Program.[14,15] Steps already taken in planning this program are the first toward a common basis of understanding climate processes and their possible consequences. Such a common base is needed for any international measures to mitigate the adverse impacts of climatic variability or changes. Our concerns can stimulate a new appreciation among nations that climate is a common resource essential to all and its protection a common responsibility.

VIEW FROM SPACE

The advent of the space age two decades ago put the earth's landmass, oceans, and atmosphere in a new perspective. What earth-borne sensors were piecing together as a puzzle, space-borne sensors began to integrate as a whole. The view from space, whether of the earth or of other planets, has contributed to our understanding of the physical processes of our planet, our philosophical concept of the earth as our habitat, and is beginning to help us to monitor the state of certain earth resources and the earth's environment.

The technology for remote sensing of the earth and other planets from space is evolving rapidly. During the next decade, the Space Shuttle will help to advance this technology; and new earth-application satellites will expand the range, diversity, and utility of data from space. Science missions to other planets will offer additional prospects for new understanding of the earth.

LOOKING AT THE EARTH

The events and features of the earth we wish to monitor occur on many scales of time and space. Tornadoes and squalls have lifetimes of minutes and hours, and continuous observation is necessary. Large-scale storms can be monitored with observations taken only once or twice a day; ocean currents and temperatures, once a week; crops, once every two or three weeks; and some geological processes, from months to decades.

Designing a space-monitoring system to detect specific events is a difficult technological and scientific task. The system includes sensors ranging from a simple camera to advanced electro-optical scanning devices or radar systems to collect data; satellite platforms with support functions such as power and thermal control and communication links to send data to ground-receiving terminals; data-processing systems, either on the satellite or on the ground or both, to convert sensed information into usable form; and data-dissemination devices to provide useful information to specific users. Each system is a chain from observation to use of information; and it is essential that all links be planned and implemented in concert if the observations are to be of practical use.

Types of Satellites

Frequency of observation from satellites depends on orbital conditions such as altitude, eccentricity of orbit, inclination to the equator, and the local times at which the satellites pass over a point on the earth. The orbit depends on the specific mission. For example, geosynchronous satellites are enormously valuable for tracking global weather patterns continuously. These satellites, when placed at an altitude of approximately 35,000 kilometers over the equator, appear to hover at a fixed point above the earth. Such a satellite can continuously monitor about one-fourth of the earth's surface but cannot adequately view high latitudes. Four geosynchronous satellites can continuously cover most of the earth but, because of their high altitude, may not provide the detail required for many applications.

Low-altitude satellites can reveal small-scale features, but cover a particular area on earth less frequently.[16,17] For example, Landsat has a circular, near-polar orbit at an altitude of approximately 915 kilometers. It completes 14 orbits per day, and observes a given area at the same local time every 18 days (Figure 3). For some uses, such as forest and range inventories, observations every 18 days may be adequate. More frequent observations may be necessary to assess storm damage or episodes affecting yields of food crops.

Remote measurements of our planet from space record electromagnetic radiation reflected, scattered, or emitted by the earth and its atmosphere.[18] Visible light represents only a small fraction of the electromagnetic spectrum that is important for remotely sensing the earth and its atmosphere. Instruments are available that can sense across a wide range of the electromagnetic spectrum, from the very short ultraviolet wavelengths to long radio wavelengths. Interpreting these observations requires an understanding of the characteristics of solar radiation at different wavelengths and the variations with wavelength of the reflection, absorption, and emission properties of the land, its vegetation, the seas, human objects, and the atmosphere.

Using television cameras, scanning devices, microwave radiometers, and other sensing devices, we have been able to detect a wide range of natural and human phenomena from space. Volcanoes, movements along geological faults, coastal erosion, flooding, storms and hurricanes, and forest fires all have been monitored by satellite imagery or data-collection systems.

Earth Applications

Events of the past two decades have demonstrated practical applications of information from space.

Landsat satellites have provided information useful to agriculture. Experiments in forecasting yields of wheat in the United States, the Soviet Union, and Canada using satellite observations have been encouraging. The conditions of crops and forests have been assessed.

Information from Landsat and meteorological satellites, such as areas of water impoundment and snow cover, has helped in managing the country's water resources.[19] Snow mapping by satellite can improve

FIGURE 3 Cape Cod region seen by LANDSAT in orbit some 900 kilometers above the earth's surface. (USGS EROS Data Center)

predictions of runoff from melting snows. Operational environmental satellite data are used routinely to monitor snow cover in scores of river basins in the mountainous western United States.

Landsat data represent a new source of information for geologists looking for minerals, petroleum, and geothermal energy sources.

Weather services depend in many ways on satellite data for essential weather and severe storm warnings. Geostationary weather satellites monitor hurricanes and tornadoes and provide the basis for improved national and international severe-storm-warning systems. Polar-orbiting weather and environmental satellites provide data about global cloud cover, temperature, and moisture, enabling major improvements in worldwide computer forecasting.[20]

EARTH SCIENCES

The earth sciences are among the major beneficiaries of earth-viewing satellites and other space technology. New Landsat systems will carry a new generation of instruments for geological studies. These instruments include multispectral remote sensing for discrimination among rock types, active and passive microwave systems, and stereoscopic imagery for the third-dimension information necessary to make structural interpretations of space-sensed data.

The solid earth will be studied as a global entity, its geological aspects with the new Landsat, its geophysics with special satellites to measure the earth's gravitational and magnetic fields, and its geodynamics with two other satellite systems that have been under development over the past decade. Both systems are capable of measuring positions on the surface of the earth within a few centimeters, the accuracy needed to measure plate motions. They will provide laser ranging to natural and artificial satellites and very long baseline microwave interferometry. These measurements are important in understanding the nature of earthquake occurrence and, hence, in the design of practical methods of earthquake prediction.

Ocean sciences have benefited from many earth-viewing satellites. Geodetic satellites, such as the Geodynamic Satellite-3, yield measurements of sea level and the earth's gravitational field. Meteorological satellites and the Seasat satellite provide a synoptic view of global ocean conditions.

The meteorological sciences are midstream in the international Global Atmospheric Research Program in which satellites play a central role. The objective is to examine the predictability of daily weather events and provide data to examine the dynamics of seasonal and annual variations in climate. The nations of the world are conducting the Global Weather Experiment during 1979. This experiment involves the deployment of five geosynchronous satellites launched by the United States, Japan, and a group of European nations; polar-orbiting satellites launched by the United States and the Soviet Union; and fleets of aircraft, surface ships, and unattended buoys. Only earth-orbiting satellites could have made such programs possible.

EARTH AND THE OTHER PLANETS

Additional insight into atmospheric and solid planetary processes is being obtained by studying our sister planets. Evidence of major past climatic changes has been discovered on Mars in the form of the channels on the surface and the extensive layered deposits that surround its north polar region. The knowledge that such changes have taken place on other planets may help us to understand causes of climatic change on earth. As other planetary systems are examined, we should gain further helpful knowledge. For example, scientists believe that the very high surface temperatures on Venus are the result of a runaway greenhouse effect. More information on that planet and its atmosphere could help us understand the potential hazards of similarly increasing atmospheric CO_2 on earth. Satellite studies of Venus are just beginning.

By studying the atmospheric circulation on Mars and Venus, each with its unique atmosphere, rotation rate, inclination, and distance from the sun, and by comparing what we learn with the general circulation of the earth's atmosphere, we may be able to construct computer models that can predict the atmospheric circulation of any given planet.

Insight into planetary volcanism and plate tectonics has been important to the geological sciences (Figure 4). Basaltic volcanism is common to terrestrial planets and the moon. Planetary basaltic lavas from the mantle make their way to the surface by the same mechanisms as on earth—from fissure and central-vent eruptions. Basaltic volcanism on the moon was confined to a restricted period of lunar history between 3 and 4 billion years ago. If it is assumed that the Martian lithosphere is so thick and rigid that movement is difficult, it is possible to deduce that basaltic volcanism lasted much longer on Mars, a bigger planet, with eruptions from single central vents persisting much longer than on Earth.

Our knowledge of terrestrial processes is augmented as information on other planets accumulates. We are also beginning to gain further knowledge of such diverse topics as early planetary evolution, early crust formation, earthquake mechanisms in nonplate-tectonics settings, tectonic processes, and many others. Terrestrial application of planetary discoveries is just beginning. Those benefits can be expected to continue as we visit and revisit the planets. The remarkable and

FIGURE 4 Volcanic eruption on Io, a satellite of Jupiter. This plumelike eruption rising almost 100 kilometers above Io's surface indicates that the satellite may have the most active surface in the solar system. The volcanism is very explosive with velocities above 3,000 kilometers an hour, or faster than any observed on terrestrial volcanoes, including Etna, Vesuvius, or Krakatoa. (NASA)

unexpected results from the Voyager mission to Jupiter and the recent measurements on Venus are tantalizing indications of what lies ahead.

SPACE SHUTTLE

The introduction of the Space Shuttle[21] in the next decade will mark a major operational change in the U.S. space program. Its advent as a relatively inexpensive transportation system will open an era of more ambitious operations in space. The shuttle should reduce the cost of transporting satellites into earth orbit. With its 60-foot by 15-foot diameter compartment, larger and heavier payloads can be lifted into space. This large volume may be essential for earth observations that require large antennas. The shuttle will return payloads to earth for refurbishment and reuse.

An important attribute of the shuttle, which is to be flown in the 1980's, is its potential for reducing the cost and time required to test new remote-sensing systems. Development tests of such systems now require expensive, self-contained satellites that must include power, propulsion, pointing, command, and telemetry subsystems. With the shuttle and its manned onboard module, the Spacelab, development prototypes can be flown to investigate how useful certain measurements are, and to optimize the design characteristics of future flight sensors.

USING DATA FROM SPACE

Users face formidable problems in applying satellite data. A single frame of the Landsat Multispectral Scanner is made up of 32 million picture elements with six bits of information each; 192 million bits of information for each frame. The weather satellite of the National Oceanic and Atmospheric Administration yields about 100 billion points of new data every day. Computer processing is indispensable.

Computers can construct images to closely resemble the original scene.[22] Geometric distortions introduced by the sensor or by satellite and earth motions can be computer-corrected. Computers can calibrate and compensate for certain data errors. They can enhance images, modifying the original data to emphasize features of special interest. For many purposes, however, digital information rather than images is required, and computers process these directly.

By comparing measurements in different spectral bands at different times, the computer can differentiate between corn and wheat, determine, in conjunction with other data, the health of plants, and help predict crop yields. Further research on crop "signatures," models of growth and yield, and pattern-recognition processes is leading to automatic extraction of specific information that agricultural users need. Automated systems already are used to produce atmospheric and surface-temperature analyses and displays of upper-air wind fields.

Solid-state devices are advancing remote-sensing technology. Arrays of silicon charge-coupled devices have made possible solid-state, visible-light cameras of very high sensitivity and resolution for use on the ground. When flight modules are developed for these cameras, reduced cost, weight, power, and volume can be expected. Similar developments coming soon are solid-state array sensors for various regions of the infrared spectrum to replace mechanical scanners currently in use. In addition to enhancing sensitivity and resolution, an important added advantage of solid-state array sensors will be that signal-processing circuitry will be combined with sensors within the array to select the information to be transmitted.

Current developments in microprocessors, charge-coupled devices, memory devices, and very large-scale integrated circuits may permit the development of spaceborne instruments that can process the data and transmit to earth only the information needed by specific users. One can conceive of a spacecraft that could locate a forest fire in its early stages, sound an alarm, and transmit an image of the fire and the surrounding terrain, including the roads to be used by fire-fighting teams.

MANAGING NEW SPACE ENTERPRISES

The technological directions for the next five years appear to be well defined, although much research and development still needs to be done. Transferring the results of this rapidly advancing technology into actual practice poses some troublesome issues. Although there are basic questions about costs and benefits and, hence, differing views on the desirability of moving to an operational system, one major problem is institutional. Space missions are costly, and effective institutional arrangements are essential to ensure that their output is used and that the benefits are commensurate with costs. The operational weather-satellite system was introduced with relative ease because of the existence of a working domestic and international infrastructure—the government weather services and WMO—that could fund and coordinate the operation of the system. Also, the diverse users of weather data were already aware of the value of such data and thus receptive to data from new sources.

Such a well-developed infrastructure does not exist for earth resources or other environmental services and must be established if the transition from research and development to an operational system is to be made. An important factor delaying this transition has been the mixed population that could benefit from satellite information. This community includes scientists of many disciplines, farming groups, government agencies, fuel and mineral industries, land-use planners, and water-resource managers in this and other countries. There is a wide disparity in the needs, technical sophistication, and financial resources of these different users. In addition, the paucity of institutional arrangements that permit them to agree on information needs makes it difficult to resolve questions about what should constitute an operational system, who should operate it, and how the costs should be shared.

INTERNATIONAL ISSUES

Other nations have strong interests in many space applications. Several successful models for international participation in space activities and for making the benefits accessible to other nations have been developed. In meteorology and communications, two quite different but successful models of institutional frameworks have been devised. In satellite communications, Intelsat Corp. is the model of an independent international corporation responsible for developing and managing a single global satellite system. The meteorological model is different. A coordinated, internationally agreed-upon regional system of geosynchronous satellites has evolved, with individual or regional groupings of countries committing themselves to the development, management, and operation of individual weather satellites as part of an international system.

Applications of other satellite data are more difficult. Developing countries have meager data on which to base their resource and land-use planning and may find Landsat data particularly useful.[23] However, they have expressed deep concern about how such data will be used. In some of these cases, effective use of the data is limited by domestic political considerations and shortages of qualified personnel. The problem is further complicated because, unlike the weather and communications cases, some countries are concerned about the dissemina...n of Landsat satellite data to third parties. Such countries can obtain processed data from the United States or directly through their own or a neighboring country's Landsat ground stations. Canada, Italy, and Brazil operate substations. Memoranda of understanding for establishing stations also have been signed with Iran, Argentina, India, Chile, and Zaire and are being negotiated with several other countries.

The way in which the United States should proceed over the long stretch is not clear. Involving other nations in planning and structuring a global resources information system or any other earth-sensing satellite system is a sensitive and difficult problem that will need to be overcome. Remote sensing can generate questions of national security and national sovereignty. These issues have been and are continuing to be discussed in the Outer Space Committee of the United Nations.

AFTERTHOUGHTS

It is tempting to bewail the complexities of modern society. In earlier days, we located mineral resources by walking the land, we collected oil from natural seeps, and we drew water from abundant sparkling streams or shallow wells. Weather and natural hazards were in God's hands and their ill effects were to be borne with grace. Population was kept at modest levels by plague, pestilence, famine, and war. Society is no longer so simple nor so limited in size. Our real or perceived need for resources and our concern about the global environment call for more imaginative ways of finding resources and for more sophisticated knowledge of the earth.

Both future and past discoveries depend on the imaginative weaving together of ideas and tools; of concepts such as plate tectonics, with its unique view of how the earth's crust moves; and new technologies that enable us to drill in the deep ocean, see the earth from the vantage of space, and detect and measure the faintest tremors and movements in the earth's crust. We have achieved better understanding of the nature of climatic change by acquiring observations from space and from deep in the earth, and by using theoretical models of the climate with ever more powerful computers to simulate climatic conditions.

Moreover, the classical boundaries of the earth sciences—geology, meteorology, oceanography, and so on—are being eroded and replaced by a planetary multidisciplinary view. For example, portraying and understanding the long-term evolution of climate depends on understanding the movement of crustal plates and the interpretation of deep-sea cores and sediment samples. Understanding and predicting the shorter period changes in climate depends on knowledge of the oceans, their temperature, currents, ability to act as a reservoir, and their role in the global energy cycle.

A central theme is that the new knowledge gained by the vigor of earth sciences and by pertinent technology is now vital to the wise management of our planet. Plate tectonics is essential to the effort to understand and predict earthquakes and to improved reconnaissance for new mineral deposits. Atmospheric chemistry enables us to make a reasoned estimate of the likely future effects of trace amounts of chlorofluoromethanes on trace amounts of ozone in the stratosphere. Basic work in marine biology and ecology is indispensable to structuring effective policies for managing the living resources of the seas. Research on the chemistry of ocean water will enable us to fix more precisely the role of the oceans as a reservoir for CO_2, helping to yield, in time, precise estimates of the climatic effects of CO_2 and a more rational base on which to plan the future use of fossil fuels.

Our appraisal of recent trends in the earth sciences is dominated by the role of technology and the approach to planetary problems through organized and collaborative efforts of institutions and scientists—big science. There is a current question about big science and its relation to the science of individual investigators. It should be noted that the big science efforts described here grew from little science—the ideas of individuals—and provide to individual scientists data that could be obtained in no other way.

While it is true that further advancements will continue to be dependent upon new technology and massive efforts to acquire observations of natural phenomena, we must recognize that we would accomplish little without the creativity of individual scientists who integrate and synthesize information about these phenomena and who develop the concepts that ultimately spur advance in any field.

New concepts sometimes require many years before they are verified and accepted. Examples in the earth sciences serve as reminders of the need to support the inquisitive leanings of individual scientists. The theory of continental drift and, indirectly, plate tectonics, was first proposed by a German meteorologist, Alfred Wegener, in 1922. Its proof waited for half a century to pass and for the technology of modern geophysical research and deep-ocean drilling. The theory that long-term climate changes resulted from the earth's orbital variations was first advanced by the Serbian astronomer, Miltvin Milankovitch, in 1930. Again, it required close to half a century and the advent of the technology of ocean drilling and paleoclimatic dating to document this as a real possibility. Finally, the massive Global Weather Experiment is the culmination of proposals made by Jule Charney and a few others in the early 1960's.

Our experiences lead us to the view that neither must the support of individual scientists be neglected nor the claims of big science. To date, neither have been. We have diversity. It is necessary that this continue.

OUTLOOK

The following outlook section on the planet earth is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

Humanity is dependent on the resources and environment of the earth for its sustenance. Their wise use depends on the understanding of the planet that the earth sciences provide. The achievements of these sciences in the recent past have had a broad influence on our daily lives. Only a few aspects of the earth sciences are treated in the chapter—the structure of the earth, earthquakes, minerals, climate, the oceans, and the impact of space technology. The earth sciences also encompass vital activities such as water resources and hydrology, weather forecasting, the geology of the continents, the high atmosphere and near space, and other subjects not treated in this volume. They will be examined in subsequent reports.

WHERE TECHNOLOGY IS LEADING

Progress in the earth sciences depends upon the acquisition and interpretation of observations of the earth, its atmosphere, and its oceans. Observational information becomes available only as small samples of natural phenomena, some of which vary rapidly with time and some of which may have global dimensions. The plates that make up the ocean floor may cover half an ocean, planetary wind systems girdle the globe, and oceanic currents extend over thousands of kilometers. While laboratory experiments are important for the earth sciences, data from the natural world are essential. Much of the modern technology that can acquire these observations requires long lead times for development. For this reason, the technologies that will influence the course of earth-science research during the next five years are reasonably well known.

Because the earth itself is the laboratory, acquiring data is difficult and costly. Much of the technology used by marine geologists and geophysicists was developed by industry or government for other purposes. For example, the search for oil and national interest in the nuclear test ban stimulated the development of the technology for geophysical surveying of the earth's interior with multichannel seismic arrays. The direct recovery of rock cores from the seabed by deep-sea drilling draws much from the industrial technology of offshore drilling for oil.

Space technology provided the first truly global platforms for observing the earth. The practical achievements and scientific data about the earth obtained from them during the past two decades have influenced many fields, such as weather forecasting and environmental and resource monitoring, and have added greatly to our knowledge of the earth. Building on these achievements, spacecraft being planned for the next decade should expand the diversity of observations available for earth applications and for the earth sciences.

New generations of satellites will yield new data for more precise mapping of the earth, improved assessment of the state of crops and forests, more precise measurements of sea level and sea state, and measures of the imperceptible motions of the solid earth. New satellites will enable us to make global observations historically unobtainable in any other way such as those important to an understanding of climate—trace gases and the earth's radiation budget.

Remote sensing technology is developing rapidly both from space satellites and from earth and airborne platforms. It will yield an observational capability that will enable us to probe phenomena deeper in the earth, higher in the atmosphere, and at greater distances at sea. Multichannel seismic reflection systems will enable us to penetrate deeply into the earth's interior, acoustic ranging offers hope of oceanic measurements at depth, and remote electromagnetic probing with radars, lasers, and other devices will enable us to sound the atmosphere and measure the sea state at great distances.

We will also probe the earth directly with more powerful tools. We will have the capability to penetrate the deep ocean sediments to great depths in any part of the oceans, sample the lower atmosphere with specially designed aircraft and reach the high atmosphere with giant balloons that stay aloft for a long time.

OUTLOOK FOR SCIENCE

Trends in science may be molded by events and socioeconomic needs, often expressed in new legislation, or by pressures due to international concerns; examples include the intense interest in climatic change stimulated by the passage of the U.S. Climate Program Act in 1978 and the search for alternative energy options stimulated by the 1973 oil embargo and subsequent price rises. Such trends are usually clear. What we cannot foresee are the directions that will be taken as a result of discoveries yet to be made.

In marine geology and geophysics, one major emphasis will be to explore continental margins. Both the drive for basic understanding of plate tectonics as well as the societal need to know more about the mineral potential of seabeds will channel these sciences.

More intense studies of earthquakes, their possible prediction, and their socioeconomic impacts will be undertaken. A much stepped-up effort will be made to detect and understand earthquake precursors using the new seismic networks and geodetic measurements along faults. These studies will be parallelled by new investigations of ways to reduce the hazards from earthquakes.

Two major national oceanic research efforts of the past decade are now drawing to a close: the International Decade of Ocean Exploration (IDOE) and the Deep Sea Drilling Program. Future trends in ocean science will depend largely on the allocation of resources among competing needs of the science. Now competing for investment in ocean research are the re-equipment and replacement of ships that belong to an aging research fleet, the continuation of IDOE investigations that have examined such major problems as the major ocean current systems, study of the interaction between ocean and atmosphere, and the processes of upwelling of ocean waters, and the continuation of a program of geological and geophysical research based on the use of newer geophysical techniques and more effective drilling capabilities. Marine biological and ecological research will make an important claim to this investment based on the needs of improved fisheries and coastal zone management.

The U.S. Climate Program Act of 1978[22] has set the course of research on climate. The principal focus of this effort will be on developing more precise estimates of the consequences of human interference with climate, with special emphasis on carbon dioxide. Considerable effort will be directed at improving our ability to make climatic predictions for periods of months and seasons. Two areas of climate research that need special attention concern ocean/climate relationships and biogeochemical cycles.

SOCIAL AND ECONOMIC ISSUES

Our new knowledge of the earth frames an interesting set of social and economic issues. What is evident in these issues is the interweaving of science with wider concerns: of climatic change with future use of energy and evolving patterns of agriculture; of new space-borne sensors with more sensible land use; of analyses of tectonic processes with better protection against earthquakes and the discovery of new mineral resources. Thus, while the viewpoint in this chapter is that of scientists, the implications are universal.

RESPONSE TO NATURAL HAZARDS

Social response to natural hazards and warnings of them will continue to pose serious problems. As we learn more about the nature of earthquakes, droughts, and other hazards, we must also learn more about possible responses, for example, what to do when an earthquake is predicted. Policies aimed at mitigating social and economic effects of hazards will demand a base of sound scientific data.

OCEAN MANAGEMENT

Managing ocean and coastal resources will be a continuing and increasingly difficult issue. We will have to draw from our bank of scientific knowledge about the oceans as we turn to our continental shelves for oil and gas, increasingly assume for ourselves the right to coastal fishery resources, consider the oceans for waste disposal, and seek to protect the habitats of marine species. Over the next five years, much of the legislation enacted during the past decade will be reviewed and modified. Science can, as a minimum, assist in outlining the consequences of various ocean-management policy options.

ENERGY AND MINERAL POLICY

The questions of energy and mineral policy should attract major interest during the next five years. Our knowledge of earth sciences will contribute importantly to framing these policies. What is the potential for new oil and gas resources in the deep ocean? What are the possibilities for radioactive waste disposal at sea? What are the climatic implications of scenarios of future fossil-fuel use?

The United States, whose economy depends on ready access to nonfuel minerals, has an inadequate domestic supply of certain strategic minerals. There exists no integrated U.S. program of basic research for understanding ore-forming processes. If we are to stimulate mineral research, exploration, and development in the United States, we will also need to examine the legal and regulatory structures that govern mining activities. The policy questions are numerous and their resolution will depend on the scientific and technical information that can be provided.

AGRICULTURAL POLICY

No area of national policy is so sensitive to climatic variations as food policy. Worldwide growing conditions determine the nature of the global food market. Agricultural export policy is involved as are decisions on acreage restrictions. International disaster relief and our participation in a global food reserve system are all dependent on climate. Improved use of climatic data in such policy decisions will become increasingly important.

ENVIRONMENTAL POLICY

We will need to base environmental policies on what is known about the impact of human activity upon the environment and vice versa. Basic to such policies is the fundamental knowledge of the capacity of the atmosphere, the hydrosphere, the lithosphere, and the biosphere to disperse or concentrate pollutants.

USING HIGH TECHNOLOGY

The pursuit of information to further knowledge drives the development and use of high technology in the sciences. But different considerations apply when that same high technology finds general uses. Illustrative is the current problem in transferring the knowledge and technology of the earth-orbiting Landsats from research and development to operational use. Introduction of an operational Landsat satellite is a complex matter that depends upon costs and benefits and how those will be shared among users. It raises the question of how institutional mechanisms can effectively relate the interests of the federal government, private concerns, and state and local governments. How this will be done will depend upon the utility of the data acquired from space, the distribution of the users, their willingness to pay, and the needs and responsibilities of the federal government. The Landsat case is likely to be the first of several, and how it is resolved will affect how we approach analogous problems arising in the use of ocean or geodetic satellites.

INTERNATIONAL AFFAIRS

Developments in the earth sciences have special implications for international affairs. Much research in the earth sciences requires observations from large areas of the globe and, in some cases, from the entire planet. The acquisition of some kinds of data has been made easier by remote-sensing devices such as earth-orbiting satellites. However, satellites cannot provide all data needed for earth-science investigations. Experience has shown that satellite data need to be supplemented by ground data because of ambiguous "signatures," i.e., several different objects giving the same remote-sensing results. The study of earthquakes requires a worldwide seismic network; that of climate, a network of global observation stations; and that of oceans, the facilities of many nations. For these reasons, international cooperation in the earth sciences is already a long-standing tradition.

The mechanisms for conducting international earth science activities are diverse, and include bilateral arrangements between governments, informal arrangements between institutions, and arrangements within specialized agencies of the United Nations and within scientific organizations outside the U.N. system. Present trends in the earth sciences indicate that such cooperative activities will not only need to be continued but intensified over the years ahead. The resources required to study many of the problems are too great for any single nation to provide. Furthermore, in many cases, access to territories under national jurisdiction for scientific observation is indispensable. Of special importance during the next five years will be the issue of scientific research at sea. Developments in the U.N. Law of the Sea Conference foreshadow serious restrictions of freedom for oceanic research.

The earth sciences are unique in their planetary nature. Information about the earth and its global environment is essential to all countries of the world if they are to meet their own social and economic objectives. For this reason, the earth sciences offer a channel for cooperation among many nations of the world, both developed and developing. Because of the increasing importance of scientific information to the development objectives of other countries, the outlook is for greater interest and, in many countries, for closer collaboration in earth-sciences research. How this is done and whether it is desirable represents an area of policy and programmatic decision that will need considerable attention during the coming years.

REFERENCES

1. *Earthquake Prediction and Public Policy* (NRC Advisory Committee on Emergency Planning). Washington, D.C.: National Academy of Sciences, 1975, p. 38.

2. *A Program of Studies on the Socioeconomic Effects of Earthquake Predictions* (NRC Committee on the Socioeconomic Effects of Earthquake Prediction). Washington, D.C.: National Academy of Sciences, 1978.

3. *Review of National Mineral Resource Issues and Problems* (NRC Board on Mineral and Energy Resources). Washington, D.C.: National Academy of Sciences, 1978.

4. Rose, A.W., *et al. Research Frontiers in Exploration for Nonrenewable Resources.* College Park, Pa.: Pennsylvania State University, 1977.

5. Ocean Policy Committee. The Marine Scientific Research Issue in the Law of the Sea Negotiations. *Science* 197(4300):230–233, 1977.

6. Study on Man's Impact on Climate. *Inadvertent Climate Modification.* Cambridge, Mass: MIT Press, 1971.

7. *Energy and Climate.* (NRC Geophysics Study Committee). Washington, D.C.: National Academy of Sciences, 1977.

8. Johnston, H. Reduction of Stratospheric Ozone by Nitrogen Oxide Catalysts from Supersonic Transport Exhaust. *Science* 173(3996):517–522, 1971.

9. Crutzen, P.J. The Influences of Nitrogen Oxides on the Atmospheric Ozone Content. *Quarterly Journal of the Royal Meteorological Society* 96:320–325, 1971.

10. *Final Report of the Climatic Impact Assessment Program.* Washington, D.C.: U.S. Department of Transportation, 1974.

11. *Environmental Impact of Stratospheric Flight: Biological and Climatic Effects of Aircraft Emissions in the Stratopshere* (NRC Climatic Impact Committee). Washington, D.C.: National Academy of Sciences, 1975.

12. *Effects of Human Activities on Global Climate.* World Meteorological Organization, Technical Note 156, WMO No. 486, 1977.

13. Whillans, I.M. Inland Ice Sheet Thinning Due to Holocene Warmth. *Science* 201(4360):1014–1016, 1978.

14. White, R.M. Organizing a World Climate Program. *Bulletin of the American Meteorological Society* 59(7):817–821, 1978.

15. *Toward a U.S. Climate Program* (NRC Climate Research Board). Washington, D.C.: National Academy of Sciences, 1979, Appendix B.

16. Lintz, J., and D.S. Sinnot. *Remote Sensing of Environment.* Reading, Mass.: Addison-Wesley, 1976, p. 326.

17. Bishop, B.C. Landsat Looks at Hometown Earth. *National Geographic* 150:140–147, 1976.

18. Barrett, E.C., and L.F. Curtis. *Introduction to Environmental Remote Sensing.* London: Chapman and Hall, 1976.

19. *Practical Applications of Space Systems: Inland Water Resources Supporting Paper No. 5* (NRC Space Applications Board). Washington, D.C.: National Academy of Sciences, 1975.

20. *Weather and Climate* (NRC Space Applications Board). Washington, D.C.: National Academy of Sciences, 1975.

21. *Space Shuttle* (National Aeronautics and Space Administration Information Office). Washington, D.C.: National Aeronautics and Space Administration, 1976.

22. National Climate Program Act. Report No. 95-1489, House of Representatives. August 14, 1978.

23. *Resource Sensing from Space: Prospects for Developing Countries* (Ad Hoc Committee on Remote Sensing for Development). Washington, D.C.: National Academy of Sciences, 1977.

BIBLIOGRAPHY

Brobst, D.A., and W.P. Pratt. *United States Mineral Resources.* U.S. Geological Survey Professional Paper 820. Washington, D.C.: U.S. Government Printing Office, 1973.

Cox, Allan (ed.). *Plate Tectonics and Geomagnetic Reversals.* San Francisco: W.H. Freeman, 1973.

Tilton, J.E. *The Future of Nonfuel Minerals.* Washington, D.C.: The Brookings Institution, 1977.

Uyeda, S. *The New View of the Earth,* O. Masako trans. San Francisco: W.H. Freeman, 1978.

Van Rensburg, W.C.J., and D.A. Pretorius. *South Africa's Strategic Minerals.* Johannesburg: Valiant Publishers, 1977.

Wylie, P. *The Way the Earth Works.* New York: John Wiley and Sons, 1978.

Government and the Nation's Resources (National Commission on Supplies and Shortages). Washington, D.C.: U.S. Government Printing Office, 1976.

Mineral Resource Perspectives. U.S. Geological Survey Professional Paper 940. Washington, D.C.: U.S. Government Printing Office, 1975.

Problems of U.S. Uranium Resources and Supply to the Year 2010 Supporting Paper 1, (NRC Committee on Nuclear and Alternative Energy Systems). Washington, D.C.: National Academy of Sciences, 1978.

Special Report: Critical Imported Materials (Council on International Economic Policy). Washington, D.C.: U.S. Government Printing Office, 1974.

Technological Innovation and Forces for Change in the Mineral Industry (NRC Committee on Mineral Technology). Washington, D.C.: National Academy of Sciences, 1978.

2 The Living State

INTRODUCTION

In recent decades biological science has grown explosively, not only in scale but in depth. Initially a descriptive field, concerned with recognizing the structures and the behavior of whole organisms and their organs, biology has become increasingly analytical as it has probed ever finer levels of organization. We now know a great deal about the mechanisms responsible for the formation, function, and regulation of cells and their components.

These advances have depended in part on the development of elaborate instruments: electron microscopes, which extend a thousandfold the dimensions that can be visualized; the scanning electron microscope, which provides three-dimensional images; ultracentrifuges, which separate cellular particles and molecules for many sciences; X-ray crystallography, nuclear magnetic resonance, and other physical probes into the intimate three-dimensional structure of molecules; a great variety of radioactively labeled compounds whose fate in living organisms can readily be followed; and electronic equipment for recording the activity of nerve cells. In addition, hundreds of biochemical compounds, which investigators used to have to make themselves at the cost of much valuable time, are now available from an industry that sprang up as research expanded. Other industrial advances with polymers made possible simple and ingenious chromatographic methods, in which a single passage through a gel can separate a complex mixture of hundreds of substances, in any size range, into its components. Finally, antibiotics have made it easy to culture animal or plant cells free of contamination by much faster-growing bacteria; genetic and regulatory properties of these cells, and of infecting viruses, can then be studied in ways that would be much slower, more expensive, and often impossible in the whole animal.

With these developments, the barriers between areas within biology have broken down, and fruitful marriages have taken place. Thus the field of cell biology has emerged with the advent of techniques for exploring a range of dimensions that used to be too large for the biochemist and too small for the investigator of cell structure. Similarly, at the molecular, mechanistic level, genes are linked to origins in the past and to functions in the present; the fusion of genetics with biochemistry has brought together those biologists concerned with how organisms arose in evolution and in embryonic development and those concerned with how organisms work.

Because research in biology spans a great range of biological systems and experimental approaches, we shall review in this chapter only a few selected areas, mostly at the molecular level. We shall also add a few comments on evolutionary studies, those concerned with whole organisms and populations. Other large areas, such as population biology and ecology, which are of great importance for agriculture and for conservation of resources, will not be covered.

MOLECULAR STUDIES

The groundwork for the growth of the molecular studies that now pervade most of biology was laid in the biochemistry of several past decades. Research in this field identified the chemical components of living cells, worked out the sequences of reactions that synthesize these components from foodstuffs and that provide the energy for these syntheses, and disclosed a good deal about the enzymes that catalyze the individual reactions. By now we know the complete pathways, involving several hundred different intermediates, for constructing all the known building blocks that are common to all living organisms. Such biomedical studies continue to be extended fruitfully to other small molecules with specialized roles in various organisms, such as hormones, as will be illustrated particularly in the sections on neurobiology, and biology and agriculture

The search for a coherent conceptual framework in biology took an enormous step forward with the emergence of the field of molecular genetics. With the discovery in 1944 that the genetic material is DNA (deoxyribonucleic acid), and the finding in 1953 that this long-chain molecule is composed of two precisely complementary strands, so that the sequence of bases of either specifies that of the other, the key to gene duplication was evident. In the subsequent quarter-century, intense investigation has revealed a great deal more about how genes work: i.e., how they serve as blueprints for fashioning the cellular machinery, and how they are regulated, mutate, repair errors, recombine in sexual reproduction to yield endless diversity, and expand in number as evolution creates increasingly complex organisms. Furthermore, we can now describe the metabolic activities of a cell in a thoroughly logical, comprehensive way, proceeding from the genes to the enzymes that they code for, to the small molecules made by some of those enzymes, to the linking of building blocks into long chains by other enzymes, and to the regulatory feedback effects of various substances on the activity of specific genes and enzymes. Genetics, like biochemistry, now pervades all of biology. This point will be particularly illustrated in the sections on cell biology and immunology.

In addition to this wealth of insights into subtle mechanisms, molecular genetics has yielded a profound, germinal concept: that of molecular information transfer. Previously, biochemists viewed compounds only in terms of structure, chemical reactions, and energy relations. The most important feature of DNA, however, is that the sequence of units in its chain is a store of information, quite analogous to that in a magnetic tape or a line of type. This information specifies the chemical structures that a cell can build, and it also specifies the program of successive reactions that, over time, channel the development of a fertilized egg into a higher organism. Eventually it was realized that all specific interactions between molecules in biological systems involve information: for example, the recognition of a specific receptor on a cell by other cells, or by hormones, or by antibodies; or the feedback response of a gene or an enzyme to the concentration of a specific substance in its environment, as a means of regulating the amount of that substance to be synthesized.

In higher organisms the reception, storage, and transfer of information in the nervous system also involves molecular changes. Accordingly, we now see a continuity between two kinds of information in biological systems: the many bits inherited in the genes, and the many others acquired from the environment. These two stores of information interact to make organisms what they are.

Along with deepening understanding of how the information in genes is translated into protein sequences has come equally satisfying understanding of how the three-dimensional structures of proteins endow them with highly specific surfaces, which allow the proteins to function as the working machinery of the cell—catalytic, regulatory, and structural. Moreover, we now know a good deal about how enzymes fit other protein molecules and induce specific chemical changes in them; and how the activities of specific enzymes are regulated by interactions with specific surrounding molecules, which convey information about the biochemical needs of the organism at each moment. We also know, in principle, that the specific affinities of proteins on cell surfaces guide the cells to find their place during the growth of an organized tissue. Detailed understanding of these structures is one of the major challenges of biology.

With the recognition of all these mechanisms, it is not unreasonable to say that the secret of life has been discovered—or better, that many secrets have been discovered. Each involves known physical and chemical forces. These results make it very unlikely that any vital forces remain to be discovered in the major mechanisms still buried in neurobiology or developmental biology. However, this triumph of the mechanistic approach to biology in no way diminishes the unique qualities and the marvel of the living world and the human spirit. Rather, we can only stand in awe that life, evolving from inorganic matter and working with inorganic forces, could develop these qualities—and could develop an organism with the capacity to understand its own origins in remarkable detail. Although there is no unique vital force in living organisms, there is a unique molecular basis for their organization which is not found in the inorganic world: the molecular storage of information programming the development and function of the organism.

CHANGING ORGANIZATION OF BIOMEDICAL SCIENCES

The large recent changes in the level and the focus of biological research have affected its organization in various ways. One is the shift, in many branches of biology, from simple to complex instruments. It may cost in excess of $100,000 to equip a laboratory for a new independent investigator, and inevitably the costs will continue to mount.

Another change has been a fading of the boundaries between many branches of biology. As a result, in many universities resources and teaching responsibilities are shifting from medical schools, where many new lines of basic research originated, to basic science departments in faculties of arts and sciences. Although this trend is logical, there would be serious losses if medical schools were divested of their interest in fundamental science, and reverted to being trade schools limited to applied research.

A third trend is worrisome. Although the current generation of medical students is intrinsically better qualified than were their predecessors, fewer are pursuing careers in medical research. The percentage of M.D.'s among those applying to the National Institutes of Health (NIH) for research support declined from 41 percent in 1966 to 28 percent in 1976, while the percentage of Ph.D.'s was rising from 48 percent to 63 percent. If this trend continues, it could undermine the transfer of basic research findings to clinical practice, while research itself might be denied the special insights arising from human pathology.

PROGRAMMING OF BIOMEDICAL RESEARCH

Since the great medical advances of the past resulted in the prevention or the cure of a remarkable number of infections and nutritional or endocrine disorders, it has been only natural to expect similar rapid applications from the intellectual achievements of the present. The expectations include cures for cancer, vascular diseases and arthritis, and healing of the mentally disturbed or retarded. But the realization of these expectations may be far off, because these major medical problems today arise to a considerable extent from the inherent weaknesses, imperfections, and aging of human organisms; and these processes are not so readily changed. On the other hand, one may be more sanguine concerning the prospects for preventive measures to reduce the incidence, for example, of cancer and atherosclerosis.

Impatience and unfulfilled expectations, as well as exaggerated promises from some scientists, have led to disappointments and to complaints that scientists are more concerned with satisfying their own curiosity than with investigating matters of greatest significance to their public sponsors. But until the requisite understanding is at hand, large-scale targeting of research is all too likely to engender expensive but unproductive attempts to apply the inapplicable.

We must therefore try to advance our understanding of cell and organ function on all possible fronts, using as study objects those natural biological systems that offer special research advantages. Our interest in the life and times of *Escherichia coli*, the neurobiology of the squid, the aging of the rotifer, or the alarm reaction of the clam derives largely from the fact that each serves as an easily studied model for some process highly relevant to man.

Moreover, we must recognize why support of apparently esoteric research is especially important in biology. Breakthroughs have frequently depended on fortunate accidents and unexpected observations that revealed isolated components in the incredibly complex network of events in the living process. A battalion of eager scientists is waiting to transfer the results to medicine.

However well motivated the pressures to channel research to societally desirable targets, if the time is not right, such channeling is more likely to retard than to advance the desired outcome.

MOLECULAR GENETICS

Molecular genetics stands at the threshold of major socially important advances. The basic mechanisms elucidated in bacteria and viruses are rapidly being extended to higher organisms, where novel features are also being discovered. Since several of the main health problems in the United States today are diseases of cellular malfunction, and since these malfunctions often involve defects in gene function, fruition of research in this area can reasonably be expected. In addition, modern genetics promises new capabilities for agriculture and pharmaceutical and other industries.

Skeleton of a single-stranded SV 40 DNA molecule computer-generated from an electron microscope image. (National Institutes of Health: B. A. Shapiro, L. E. Lipkin, P. F. Lemkin, E. M. Smith, M. Schultz, National Cancer Institute; J. Maizel and M. Sullivan, National Institute of Child Health and Human Development; N. Salzman and M. Thoren, National Institute of Allergy and Infectious Diseases)

TRANSLATION OF GENE INFORMATION

The previously formal concept of the gene was given physical reality in 1944, when Avery and his colleagues showed that genes are located in long-chain molecules called DNA. Each chain is composed of only four different kinds of molecules called bases, which are denoted by the letters A, G, T, and C. The sequences of these units store the information that makes up our inheritance. The mechanism became clear when Watson and Crick discovered that DNA is composed of two complementary strands, i.e., A in either strand is always matched with T in the paired strand, and G is always matched with C. Thus, when DNA duplicates, either strand can automatically specify the sequence of the other.

A chromosome consists of an extremely long chain of DNA in which specific sequences function as genes. Most of the information in DNA is translated into the sequence of another class of long-chain molecules—proteins—which constitute most of the working machinery of the cell. Proteins are composed of 20 different kinds of amino acids. The formation of proteins begins with the process called transcription, in which one strand of a gene directs the synthesis of a complementary sequence in a similar long-chain (but single-stranded) molecule called RNA (ribonucleic acid). The sequence of bases in RNA is ordered by essentially the same base pairing as in DNA. The RNA serves as a messenger between DNA and protein; in the next step its sequence is translated, inside a complex subcellular particle called a ribosome, into a corresponding sequence of the amino acids of a protein chain.

Ten years ago, in a major triumph, the genetic code that translates the language of DNA and RNA into the language of proteins was deciphered. In messenger RNA, successive three-letter words—made up of different combinations of three of the four base letters—specify which amino acid is to be linked next onto the growing protein chain. All of the three-letter words possible with a four-letter alphabet are actually utilized. These words (including those used as the signals for start and stop) have all been identified. Moreover, the code, and the basic machinery of protein synthesis, is universal. All proteins are made in the same way in all living cells.

Gene Regulation

DNA not only codes for the structure of messenger RNA and thereby of proteins; it also contains specific regulatory sequences, which determine how much of each messenger is made. By turning specific genes on or off, these regulatory mechanisms adjust the composition of cells, and their output, to their circumstances. Even the most primitive organisms, the bacteria, have evolved elaborate regulatory mechanisms that promote the efficiency and speed of their growth by preventing the synthesis of components, such as amino acids, in excess of their requirements.

For example, *E. coli* makes the amino acid tryptophan only when it is lacking in the culture medium and its intracellular concentration is very low. This is made possible by the fact that, in the *E. coli* DNA, there are consecutive genes that code for five different enzymes needed during the synthesis of tryptophan. A separate protein called the tryptophan repressor (coded for elsewhere on the DNA) can bind to a region of DNA (called the operator) immediately preceding the five genes; when so bound, the repressor prevents transcription of all five structural genes. But to attach to the operator region of DNA, the repressor must have tryptophan bound to it; such binding occurs readily when the concentration of tryptophan is about that normal for *E. coli* but not at lesser concentrations. Thus the concentration of tryptophan governs tryptophan synthesis, mediated by a regulatory protein which senses the concentration of end product and transmits this information from the surroundings to the genes. The molecular basis of these interactions and the mechanisms of several other, more intricate regulatory mechanisms in bacteria are now understood in considerable detail.

APPROACHES FOR STUDYING HIGHER CELLS: RECOMBINANT DNA

Much less is known about regulatory mechanisms in animal and plant cells than in bacteria. We do know, however, that all the different kinds of cells in a higher organism contain essentially the same set of genes, of which only a fraction is functional in a given cell type. Accordingly, the regulatory mechanisms discovered in bacteria provide a simple analogy to the mechanism, for example, that restricts hemoglobin synthesis in higher organisms to cells destined to become red cells. However, there is a major difference. In bacteria, regulatory changes are instantly reversible, while in the differentiated cells of higher organisms they are stable and self-perpetuating, through yet unknown mechanisms.

The striking success of studies of regulation in bacteria has depended on the ability to perform genetic and molecular studies on specific regions of DNA in these simple cells, which have about 5,000 genes each and can double every 20 minutes. In higher systems, with 1,000 times as much DNA per cell and with generation times of years, these methods are ineffective. It is therefore extremely encouraging that in the past few years the experimental barriers have fallen, owing largely to the development of what has come to be known as recombinant DNA methodology.

In this approach, the DNA from any source can be cut into small pieces by a family of so-called restriction enzymes, each of which recognizes and excises a specific short sequence in the DNA chain. The fragments can then be spliced into bacterial DNA, where their behavior can be studied as precisely as that of bacterial genes. In addition, the inserted DNA, grown in unlimited amounts in the bacteria, can easily be recovered and purified. Its base sequence can be determined quite simply by recently developed methods. With viruses, which contain much shorter molecules of DNA than do cells, the restriction enzymes can be applied directly to that DNA, and the total base sequence easily determined from the fragments. Finally, viruses and DNA fragments can be subjected to agents that produce persistent changes, or mutations, in known loci rather than in random loci, as in cells.

At one time, scientists considered it prudent and desirable to avert any potential danger that might be associated with research on recombinant DNA. The NIH guidelines were a reasonable response to public anxiety, given the limited knowledge that was available at the time. Research during the intervening years has shown that the earlier concerns were without basis, and has essentially dissolved the apprehension that originally led scientists to bring the problem to public attention. As a result, the restrictions have been somewhat relaxed.

Worst-case analysis with a virus highly infectious to mice indicates that it is utterly devoid of infectious or other noxious properties when joined to other DNA inside *E. coli*. No precautions other than the conventional safety procedures of routine microbiology now seem necessary or appropriate to research with recombinant DNA.[1,2]

SEGMENTED GENES IN HIGHER ORGANISMS

The magnitude of the breakthrough provided by recombinant DNA is hard to overestimate. One example of its impact is the recent revolution in the concept of a gene.

Classical genetics defined a gene both as a unit of function (which determines a particular trait, such as the color of a flower) and as a unit of structure (located at a particular position in a chromosome). However, with the advance of molecular genetics, it became clear that each gene is not a molecule but rather a particular sequence of bases in the long DNA helix. It has no beginning or end, except as signals to start and stop transcription are encoded in its sequence. In bacteria the whole sequence is transcribed, one letter after another, into messenger RNA, which is similarly translated into the corresponding protein.

In higher cells, however, the very first applications of recombinant DNA methodology yielded a most unexpected result: The DNA sequence for a given protein (for example, hemoglobin) is not found as a single contiguous series of bases. Instead, several regions, each coding for part of the protein, are separated by long intervening sequences which are not translated. When a gene is expressed, the entire long DNA sequence is transcribed into RNA; then special enzymes cut out and splice together the several segments that together constitute the messenger. This RNA molecule is then translated continuously, as in bacteria. This discovery has led to optimistic speculation that the intervening sequences, and the joining mechanism, may play a central role in the complex gene regulation of higher organisms. Many tests of this hypothesis are in progress.

The fragmentation of genes goes even further in the synthesis of antibodies. A single antibody molecule is encoded in three libraries of DNA fragments, which are found in separate places on a chromosome. In making an antibody messenger, first one piece of DNA from each of the three libraries is physically translocated to form a single sequence, then that DNA is transcribed into RNA, and finally parts of that transcript are spliced together to make the messenger RNA.

Duplication of DNA

One of the major challenges of molecular genetics is to understand how DNA is precisely duplicated, with an accuracy that allows a sequence of a million bases to be transmitted intact from one generation of cells to the next. In this area, as in gene structure and regulation, a great deal is known about bacteria and their viruses and very little about higher cells, except that the latter have new levels of complexity. When Watson and Crick first grasped how DNA is organized, they were gratified to realize that the molecular mechanism of its duplication was inherent in its structure: Either strand can be a template for synthesizing the other. But many questions remain to be solved. How does the double helix unwind to open up the paired regions? How is synthesis regulated to assure one complete round of DNA synthesis each time a cell doubles? How is the very high fidelity of duplication maintained? At how many sites along the strand is DNA synthesis initiated?

Central elements of the answers to these and many other questions are lacking. We have learned, though, that DNA can take many physical forms and that it is duplicated in various ways. In some viruses, for instance, DNA is linear and in others it is circular, with neither beginning nor end. Some linear viral DNA have a unique sequence (e.g., A–Z), while others occur as a family of sequences derived by cutting a circle of DNA at different places (e.g., A–Z, D–C, Q–P). Both in bacteria and in higher cells not quite all DNA is in the chromosomes. There are independent smaller DNA pieces—called plasmids—that find expression and replicate autonomously like chromosomal or viral DNA but are not released like viruses. The ability of bacteria to form toxins or to resist various antibiotics is usually due to information in such DNA pieces, which can be transmitted from one organism to another by special mechanisms. Another complement of autonomous DNA is found in mitochondria, small intracellular structures that inhabit all cells, as well as in the chloroplasts of plant cells.

The synthesis of chromosomal DNA in higher cells differs in a crucial way from that in bacteria. Bacteria generally start duplication at one point on their circular chromosome, and a wave of synthesis passes from that site around the circle. In the much larger chromosomes of higher cells, duplication starts at many sites, and the resulting segments are then joined. The mechanisms coordinating this process are yet to be studied.

In summary, although the elementary act of DNA duplication is evident in the Watson-Crick model, many individual aspects vary with the specific DNA molecule and the cell type. And the picture is far from complete.

Chromosome Structure

Great progress has recently attended the study of chromosome structure in higher organisms. Small strands of the DNA double helix wind around balls of a protein called histone. These balls then clump together to form the compact chromosome.

This mode of packaging for DNA raises many questions. Does this arrangement break apart to allow the DNA to duplicate, or can the DNA replicate within this chromosomal organization? How does this organization change in different stages of the cycle of cell division? How can the DNA specify RNA sequences when it is so compactly packaged? Does this package mechanism help to regulate DNA expression, or is it simply a good way to pack a lot of DNA into a small nucleus? Investigation of this field has just begun.

GENETIC RECOMBINATION

Another goal of molecular genetics is to understand how the DNA molecule, in all forms of life, can recombine as it does during reproduction of sperm and egg cells, breaking at identical points in a pair of homologous chromosomes and then resealing with fragments exchanged. Genetic recombination is pivotal to evolution, for while mutations are the ultimate source of genetic variation, their accumulation within a succession of progeny would be a very slow source of the variation on which natural selection acts. Genetic recombination, by reshuffling the mutations accumulated in different members of a species, enormously increases the number of variations.

In higher cells, recombination seems quite complex, and even in bacteria, where recombination is essentially the same, the details are not clear. Because of the simplicity of viruses, their recombination mechanisms are better understood. Recently it has become possible, using viruses, to carry out recombination in solution, outside cells. These new systems promise to provide models through which recombination in cells may be understood.

Mobile DNA

One important new discovery is the presence in bacterial DNA of sequences that "hop" from one region to another rather frequently, through high-frequency recombinations at specific short sequences near their ends. How they do this is still debated, but the process clearly plays an important role in bacterial evolution and in certain regulatory processes. This finding should help us to understand similar obscure processes in the cells of plants and animals. For example, as mentioned earlier, the production of antibodies requires pieces of DNA to move within the chromosomes.

MUTATIONS AND RADIATION DAMAGE

One area of genetic research with great practical significance is the study of mutations. Errors are incorporated into DNA, either by mistakes made during copying or by chemical alterations in existing DNA. Mutations can have either positive or negative consequences for organisms. The negative aspect is obvious; most changes in DNA impair normal functioning by altering or inactivating a protein or by changing the control of its synthesis. But mutations are also the ground substance of evolution, for occasionally they improve the adaptation of an organism.

Spontaneous mutations cannot be totally eliminated, because occasional errors are inherent in the copying process, despite its remarkable accuracy. These errors can occur at random in any gene in an organism. Errors would be much more numerous if it were not for repair processes, which eliminate most, but not all, of the mismatches in base pairs. The mutation rate per unit length of DNA, which varies widely among species, depends on two mechanisms—the precision of the replicating enzyme and the effectiveness of various repair systems.

In addition to spontaneous mutations, others can be induced by agents that increase the mutation rate in various ways. These agents include ultraviolet and high-energy radiation. Many chemicals, both naturally occurring and man-made, are known to cause mutations in intact living cells; only a few have, to date, been seen to so affect DNA *in vitro*. With the very sensitive bacterial systems now available, it is possible to detect slight mutagenic effects of certain normal constituents of food, such as caffeine, or even of moderately elevated temperature.

Closely related to the study of mutations are the lethal effects, on cells and on organisms, of damage to DNA by radiation and by some chemicals. Many kinds of damage cause errors in replication (mutation); but others, which prevent further replication, are lethal to the cell. Cells have developed systems for repairing such damages just as for repairing mutations. The effects of radiation on organisms, long studied at a largely descriptive level, have now been translated into specific chemical terms in some detail.

While mutations in germ cells are passed on to offspring, it appears that some mutations in normal body cells (somatic mutations), passed on to daughter cells (but not to the offspring of the organism), can result in cancer. Knowledge of the mechanisms of mutation and repair is therefore fundamental to understanding how cells become cancer cells and how we may prevent this process.

Because of the apparently close links between mutation and cancer, inexpensive and rapid tests for mutations in bacteria are used to screen for potentially cancer-causing substances. The Ames test was the prototype of this approach. These tests, now widely used in industry, are a landmark in the application of molecular genetics to public health problems.

VIRUSES

Until the advent of molecular genetics, viruses could be studied only descriptively, as submicroscopic agents of disease. Virology has now become an intensely active, sophisticated field, as modern methods, gradually extended from bacterial viruses to those of animals and plants, have made it possible to analyze the properties of various viruses in great detail. Viruses are not cells smaller than bacteria; they are autonomously replicating blocks of DNA (or in some cases RNA) encased in a protein coat. The protein, by binding to a specific receptor in a host cell membrane, may facilitate its entry into that cell.

Like plasmids, viruses multiply within cells. The virus sheds its protein coat and its nucleic acid enters the cell, where it appropriates the cellular machinery it needs to duplicate itself and to synthesize its protein components. Then the nucleic acid and its coat are assembled and exit from the cell as complete virus particles.

Because of the extreme simplicity of viruses (some have only three genes), virology has contributed much to molecular genetics, as well as to the development of improved vaccines and the beginnings of effective chemotherapy of viral infections. The present stage of the field is one of dynamic progress.

Some viruses have chromosomes made of RNA instead of DNA. These viruses generally induce in the cell a novel system for copying (transcribing) RNA from RNA. But in one group called retroviruses the virus codes for an enzyme that makes a DNA copy of the viral RNA; this DNA copy can be spliced into the DNA of cellular chromosomes. These viruses can thus alternate between being independent entities and being cellular genes. This phenomenon is somehow crucial to the induction of tumors by viruses of this class.

PROTEIN SYNTHESIS

The ultimate role of DNA is to direct protein production. The conditions required for making protein outside cells, in extracts, were worked out about 20 years ago, and with this development the components of the system could be separated and identified. Progress was rapid. Proteins are now known to be synthesized in complex particles called ribosomes, with the help of several dozen enzymes as well as several dozen RNA molecules called transfer RNA, all of which are found in the cytoplasm. Each transfer RNA provides a link between a three-letter word in messenger RNA and the specific amino acid that it codes for.

The ribosome is an extraordinarily complex structure, consisting of 54 different proteins and 3 different RNA molecules, fitted together in a specific way. These components provide many different binding sites, and go through an orderly cycle of complexing with other components of the system in the course of adding each amino acid as the ribosome reads the RNA message. A major triumph has been the reassembly of active ribosomes in the test tube from their dissolved components.

Observing the functions blocked by various antibiotics that act on bacterial ribosomes (e.g., streptomycin), has made it possible to work out the mechanisms of action of these drugs, and to clarify aspects of ribosomal functions. Bacterial mutants resistant to streptomycin contain altered ribosomes, and by reassembling ribosomal components from sensitive and resistant cells, researchers have also identified the basis for resistance.

Protein Structure

The specific surface of each protein enables it to play its particular enzymatic or structural role by closely fitting to the surface of some other molecule. Accordingly, the amino acid chain translated from the one-dimensional information in DNA must fold into a three-dimensional structure with a highly specific surface. The mechanism has turned out to be both remarkably simple and elegantly subtle. Not only the information but the forces for the correct folding are built into the protein chain, through the specific affinities of the 20 different amino acids in that chain for each other. The resulting internal binding causes each protein chain to fold up on itself spontaneously in a unique way, without requiring an external template or enzyme. On the resulting surface, unique to each protein, will be found regions with affinities for various other types of chemical structure such as substrates (for enzymes) or antigens (for antibodies). Moreover, the spontaneous aggregation of specific structural proteins with each other and with other cellular components initiates a cascade of steps, in an increasingly complex assembly, that accounts for the formation of membranes, cells, and even organs and organisms.

GENETICS AND HUMAN DISEASE

Folklore has long suggested that some diseases are transmitted from generation to generation. Current understanding began in 1908 with the description by Garrod of half a dozen "inborn errors of metabolism." The subsequent development of biochemistry and genetics continued to expand that understanding; over 2,000 hereditary disorders are now known, and the list continues to grow. Most are seriously disabling, and many are fatal in early life.

To date, the specific defects in more than 150 specific hereditary metabolic abnormalities have been learned. Examples include cystic fibrosis, Tay-Sachs disease, sickle-cell anemia, various hemophilias, and phenylketonuria. In each a specific protein—an enzyme or other functional category—is synthesized either in a defective form or not at all. Such single-gene diseases are inherited in the manner predicted by classical Mendelian genetics. In sum, they constitute a massive burden of illness for which cures in the ordinary sense are generally unlikely.

At the same time, evidence has mounted that other disorders are genetic in a more complex sense, as in diabetes, for instance, where the defective biological process is affected by a multiplicity of genes. Thus, while juvenile diabetes is surely the expression of an individual's genetic constitution, the genetics are subtle and complex and differ in different individuals with much the same disorder. This may be the case for many endocrine disturbances.

More recently, the essentially genetic character of yet another broad set of diseases has begun to be understood. These include ankylosing spondylitis, rheumatic fever, rheumatoid arthritis, lupus erythematosus, thromboembolic purpura, and Reiter's disease. The individuals who acquire these diseases usually (with some diseases, always) bear certain genetic markers, a specific one of the dozens of antigens on the surfaces of red blood cells, of lymphocytes, or of tissue cells generally. The converse does not hold true: Only a fraction of those who bear a given tissue antigen ultimately develop the associated disease. The relationship between the specific antigen and the manifestations of the specific associated disease is unknown. In some cases, occurrence of the disease also requires previous exposure to some environmental agent, for example, streptococci for rheumatic fever, dysentery-causing organisms like Shigella for Reiter's disease. Yet it is clear that individuals so affected are genetically predisposed to these diseases.

Lymphocyte and tissue typing of humans has barely begun. It is clear that as an increasing variety of tissue antigens is typed in larger numbers of individuals, additional correlations of this character will surely be found. They perhaps will predict which of us are most likely to develop cancer when exposed to an environmental carcinogen, which are most likely to develop atherosclerosis, multiple sclerosis, and so on. In fine, from this standpoint, all disease has a genetic component. But that should not be surprising, since that is true of all significant traits. What is new and hopeful is that it may become possible to ascertain, in early life, the special vulnerabilities of any individual human and, hence, to minimize exposure to the particular environmental factors that might trigger disease.

FUTURE CHALLENGES

Molecular genetics, born 25 years ago, has made great progress in its short lifetime. But in contrast to earlier developments in biochemistry, which identified small molecules, such as vitamins and hormones, that could immediately be put to use in medical practice, discoveries in molecular genetics have not been so readily converted into applications. Rather, this field has dealt with the giant molecules that lie at the heart of the cell, and these were often best studied best in bacteria and viruses. However, molecular genetics is now analyzing problems of health and disease in human cells, and also the influence of small molecules such as antibiotics on DNA synthesis and function. As this understanding progresses, so will our understanding of such diseases as cancer, autoimmune disorders, thalassemia, and formation of abnormal hemoglobins. Other rapidly growing applications include the ability to determine the presence or absence of various fetal defects by amniocentesis (sampling of intrauterine fluid), and to detect mutagenic compounds.

Even more predictable than contributions to health will be the contributions to our understanding of our own physiology. We can confidently expect that in a few decades we will understand in some detail how the growth of a specific organ system and the harmonious growth of an entire organism are regulated. Moreover, though the particularly interesting behavioral and physical traits in human beings must involve huge numbers of genes, it may not be too long before we begin to unravel the genetic components of much of human individuality.

Beyond the new understanding of the human body in health and disease, other valuable consequences can be foreseen. One is the harnessing of biological processes for human welfare. Recombinant DNA methods are already being applied to convert bacteria into factories for synthesizing specific products, such as a brain hormone (somatostatin) and insulin. Synthesis of the powerful natural antiviral compound, interferon, is a goal in a number of laboratories. Synthesis of a wide range of pharmaceuticals is expected.

Other possibilities have yet to be examined. Will recombinant DNA methods make it possible to design and synthesize new enzymes to be used in place of chemical synthesis of many organic compounds? Are there other transformations of biological systems that can generate new energy sources, such as improved biological fixation of solar energy? Can other food plants be rendered, like legumes, independent of the need for nitrogen fertilizer? Industries may come to rely on genetic technologies.

Another consequence might be gene therapy—the correction of genetic diseases by replacing defective genes with normal ones. The possibility of success is hard to evaluate. We will soon be able to isolate or synthesize any gene we wish, but the obstacles to inserting them into body cells in a useful way are large. It seems likely that this step may become possible before too many years for the defective precursor cells of blood cells, since these are located in the bone marrow in a way that allows them to be replaced from the bloodstream. With more highly organized organs, however, the problems seem insuperable.

This type of genetic engineering would benefit the afflicted individual; it is much less likely that a gene replacement can be carried out in sex cells, as would be required to eliminate further inheritance of a genetic disorder.

CELL BIOLOGY

SCOPE AND DEFINITION

It has been known for 140 years that all higher organisms, regardless of size or complexity, are constructed of cells of about the same size (about 1/100 millimeters diameter). Only in the past two decades has the electron microscope permitted studies of visible structure to penetrate to ever finer dimensions. Meanwhile, biochemistry has moved up from small molecules to the detailed structure of giant molecules, such as proteins and nucleic acids, and indeed to the isolation and analysis of even larger components of the cell. Biochemists and morphologists have thus met in what was formerly a no-man's land between their domains.

The dynamic, functional properties of the cell have been advanced by new biophysical techniques, such as study of the electrical potential across cell membranes, and by the use of mutations in microorganisms to produce a variety of sharply defined changes in specific cell components. We are beginning to understand how many components of cells fit together, and even beginning to carry out such assembly in the test tube.

UNIFYING PRINCIPLES OF CELL FUNCTION AND STRUCTURE

In the 3 billion years since life arose on this planet, millions of species have evolved, from relatively simple bacteria to highly complex plants and animals. Contemporary cells thus vary greatly in their properties; the structural organization, functions, and life styles of, for example, a bacterial cell, an amoeba, and a human liver cell might seem to have little in common. But despite the glaring differences, these cells share profound similarities in the principles underlying their structure and function. All cells store their genetic information in DNA, and they use the same genetic code and much the same kind of machinery to transfer this information into proteins. Every cell has an outer membrane that determines which materials may pass in and out of it. And though various cells derive energy from very different foodstuffs—for example, some bacteria oxidize sulfur to sulfuric acid—this energy is utilized, by common final pathways, to synthesize ATP (adenosine triphosphate), which provides the energy needed to drive reactions within all living cells.

Although it is, thus, justifiable to speak of "the cell," it is also often useful to distinguish two groups with major differences in their complexity: the prokaryotes (bacteria) have only one chromosome and no enclosed cell nucleus; the more complex eukaryotes (all species of animals, plants, and higher microorganisms) have multiple chromosomes and a nuclear membrane.

A bacterial cell is about 1/1,000 as large as most animal or plant cells. It is a good deal simpler (containing perhaps 5,000 genes, whereas the DNA in a human cell is equivalent to 1 million genes), and it has been much easier to study genetically. Therefore, the mechanisms that regulate and integrate the activities of the genes and their products are much more completely understood in bacteria than in the cells of higher organisms. However, with the knowledge derived from these simpler cells, and with the recently developed possibility of carrying out genetic studies in cultures of eukaryotic cells just as in cultures of bacteria (e.g., isolation of mutants; genetic recombination between different mutants), knowledge of the cell biology of higher organisms is just beginning to reach a similar level of sophistication.

Cell Adaptation and Evolution

The cells of unicellular organisms, such as bacteria, can to some degree adapt to changed environments by turning appropriate genes on and off. There is a counterpart of this process in multicellular organisms. Liver cells, for example, respond to the chronic presence of drugs, alcohol, and other toxic substances by increasing their synthesis of a limited group of enzymes that destroy such substances. Muscle cells respond to their own exercise (contraction) by synthesizing more contractile proteins, and immune cells that synthesize a given antibody multiply faster in response to stimulation by the corresponding antigen.

In addition to such reversible adaptations, cells of unicellular and multicellular organisms can undergo changes in genes that are inherited by the descendants of those cells. A particularly striking case of such cellular evolution occurs among tumor cells. We have long known that these cells may undergo a genetic change as evidenced by visible changes in their chromosomes. We have also learned that tumor cells exposed to a chemotherapeutic drug sometimes become resistant to the drug, in some cases because of the emergence of genetically altered tumor cells endowed with a means of resistance. For example, the drug methotrexate blocks an enzyme necessary for cell duplication. Recent research has shown that cultured cells can become resistant to methotrexate by overduplicating the gene that codes for that enzyme. A normal cell has one or a few copies of that gene, but a resistant cell has several hundred copies.

This development of resistance to methotrexate provides a guide for investigating the evolution of resistance to other drugs; it also reveals an unexpected instability of the genes, a problem of the broadest biological significance. Since the methotrexate has simply selected for those occasional spontaneous mutants that have duplications of a particular gene, similar duplications must be appearing for all kinds of genes as cells multiply, but the changes are perpetuated only when they are useful.

Cell Reproduction and Its Regulation

Every organism requires cell reproduction for its long-term existence, and every cell arises from a preexisting cell by division. A human, for example, starts as a single fertilized egg cell and grows to 100 trillion cells in adulthood. Cell reproduction continues throughout life. Although brain and muscle cells are long lived, certain kinds of white blood cells are replaced every day.

The division of a cell is preceded by a period of growth, during which all of the structural components increase in size. A striking feature of this growth is its balance: Each part increases in proportion to all other parts. The kinds of intricate regulatory mechanisms that integrate thousands of gene and enzyme activities to achieve this balance in prokaryotic cells are reasonably well understood, but a great deal of research will be needed to reveal the full nature of the regulatory interactions of our own cells.

Another major preparation for cell division is the duplication of the chromosomes, so that a full set of genes can be distributed to each of the two daughter cells at cell division. But the molecular control of the switch that regulates the initiation of DNA replication remain unknown.

These are critical areas, not only because of their importance for normal cell reproduction, but also because the conversion of a normal cell into a cancerous one involves loss of its ability to control its reproduction. Tightly coupled to this loss is the cell's failure to continue to perform its normal, specialized functions. Thus, cancer cells are characterized by uncontrolled reproduction and by functional defects. Both of these changes are considered to derive from one proximate cause, still unknown.

Carcinogens and Mutagens

A paramount question is how certain agents, in whole animals or in cell cultures, convert a normal cell into a cancer cell. Particularly puzzling is the extraordinary diversity of chemical structures active in this regard. It is difficult to imagine some common aspect among the small and large molecules, which can be water-soluble, nonpolar, acidic, basic, electrophilic, or nucleophilic. Since the change is passed on through cell division to all of the descendants of a single original cancer cell, in the absence of the original causative agent, the original cell must have undergone a change in gene structure or a persistent change in gene regulation. Indeed, radiation and many cancer-producing chemicals can cause mutations, and tumor viruses (oncogenic viruses) bring additional genes into cells. We may therefore reasonably hypothesize that cancer is caused by persistent changes (mutations) in a limited number of genes that affect normal cell reproduction.

According to this hypothesis, radiation and chemicals are carcinogens largely because they are mutagens. The case for radiation is strongly documented: All forms of ionizing radiation and ultraviolet light radiation are mutagenic and all can cause cancer. Admittedly, some evidence suggests that radiation may also cause cancer in animals by activating a latent oncogenic virus already in the cell or by activating a chemical carcinogen.

A particularly puzzling aspect of carcinogenesis is the long lag—sometimes decades—that usually intervenes between initial exposure to the carcinogen and the appearance of the cancer. Probably, this delay will be understood only when the origin of cancer cells and the nature of the biological defenses against such cells are known in greater molecular detail.

A major question about radiation-induced cancer is whether there is a lower limit below which radiation does not cause either irreversible cell damage or cancer; that is, is there a safe level of radiation? Some evidence suggests a threshold below which radiation does not permanently damage or change cells, perhaps because the cellular DNA repair mechanisms can keep up with DNA damage when it is infrequent enough. Other evidence suggests that the mutagenic and carcinogenic effects are proportional to the lowest levels tested. The question remains open because human exposures, to both radiation and chemicals, are generally to dosage levels far below those tested in the laboratory. This problem is difficult and it continues to be the subject of research.

The proposition that chemical carcinogens are mutagens is currently under extensive study. Results of the Ames test, which measures the capacity of a substance to cause certain mutations in bacteria have been positive for about 90 percent of the hundreds of chemicals that are known to be carcinogenic.

Because of this striking evidence, any chemical that is mutagenic in the bacterial test must tentatively be considered a possible carcinogen. Yet there is doubt that mutagenesis may be equated to carcinogenesis. Considerable effort is being made to substitute mammalian, including human, cells in culture for bacteria in order to develop a more directly applicable test that is still fast and inexpensive.

While recognition of the role of environmental carcinogens has been a major advance, it would be a serious mistake to think of these agents as the unique cause of cancer. Just as in the simpler case of mutations, there is undoubtedly a background rate of carcinogenesis that arises through the inherent production of errors in the process of gene duplication, and by background radiation. Mutagens/carcinogens increase the frequency of such errors.

Tumor viruses also cause cancer in various animals by altering a cell's genetic makeup, but in this case the change is not mutational. Rather, part or all of the viral chromosome becomes integrated into a chromosome of the cell. One or more of the integrated viral genes functions somehow to override the cellular mechanism that regulates cell reproduction.

Studies of the viral induction of cancers have thus made major contributions to understanding virus–cell interactions. It may also be easier to trace the mechanism of deregulated cell growth when it is caused by one or very few added viral genes, rather than by the action of a mutagen that can change any gene in the cell. Moreover, there is strong evidence that some human cancers are of viral origin—for example, Burkitt's lymphoma, which is prevalent in parts of Africa.

The concept that cancer is the result of a genetic change in a cell bears on the question of a hereditary disposition to some kinds of cancer. Cancers of the eye (retinoblastoma) and of the kidney in children involve inheritance of a specific defective gene from one parent or the other. The basis for the less rigid but seemingly real inheritance of susceptibility to some other kinds of cancer, such as breast cancer, remains to be uncovered.

In a sense, any research that increases knowledge of the normal and abnormal behavior of cells may have the potential to increase our understanding of the cancer cell. But in particular we must continue the large effort to elucidate the molecular mechanisms that underlie normal cell reproduction, regulation of that reproduction, and cell differentiation, as well as the transformation of a normal cell to a cancerous one.

Cell Membrane and Transport

The various interactions of cells with their environments intimately involve the cell's outer membrane. This extremely thin, continuous barrier prevents cell components—even very small molecules—from leaking out of the cell, and it similarly prevents substances from entering the cell except by special uptake mechanisms.

The membrane is essentially a continuous layer of lipid (fatty) molecules, much like a soap bubble but stronger. These molecules are closely packed, creating the impenetrable barrier that separates the cell from its environment. Special protein molecules embedded in the membrane between the lipid molecules enable selective entry and exit (transport) of appropriate substances.

Some of these transport systems act as pumps, actively concentrating in the cell interior substances that are present in the environment only in low concentration. As a result, bacteria growing in very dilute external environments, such as pond water, have concentrations of substances inside their cells just as high as bacteria grown in a rich culture medium in the laboratory.

In the past two decades, a number of proteins that bind specific small molecules and transport them across the membranes has been isolated. More recently, some insights have been developed into the mechanism that converts metabolic energy into the work of active transport by such systems; however, the details of this conversion remain a great challenge.

Membrane pumps are responsible for the remarkably constant concentration of ions (e.g., potassium, magnesium) in all cells of the mammalian body. Other molecular pumps in kidney cells, selectively excreting ions and waste products, ensure that the fluid surrounding body cells (and in the circulation) enjoys an equally constant, though very different, composition. In addition, the mechanism responsible for the electrical impulses characteristic of nerve cell activity exploit transient alterations in membrane permeability.

Nutrient and ion transport is only one aspect of research into the interactions of cells with their environments. We are only beginning to understand chemotaxis, the process by which motile cells sense and respond to the presence of substances that cause them to change direction of their motion. In bacteria, for example—in a striking breakthrough—some molecular components of this process have been recognized.

In multicellular organisms, such directed movements of cells, far more complex than in bacteria, are extraordinarily important and still very poorly understood. White blood cells, for example, find their way toward sites of infection, where they engulf the invading organisms and destroy them. Similarly, during embryonic development a myriad of directed cell movements occurs as tissues and organs are being formed, guided by cell-to-cell interactions. Regulated cell movements also occur in tissue repair after damage and in normal tissue maintenance in an adult organism. The response of the immune system (including the chemotaxis of white blood cells just mentioned) similarly depends on a complex set of specific surface interactions among the various kinds of cells that make up the immune system. The detailed molecular mechanisms of signal and response are a major challenge.

Membrane Receptors

Until recently, only molecules present in soluble form in cells or extracts could be purified and identified by biochemical procedures; organized microscopic structures that remained after cells were broken up (membranes, for the most part) were sometimes labeled "cell debris"; their components could not be separated and manipulated by available techniques. Recently developed methods, however, can isolate cell membranes, separate their components, and identify specific protein molecules that have the properties of receptors, *viz.*, they specifically bind to some compound.

This extension of biochemical analysis has revolutionized several fields. In endocrinology, which is concerned with the formation of various hormones in certain cells, the molecular basis for hormonal actions on the development and the function of various target cells is being scrutinized. For example, when a steroid hormone such as an estrogen reaches its destined target cells, it binds to specific membrane receptor molecules and enters the cell as a complex. This complex migrates to the nucleus where its binding then turns some genes off and others on, resulting in major changes in the cell's metabolic activity. Similarly, several powerful bacterial toxins, such as those of diphtheria or of cholera, are now known to complex with specific membrane receptors, enter the cell, and alter the activity of specific enzymes.

Finally, the focus on cell surface receptors has also led to the discovery of a number of novel, hormonelike transmitters in the brain, which stimulate or inhibit nearby target cells. Awareness of membrane receptors and the capability to study them is but a decade old; a wealth of understanding must surely lie immediately ahead.

Chromosomes

Chromosomes, whose DNA constitutes the genes, play a most important role in determining the structure and function of the cell. In any mammalian cell only a small fraction of the genes is active. An important advance of the last few years has been the discovery that the DNA is coiled in a highly regular pattern. Whether this coiling pattern is related to the activity (expression) of genes in the DNA, and what role other chromosomal proteins play in chromosome function, are central questions now receiving major attention.

In ordinary cell division by mitosis all the chromosomes in the cell nucleus duplicate. The resulting pairs (still connected) condense into tightly packed bodies, and then the two products of each duplication are pulled apart to form two groups of daughter chromosomes. In the next step of this process these two groups are drawn to opposite sides of the cell, which then splits into two halves. An identical group of chromosomes is distributed to each daughter cell.

The essentials of this process, visible in the light microscope, have been known to cytologists since the beginning of the century. Cell biologists are now trying to clarify the underlying molecular mechanisms: What makes the duplicated chromosomes condense? How and by what are they moved to opposite ends of the cell? How does the cell split into two daughters? What is the motile mechanism so busily engaged in the division process? (It begins to appear that this last has features in common with other aspects of cell motility.)

The Cytoskeleton and Cell Motility

The most obvious form of cell motility is the contraction of muscle cells to perform mechanical work. However, cell motility takes a variety of other forms, such as the beating of flagella (as in sperm, flagellated protozoa, algae), the waves of movement of cilia (which, for example, sweep mucus along the lining of the bronchial tubes of the lung), the streaming of cytoplasm in a cyclic manner (as in nerve cells, many plant cells, and algae), and the migration by protrusion of regions of the cell (as in amoebae, white blood cells, some types of cancer cells, and all animal cells in culture). All these forms of motility appear to be based on specialized intracellular protein molecules that are capable of sliding along one another to produce contractions, but only in muscle is the role of proteins in the contraction understood in any detail. It will be necessary to unravel the various mechanisms of cell motility if we are to understand such phenomena as cell migration in development, migratory movement of white blood cells, migration of cancer cells, etc.

Cell motility involves the internal cytoskeleton of a cell. This structural component, scarcely known to exist a decade ago, is made up of a system of extremely thin fibers that determine cell shape and give the cell its mechanical rigidity. It is not a skeleton in the usual sense, since the cell constantly assembles and disassembles the fibrous components, changing its shape from minute to minute.

Cytoskeletons are built of microtubules and microfilaments. Microtubules are rigid fibers that are constantly growing longer in one part of a cell and disappearing in another, as their specific component protein molecules are being assembled or disassembled. Microfilaments are built of a protein, actin, that participates in muscle contraction, and also creates a gelatinous texture in parts of virtually all eukaryotic cells. They are often assembled into fiber bundles that function as miniature muscles, providing the force, against the microtubular network, that changes the cell shape and causes movement.

The plasticity of the cytoskeleton is dramatically demonstrated by changes that occur during cell division. Immediately prior to division, the whole cytoskeletal system is disassembled, and the molecular parts are recruited to build the apparatus that moves chromosomes and splits the cell in two. When division is complete, the division machinery is disassembled, and the molecular parts reassemble into new cytoskeletons in the daughter cells.

Finally, the cytoskeleton is also modulated by environmental signals that induce the cell to change directions. Study of the molecular basis of the cytoskeleton's dynamic nature is enjoying rapid progress.

Cell Aging

Normal animal cells in culture undergo aging; they can continue to reproduce for only a finite period, usually a few months. In contrast, cancer cells grown in culture are very nearly immortal, reproducing without limit. The prime example is the human cervical carcinoma cell known as HeLa, which was removed from a patient in the early 1950's and is still reproducing as vigorously as ever.

Normal cells taken from an elderly individual have shorter life-spans in culture than those derived from a young individual. This finding suggests that the limited life span of normal cells in culture reflects the normal aging process of cells in an animal. Much current research is aimed at relating the possible role of the change observed in cultured cells to the processes of senescence of animals and at identifying the molecular bases of the immortality of cancer cells.

Cell Differentiation

In the development of a multicellular organism from a fertilized egg, cells become differentiated, that is, specialized to perform particular functions. Since all these cells contain an apparently full set of genes (though recent developments suggest that there may be fine differences), the great differences in their function and structure are due to differences not in genetic makeup but rather in the ensembles of genes that are expressed and their degree of expression. Thus muscle cells produce large amounts of contractile proteins, endocrine cells are specialized to synthesize hormones, certain white blood cells produce antibodies, etc. A key element in differentiation is therefore the molecular events that regulate gene expression differently in different cells, and that perpetuate the differentiated state. In contrast to gene regulation in bacteria, we know virtually nothing about this process, or about the mechanisms by which differentiation prescribes different lifetimes for different animal and plant cell types. For many years to come, the problems of cell differentiation will present a virtually endless frontier.

CELL FUNCTION AND STRUCTURE IN HEALTH AND DISEASE

All diseases in plants and animals are expressions of the malfunctioning of cells of one kind or another, or of a tissue or organ made of specific cells. For example, circulatory diseases, which currently account for 48 percent of deaths in the United States, are mostly caused by the behavior of certain cells in the walls of arteries. As a result, patches of deposited lipids, called plaques, form on the inner surfaces of arteries (atherosclerosis), block the flow of blood, and serve as foci for formation of blood clots. Recent research has suggested a resemblance between this process and cancer; plaques appear to start as tiny overgrowths of smooth muscle cells in the arterial wall—something like a benign tumor.[3] The early stages of this process are visible in a high proportion of the aortae of young Americans who die of trauma.

Cancer is a particularly clear example of the cellular basis of disease: Its essential feature is a persistent loss of some, as yet unidentified, normal genetic regulatory mechanism, resulting in the overproduction of cells that are functionally defective. Aging, not a disease process in the usual sense, is a result of poorly understood losses in the functions of cells.

To increase our understanding of disease processes (and aging) we must continue the current, broad-based research on the normal structure and function of cells; continue to investigate the molecular basis of degenerative changes or defects in structure and function of diseased cells; and identify the specific causes of such degenerative changes or defects. Collectively, these activities will continue to account for a major thrust of research in cell biology for the next decade or more.

IMMUNOLOGY

Immunology is concerned with a broad range of phenomena relating to human disease, some beneficial and some harmful. Best known is resistance to infectious disease, induced either by natural infection or by artificial immunization. The latter practice has virtually eliminated diphtheria, whooping cough, poliomyelitis, and tetanus in the United States, and smallpox worldwide. On the other hand, the immune system can also occasion harmful reactions, including allergies, rejection of transplants, and various autoimmune diseases in which the body reacts against its own components. The major medical purpose of immunological research is to enable manipulation of the immune system so as to augment the beneficial effects and curtail the harmful ones.

After more than half a century of descriptive observations, the past two decades have witnessed the development of deep insights into the cellular, genetic, and molecular mechanisms of the immune system. Immunology has also become of great interest to other biological sciences, more broadly taken. It provides a model system for studying the challenging problems of selective gene activation during development of the fertilized egg, and it has provided extremely sensitive, highly specific analytical tools for quantifying many compounds.

NATURE OF THE IMMUNE SYSTEM

The main elements in the immune system are lymphocytes, small white cells found in the bone marrow, lymph nodes, spleen, and circulating blood. One class of lymphocytes secretes protein molecules called antibodies or immunoglobulins (Ig). Each individual lymphocyte makes many copies of one specific antibody, each molecule of which can bind very tightly to two molecules of a given foreign substance (antigen). As a consequence of binding to antigens, antibodies can neutralize toxic products of bacteria, or promote the uptake of bacteria by scavenger cells that destroy them, or prevent the penetration of viruses into cells.

Another class of lymphocytes also bears antibody-like molecules on their cell surface, but does not secrete them; instead, by means of these molecules they can attach to, and kill, cells that carry the corresponding specific antigen on their surface: for instance, infecting microbes, foreign (transplanted) animal cells, and host cells infected with a virus.

Antibody Molecules

The binding sites of antibodies are relatively small regions, each of which fits the combining site of the corresponding antigen as precisely as a lock matches a key. The repertoire of antibodies in an organism is enor ous: There are perhaps a million.

The total amino acid sequences of several antibodies are known and their general three-dimensional structures are apparent (Figure 5). Antibodies of different specificity hold large regions in common but differ at their binding sites.

Several distinct classes of human antibodies exist, each with a unique structure and biological function, and a set of binding sites for the full range of antigens. For example, the class called IgE, constituting less than 0.1 percent of the total antibody population, is

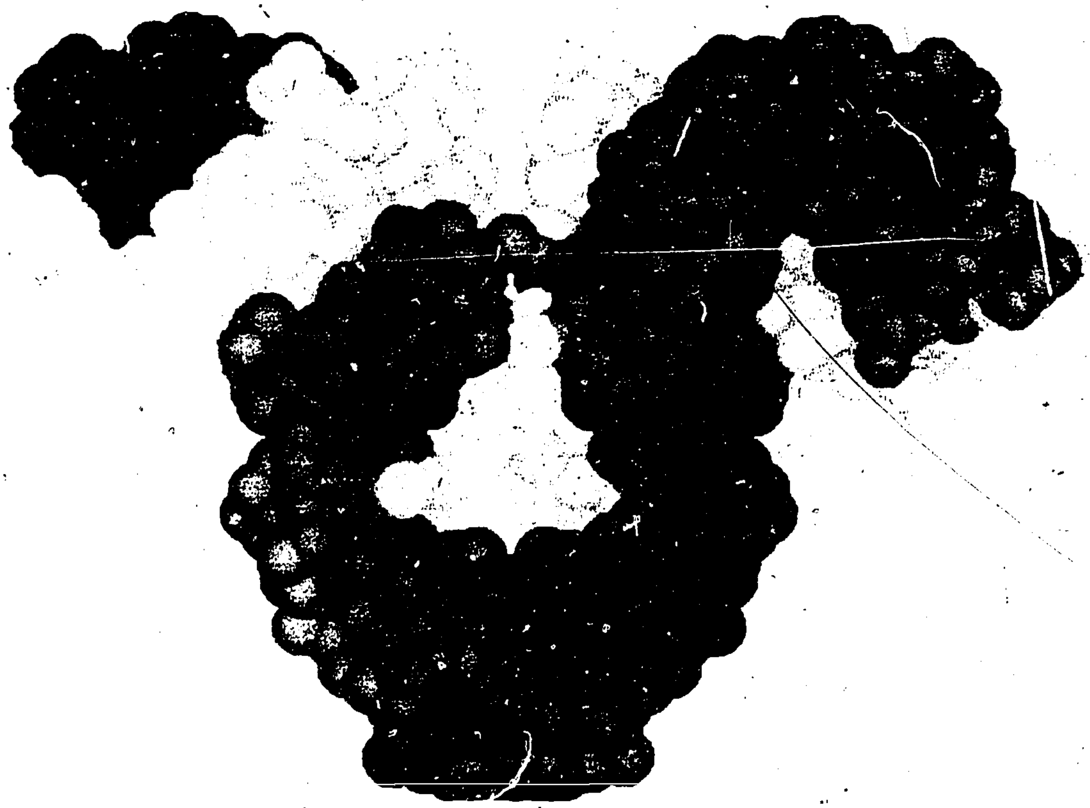


FIGURE 5 Computer-generated model of intact human antibody molecule. The two heavy polypeptide chains appear dark gray and almost black, both light chains are lighter gray. The almost-white circles near the center make up a complex carbohydrate. (M. A. Navia. Coordinates from E. W. Silverton, M. A. Navia, and D. R. Davies, *Proceedings of the National Academy of Sciences* 74:5140-44, 1977.)

responsible for allergies. Contact with allergens such as ragweed pollen or penicillin may provoke formation of several classes of antibodies, but only the IgE antibodies cause allergic reactions.

Lymphocytes

A number of different classes of lymphocytes, with major differences in function, have been recognized in the past two decades. This knowledge should lead, perhaps in a very few years, to beneficial control over formation of these alternative responses to an antigen.

B cells, arising in the bone marrow, are the lymphocytes that produce antibodies. A mammal contains hundreds of thousands of different B cells, each specific for a particular antigen. Each kind of cell is normally present in very small numbers, but when such a cell comes into contact with the corresponding antigen it divides rapidly, and all of its descendants secrete the same antibody molecules.

T cells are also lymphocytes of bone marrow origin, but they are processed further, in some unknown way, in the thymus gland—whose function was unsuspected 20 years ago. As a result of recent research, three major subclasses of T cells are now recognized, each with distinct biological functions; within each subclass, different T cells have specificity for different antigens.

The *killer T cell* is one of our principal defense mechanisms; it can attack and destroy tissue cells that have been invaded by certain viruses. Killer T cells can attack tumor cells, too, at least in experimental animals; and they are largely responsible for the rejection of skin or organs transplanted from one individual to another. The success rate in kidney transplantation has been increased markedly through the development of drugs that prevent the multiplication of killer T cells, as well as through more careful matching of recipients and donors so that the antigens present on cell surfaces of tissues from donor and recipient closely resemble one another.

The *helper T cell*, once activated by a particular antigen, stimulates the corresponding B cell to divide and release antibodies.

The *suppressor T cell*, discovered very recently, plays a major role in controlling the level of the immune response, including the tolerance that precludes us from forming antibodies to our own constituents. An excess of suppressor T cell activity can shut down the immune response to an antigen, whereas a deficiency can lead to an exaggerated response. One of the most promising areas in immunology deals with the development of methods for controlling the level of suppressor T cells specific for particular antigens.

Molecular Basis for Antibody Diversity

When molecular genetics showed that each protein chain is coded for by a corresponding gene, immunology faced a major problem: Could the chromosomes of a mammal carry a separate gene for each of hundreds of thousands of specific antibodies? There is increasing evidence for a quite different mechanism, in which the individual inherits a much more limited number of genes for antibodies, and mutations in these genes, or recombinations between them, occur at an exceptionally high rate in the precursor cells, thereby giving rise to the great variety of final, differentiated, antibody-producing cells. This unusual behavior of the genetic material in these special precursor cells is under intense investigation.

Cell Fusion; Development of Permanent Cell Lines Producing Antibodies

Most large antigens possess many different combining sites, recognized by different antibodies. Moreover, even a single such site can be recognized by several different antibodies of overlapping specificity. Consequently, when an antigen is injected into an animal, the antibodies evoked are highly heterogeneous.

This heterogeneity has been a serious obstacle to molecular studies on antibody formation, structure, and function. However, an elegant solution has recently been developed. If an antibody-producing cell (which has a life span of only a few days) is fused with a cell from a certain tumor line (which can be cultivated in the test tube indefinitely), the resulting hybrid product—which encloses the chromosomes of both in a single membrane—can multiply and secrete the desired single antibody indefinitely. These homogeneous antibodies promise immense variety of applications, in research and in practice.

APPLICATIONS

Vaccines

Since the nineteenth-century discovery that infectious agents killed in certain ways retain the capacity to elicit antibodies, such preparations have been used as vaccines with great success in preventing a number of diseases. Live (attenuated) vaccines have also been developed against viruses (and the tubercle bacillus), through the use of harmless mutants. Immunization against viruses (e.g., polio) became much more feasible after methods were developed for growing the virus inexpensively and in large quantities in cultured animal cells, rather than in infected animals.

Some killed organisms have proved too toxic for use in immunization. However, the range of effective vaccines is now being extended by purifying the surface antigens of bacterial cells (which are the components attacked by protective antibodies) and eliminating the more toxic, deeper constituents.

New vaccines are contributing to the control of a variety of disorders. Pneumococcal (lobar) pneumonia still kills 25,000–50,000 Americans annually (especially older or debilitated persons), and a vaccine against the 14 commonest types was licensed in 1978. About 20,000 Americans contract bacterial meningitis annually, with around 3,000 deaths, and antigens against two of the three meningococcus types (A and C) have been licensed and are now given routinely to members of the armed forces. German measles (rubella) is a mild disease, but in pregnant women the virus is often lethal or damaging for the fetus. Following the 1964 epidemic, some 40,000 infants were born with mental or other defects. The virus was first isolated in 1962, and an attenuated vaccine has now been licensed.

Viral hepatitis A (infectious hepatitis), with about 30,000 cases reported annually in this country, is usually transmitted by fecal contamination. We have not learned how to cultivate the virus, and so it has not been possible to produce a vaccine against this debilitating disease. However, with the similar hepatitis B (serum hepatitis), an antigen from the virus, designated as HBs, was accidentally discovered in the blood of Australian aborigines, in the course of a search for new blood types (for which it was at first mistaken). The incidence of this disease, which is usually transmitted by blood transfusions, has been markedly decreased by screening all blood being used for transfusions for this antigen. Moreover, HBs antigen, isolated from the blood of carriers and then inactivated, appears promising as a vaccine.

A warning note is in order. When a sufficient fraction of all children is immunized against a disease such as pertussis or measles, the incidence of the disease markedly declines. One result is that nonimmunized children are very much less likely to contract the disease and develop active immunity. When they, as adults, travel to countries where the disease is prevalent, they are highly susceptible to an infection that is far more serious in adults than in children. Similarly, if the very success of a vaccination campaign engenders subsequent carelessness, with a decreasing fraction of children being immunized, the stage can be set for a future epidemic in the adult population.

Parasitic Infections

Parasitic infections of man constitute the main health hazard in tropical regions, where they affect hundreds of millions of individuals. In Africa alone an estimated million children die annually from malaria. A number of drugs are effective, to varying degrees, against some of these infections, and vaccines have been developed against several parasitic infections of domestic animals. So far, however, vaccines against parasites in man have not been successful.

Current work on malaria is very promising. The parasite that causes most human malaria has recently been cultivated in the test tube. This development provides a rapid test for antimalarial drugs, and it should also make many other kinds of research on this organism more effective and less expensive. The prospect for a vaccine against malaria in man now seems bright.

Rh Incompatibility (Hemolytic Disease of the Newborn)

All red blood cells have numerous antigens on their surfaces, including those of the Rh series. If a fetus inherits from its father antigen Rh-D, but the mother lacks that antigen, she may form antibodies to red cells that leak from the fetus into her circulation. These antibodies, in turn, can reach the fetus and destroy its red blood cells. The risk of this disease increases with each pregnancy: Fetal cells are particularly likely to invade the mother's circulation during childbirth, and a second antigenic stimulus elicits a larger response than the first. Prior to 1968, an estimated 10,000 infants died of this disease each year in the United States, and about 10 times as many were affected. However, the disease can now be almost completely prevented: Anti-Rh-D antibodies, injected into the mother immediately after each delivery, prevent her from forming antibodies to this antigen.

Organ Transplantation

Immune responses to foreign tissues are the main limitation to effective organ transplantation. Kidney transplants are the most common, but some progress is also being made in transplantation of bone marrow, heart, and liver.

Approximately 40,000 Americans with kidney failure are at present receiving regular dialysis treatments, at a total cost of over $500 million annually. The potential economic benefit of transplantation is thus very large.

Most kidney transplants come from recently deceased individuals, but only about 50 percent of these grafts have survived for four years, compared with 75 percent of those from properly matched related donors. However, major developments in the past three years promise to greatly increase the survival rate of cadaver grafts. Two sets of antigens that are present on all tissues, designated HLA-A and HLA-B, have generally been used for matching. In one study in which tissues were matched as well to a third set (DR), which is present on lymphocytes, 90 percent of cadaver grafts survived for at least two years.

There has also been progress in suppressing the immune system of patients receiving transplants, by using drugs, irradiation, or antibodies to killer cells.

Inherited Diseases of the Immune System

In several inherited diseases, the production of B and/or T cells is abnormal, and the patients often succumb to infections early in life. Some patients can now be treated with transplants of bone marrow, thymus tissue, or fetal liver.

Another serious abnormality of the immune system, hereditary angioedema (a type of giant hives or swelling), has been traced to alteration in the function of a component of the complement system (a set of proteins that interacts with antibodies to increase resistance to infection). This disorder can now be successfully managed with a synthetic steroid, which usually returns complement function to normal.

Allergies

The most frequent disorder of the immune system is allergy. A serious form, asthma, afflicts about 9 million persons and causes over 2,000 reported deaths a year.

Allergic reactions are caused by antibodies of the IgE class, which thus far are known to have only deleterious effects. Certain white blood cells (basophils) have receptors that can bind complexes of IgE with corresponding allergens. Binding triggers the release of substances, such as histamine, that then act on various other tissues, causing such symptoms as sneezing, rash, or hives. Identification of these agents can lead to the development of specific drugs, such as the antihistamines, that minimize their formation or interfere with their action on target tissues. Moreover, the receptors for IgE on basophils have been identified in the past two or three years, and this breakthrough may permit the design of chemical agents that will interfere with their interaction with IgE.

In another recent advance, certain suppressor T cells and helper T cells have been found to act specifically, in opposite directions, on the IgE response. Since the balance of the two cell types can be influenced (e.g., by the dosage and the form of antigen), this finding offers promise of a rational, guided approach to desensitization. Another new approach seeks to block the IgE response, on the basis of the finding that some antigens induce tolerance, instead of antibody formation, if they are linked to certain nonantigenic large molecules.

Purified specific components of many pollens that cause allergy are now being used to improve diagnosis. Moreover, these antigens can be used to measure the amount of corresponding IgE antibody in the blood, a procedure that has some advantages over the common skin test for allergy. These purified antigens are also being used in immunotherapy (also known as hyposensitization, or allergy shots), which presumably works by shifting the balance of immune responses.

Similarly, purified venoms of bees or hornets are proving much more effective in hyposensitization than the previously used extracts of whole insects. Unfortunately, because the market is limited, and extensive tests are now required before a new pharmaceutical product can be licensed, this advance is unlikely to be made generally available through the normal marketing mechanisms.

Autoimmune Diseases

Under ordinary circumstances one does not develop an immune response to constituents of one's own body tissues. However, occasionally regulatory mechanisms break down and an individual produces antibodies or active lymphocytes that react with specific tissues of his body. Such autoimmunity has been implicated in recent years in an increasing number of important diseases, including lupus erythematosus, rheumatoid arthritis, myasthenia gravis, hemolytic anemia, thrombocytopenic purpura, Addison's disease, hyperthyroidism, multiple sclerosis, Guillain-Barré syndrome, chronic glomerulonephritis, and possibly pernicious anemia and certain forms of diabetes mellitus. Autoimmunity may also appear after tissue damage or an infection. For example, after a myocardial infarction, antibodies to heart tissue can be detected, and in rheumatic fever or glomerulonephritis, which may follow a streptococcal infection, antibodies to certain streptococcal antigens cross-react with antigens on the specific tissues.

The mechanisms that normally prevent autoimmune responses are referred to as immunologic tolerance; that is, one is normally tolerant to one's own tissues or body constituents. The recently discovered suppressor T cells play an important role in mediating tolerance, and a major advance in understanding is promised by the finding that the level of specific suppressor cells and of corresponding antigens varies with the activity of certain diseases, such as lupus erythematosus and juvenile arthritis. In addition, it has been found that killer T cells, which may be involved in autoimmune processes, can be inactivated by antibodies directed against specific receptors on them.

This newly acquired and rapidly increasing understanding of the mechanism of tolerance offers promise of leading, relatively soon, to improved methods for identifying and controlling autoimmune processes. For example, many patients with myasthenia gravis have shown improvement after alteration of immune function by removal of the thymus gland, or by treatment with steroid drugs.

Immunology and Cancer

It has been possible to transplant tumors, without rejection, by using highly inbred, genetically almost identical mice, which no longer react immunologically to each other's normal tissues. Certain tumors were thus found to have characteristic antigens that elicit immune responses in the recipient host. These responses are detected because they result in rejection of a second transplant from the same donor. The occasional unexpected spontaneous regression of a malignant tumor in man may well be due to an immune response. It has been hoped that immunotherapy would completely eliminate any cancer cells remaining after chemotherapy, just as the two procedures complement each other in the therapy of infections. So far, however, results have been disappointing.

Immunological Analytical Techniques

The extreme specificity of antibodies, combined with radioactive labeling, has recently led to their widespread use in radioimmunoassay as analytical tools to measure very low concentrations of many important biological substances, including hormones, drugs (such as digoxin, gentamicin, or penicillin), and metabolic intermediates. A modern clinical laboratory may run 30 different tests of this type routinely, and hundreds are used in research. This approach has revealed, for example, that many patients with diabetes produce adequate quantities of insulin but inactivate it with their antibodies.

A powerful derivative tool has been the attachment of fluorescent compounds to antibodies. These reagents can single out a specific bacterial cell in a population of other bacteria, or identify organisms in tissues, making possible much more rapid diagnosis than that dependent on cultures. (In Legionnaire's disease, diagnosis depends entirely on immunofluorescence.) In addition, immunofluorescent staining of normal cell constituents has become a major tool in studying the organization and dynamics of animal cells. For decades antibodies have been used as diagnostic tools to identify infectious microbes: Known antibodies are used to identify an organism recovered from a patient, and known organisms are used to detect a rise in antibodies in a patient.

Interferon

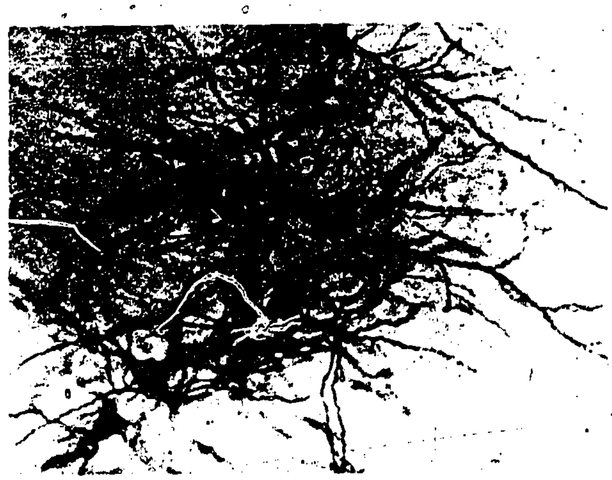
Interferon is a powerful antiviral protein manufactured in tiny amounts by the body's own cells when they are infected by a virus. When this substance is released, it renders nearby cells more resistant to infection by any virus. Interferon has proved therapeutically effective in preliminary tests. However, both research and application are limited by the extremely small quantities available. Recombinant DNA technology may soon alleviate this problem.

FUTURE DIRECTIONS

The past 20 years have witnessed a revolution in research in immunology, with improved tools for study at a cellular and a molecular level. The results have revealed a cascade of reactions, and a balance between antagonistic processes, which offer many opportunities for therapeutic intervention. In addition, the recombinant DNA technology should soon make available unlimited quantities of any specific component of the system. The next years therefore offer strong promise of marked increase in our ability to control both desirable and untoward immune reactions, which are relevant for a large fraction of all human disease.

NEUROSCIENCE

Neuroscience is the comprehensive research endeavor that seeks to understand the biological bases of



Photomicrograph of nerve cells. (Fritz Goro, Courtesy of Polaroid Corporation) behavior and experience. The human brain is the most remarkable and complex structure in the known universe. Composed of more than 50 billion nerve cells, each connected with as many as a thousand other nerve cells, a single human brain has a greater number of possible connections among its nerve cells than the total number of atomic particles in the entire universe. This three pounds of tissue is the structure that has allowed the human species to dominate its environment, write poetry, compose symphonies, make war, and undertake scientific understanding of the cosmos and of itself.

Whatever the mind is, it is in the brain. Application of physical, biological, and behavioral techniques to the study of the nervous system has just begun to reveal the way individual nerves function, the functional organization of the brain, the nature of sensations and perceptions, the way language is organized, and the physical basis of consciousness. Indeed, by the twenty-first century, neuroscience may well become the dominant science.

Study of the brain and nervous system developed as a special field within each of a number of traditional disciplines; anatomists worked on the structure of the brain, chemists focused on its chemistry, physiologists studied the electrical activity of nerve cells, psychologists analyzed the behavioral functions of various parts of the brain, and so on. Only recently, these efforts began to merge as the single enterprise called neuroscience.

The goal of neuroscience is a comprehensive scientific understanding of the nervous system and brain, particularly the human brain, extending to the final product of the brain's activity: behavior and experience. With this understanding should come treatment for many mental and neurological disorders; psychoses such as schizophrenia and depression, mental retardation, learning disabilities, stroke, blindness, Parkinson's disease, multiple sclerosis, head and spine injuries, drug addiction, muscular dystrophy, myasthenia and even, if we are very lucky, more effective approaches to the prevention of crime and war. These are, of course, long-term goals. The young field of neuroscience has made impressive advances in the past few decades and will accomplish much more in the next few years. This report is focused on the immediate past and future.

THE NERVOUS SYSTEM—A BRIEF OVERVIEW

To facilitate discussion of recent achievements and prospects, it may help to begin with a brief summary of the subject matter of neuroscience: the nervous system. In vertebrates, the brain and spinal cord are a continuous structure that is termed the central nervous system (CNS). The peripheral nervous system comprises the various nerves that lie outside the CNS, mostly sensory nerves that convey information from receptors to the CNS and motor nerves that convey commands from the CNS to muscles.

Although many invertebrates have rather differently organized nervous systems, some of them relatively quite simple, the basic plan of the human nervous system can be traced back at least to the worm, with its tubular nerve cord. Even in the worm, the head end of the nerve cord is enlarged to accommodate the specialization of the front end of the worm for sensing and feeding. The human CNS maintains this basic tubular organization from the spinal cord up to about the middle of the brain. However, the tube's front end, the forebrain, is enormously expanded and laid back over the core tube to form most of the matter we call the "brain."

The Neuron

During the nineteenth century the gross structures and subdivisions of the brain were described by anatomists. However, the nature of these structures was not at all clear. The Spanish neuroanatomist, Ramon y Cajal, mapped out the detailed microscopic structure of the subdivisions of the brain and established the neuron doctrine, *viz.*, the nervous system is composed of a great many discrete cells, the nerve cells or neurons, that influence one another by specific connections in a highly patterned manner. These connections are termed "synapses." The electron microscope revealed that a synapse is not a direct or continuous connection but rather a very close approximation (about 20 millionths of a millimeter) between a terminal of one neuron and the cell membrane of another neuron. On a typical neuron in the brain are hundreds of synaptic endings from other neurons.

The characteristic neuron has a central cell body, containing the nucleus and much of the biochemical machinery of these specialized cells. From the cell body extend fibers as long as 3 feet in man and over 60 feet in the whale. Most chemical substances in the neuron are synthesized in the cell body and then transported out to the fibers. The principal fiber, the axon, conducts information to an end-organ, for instance, muscle, or, in the brain, the end of the axon subdivides into numbers of small fibers that end in synaptic terminals on other neurons. Other fibers, dendrites, are specialized to receive information from synapses with other neurons, as is the membrane of the cell body itself.

The brain, then, is essentially a vast network of interconnecting neurons. Each neuron receives information from other neurons or from sensory receptors and transmits information to still other neurons or to muscles or glands to produce behavior. All stimuli impinging upon us, all sensations, thoughts, actions, and feelings, must be "coded" into the languages of the neuron. A particular neuron is affected at all times by many other neurons, some of which activate, some inhibit, in a graded-decision process. When the sum of these inputs activates sufficiently, the nerve cell transmits this information along its axon fiber by propagating an electrical action potential (the all-or-none digital language of the nerve fiber), which reaches to all the distal presynaptic endings of the axon. The basic processes of neuronal activity thus are synaptic input, the action potential and synaptic transmission.

Research in the 1940's and 1950's, primarily in England and the United States, led to an understanding of the nerve action potential. This electrochemical process, a wave of altered potential across a localized region of cell membrane, travels down the axon at a speed of 10–50 meters per second to all the synaptic terminals of that neuron. Once started, the action potential runs its course and is not normally influenced by other processes in the nervous system.

When the action potential arrives at a presynaptic terminal—the end of the axon that forms the synaptic connection to another neuron or muscle fiber—the presynaptic terminal releases a small amount of a chemical substance. This substance, called a transmitter, diffuses across the narrow synaptic space and binds to specific receptors, protein molecules lodged in the membrane of the second or postsynaptic neuron. This binding causes a small change in the electrical potential across the postsynaptic membrane, most prominently in the immediate vicinity of the synapse. For the entire neuron so affected, the change is also graded; the more synapses that are thus active at any moment, the greater the total change.

Several different compounds are now known to serve as synaptic transmitters. They are all relatively small molecules (molecular weights 150–3,000) that are either normal constituents of most cells, e.g., amino acids, such as glycine and glutamic acid, or are made from such, e.g., noradrenaline, acetylcholine, peptides, γ-aminobutyric acid (GABA). A given synapse uses only one chemical transmitter, and a given transmitter affects the receiving neuron in only one of two modes: excitation or inhibition. Excitation stimulates the neuron, ultimately to the point where it generates an action potential that travels down the axon to influence other neurons. Inhibition makes the neuron less likely to generate an action potential in response to excitation from other synapses. Importantly, at all synapses there is a mechanism that rapidly removes the transmitter, terminating its effect within a few milliseconds after it has been released.

With the exception of "electrical" synapses in certain invertebrates and in a very few locations in the mammalian CNS, information is transferred from the axon terminal of one neuron to the cell membrane of another only by a chemical transmitter. This fact, which has been clearly established only for the past 25 years, has extremely important implications. A chemical synapse is, by its very nature, modifiable; other chemicals, hormones, drugs, the local environment of the neuron, and perhaps even its past experience, can influence the synapse.

The chemical synapse has become a unifying concept in neuroscience. Most drug actions on the nervous system occur at synapses. Different types of drugs can have different types and modes of action at the same synapse; one drug might block the release of chemical transmitter, another might occupy the binding site on the receptor in place of the normal chemical transmitter substance, and so on. In a similar vein, various hormones appear to exert their influences on the brain by attaching to specific hormone-binding receptors on the surfaces of certain sets of neurons, particularly in the brain structure termed the hypothalamus, which is closely connected to the pituitary gland.

Important behavioral phenomena like motivation appear to rest on such mechanisms. The experience of "thirst" is due to the action of a "thirst" hormone, activated by the kidney under appropriate conditions, which attaches to and activates certain sets of neurons in the brain. Although adequate evidence is lacking, it seems likely that learning and memory, the bases for behavioral phenomena in humans, rest on persistent changes in chemical synapses in certain regions and systems of the brain.

General Structure of the Brain

The brain is traditionally described as consisting of several major levels. The lowest level—the brain stem—is, evolutionarily, the oldest part of the brain, shared relatively unchanged with primitive vertebrates. It is essentially an upward elaboration of the spinal cord, important to certain specialized senses and to such vital regulatory and reflex functions as breathing, cardiovascular activity, and digestion.

The next major step in brain evolution was the development of the limbic system, about 150 million years ago. It controlled primitive behaviors—anger and attack, fear and flight, hunger, sexual activity—and has been the highest region of the brain for reptiles like the crocodile. The limbic system remains well developed in the human brain, although it is in some ways much changed and may serve rather different functions. However, it is overshadowed by the cerebral cortex, even in lower mammals like the rat. Over the course of evolution of the mammals there has been a steady increase in the relative size of the cerebral cortex. The cortex is a complex neuronal structure about two millimeters thick, that overlies most of the brain, much like the rind on an orange. In higher mammals it becomes increasingly folded and convoluted, so that its total area and extent are much increased. It is the dominant brain structure in primates.

The evolutionary development of the cerebral cortex that separates human from ape began only about 2 million years ago and proceeded at a rate unprecedented in the earlier evolutionary history of the brain. The human cerebral cortex has existed in its present form for only about 100,000 years. And the rational and abstract thought and subtle feelings of the cortex do not always dominate the surges of cruder feelings from the reptilian core, the limbic system, within.

Another way of looking at the organization of the brain is in terms of systems that have been identified by observing the effects of experimental lesions in animals or accidental lesions in humans. The sensory systems convey information to the brain from the various receptors; the motor systems control the muscles (and glands) and hence generate behavior.

For humans, the most important sensory systems are the visual and auditory. To take the visual system as an example, light entering the eye activates receptors in the retina at the back of the eye. This visual information undergoes considerable processing by succeeding layer of neurons in the retina. This processed information is then transmitted to the brain by the optic nerves. Within the brain, the information receives further processing in a portion of a region called the thalamus, which then transmits the information to the visual region of the cerebral cortex, where still further processing is done. Thus, sensory systems extend from the receptors, through various brain levels, to the cerebral cortex.

Motor systems of the brain are those sets of neurons and pathways most directly involved in the control of muscles. Most behaviors that we observe in others are muscle movements, sequences of contractions and relaxations controlled by the motor regions of the brain. Motor neurons, with cell bodies in the spinal cord and brain stem, send their axons out of the CNS to innervate muscles of the body and head. Higher order systems and structures in the brain—ranging from reflex pathways in the spinal cord and brain stem to elaborate, specialized structures like the cerebellum, the basal ganglia, and the motor region of the cerebral cortex—play on these motor neurons. The microstructure of the cerebellum is surprisingly constant from reptile to human. It plays an important but not fully understood role in regulation and coordination of movement. The basal ganglia, large structures lying mostly under the cerebral cortex in the forebrain, are somehow involved in the control of movement—Parkinsonism is a disorder of the basal ganglia—but their functions are very poorly understood.

The limbic system, noted above, is generally believed to be concerned with the motivational and emotional aspects of behavior. It consists of several large and ancient structures—hippocampus, amygdala, septum—closely interconnected with each other and with the hypothalamus, a very ancient structure at the upper end of the brain stem consisting of a number of small groups of nerve cells that exert rather direct and powerful control over hunger and eating, thirst and drinking, sexual behavior, pleasure and pain. The hypothalamus is adjacent to and closely connected with the pituitary gland, the master control gland of the endocrine system.

Another very important brain system is the reticular formation, a collection of groups of nerve cells and fibers that ascends in the core of the brain from the spinal cord up to the level of the thalamus and has close conections to the hypothalamus. The reticular formation appears to play a critical role in the control of sleep and waking. Also, those neurons that utilize noradrenaline and serotonin as transmitters have their cell bodies only in certain regions of the reticular formation.

It must be emphasized that the various brain "systems" are in part useful abstractions. They are categorized as systems largely on the basis of what appear to be close anatomical interconnections or apparent functional similarities. So little is yet known about many regions and structures of the brain that current descriptions of brain systems must be viewed with reservations. For example, the hippocampus is a large structure in the limbic system, whose functions are believed to be motivational and emotional. Yet recent evidence indicates that the hippocampus plays a critically important role in learning and memory in humans and other mammals. Much of the excitement in neuroscience today stems from the fact that so little is yet known—it is a field of science that has only begun.

Some Aspects of Neuroanatomy

Neuroanatomy is in the midst of a revolution, due largely to techniques developed in the past 5–10 years for the tracing of "pathways." Methods for tracing the connections of neurons were difficult, unreliable, and time consuming until the development, in the early 1950's, of a technique based on the fact that, if a cell body or axon in the brain is destroyed, the axon terminals that connect to other neurons will degenerate. At a certain phase of this degeneration, the terminals selectively absorb silver salts, which can then be visualized with a microscope, revealing terminals far removed from the cell body, thus permitting tracing of remote connections.

Recent tracing techniques, which work with undamaged brain, make use of the fact that neurons transport various substances along their axons. Thus, radiolabeled amino acids, injected in the vicinity of cell bodies, are absorbed and transported to the axon terminals, where they can be visualized by appropriate techniques. Conversely, the enzyme peroxidase is taken up by nerve terminals and transported back to the cell bodies, where it can be stained and visualized. Many laboratories are now using these methods to trace out the wiring diagrams of the brain. Not only is new circuitry being plotted; some of the classical connections in the brain are being shown not to exist. More has been learned about the circuitry of the brain in the past 10 years than in all previous history.

A recent technique of great promise depends on the fact that neurons, like all other cells, use glucose as their fuel. A radiolabeled substance (deoxyglucose) that is chemically similar to glucose but not metabolized by neurons is injected. Neurons take up this substance and accumulate it in proportion to their levels of metabolic activity. It is thus possible to determine the effects of various conditions, such as sensory stimulation, on the activity of specific regions and neurons in the brain. Moreover, if the isotope used for labeling the deoxyglucose decays by positron or γ-ray emission, localization in the living, intact brain can be achieved from without.

A method that is just being developed utilizes antibodies to label certain classes of neurons in the brain. As an example, γ-aminobutyric acid, a transmitter operative as an inhibitor at certain synapses, is synthesized by the action of the enzyme glutamic acid decarboxylase (GAD). Application of radioisotopically labeled antibodies specific for GAD permits identification of all neurons that use this transmitter. In principle, analogous procedures can be developed for most of the known transmitters.

Another important approach to analysis of the structure of the nervous system is electron microscopy. The ultrastructure of neurons and synapses is being catalogued at a rapid rate. New methods such as freeze fracture permit ever better visualization of fine structure, which, one day, will enable better understanding of the functional properties of neurons.

FUNCTIONING OF INDIVIDUAL NERVE CELLS

Axonal Conduction

The achievements of neurophysiology in the past 30 years have been monumental. As indicated above, considerable understanding has been gathered concerning the two basic processes of the neuron: axonal conduction and synaptic transmission. These advances depended upon the development of electronic apparatus for amplification, display, and analysis of the tiny electrical signals generated by individual neurons, and upon the use of the microelectrode—a very slender wire that can be placed close to or inside a neuron.

The action potential generated by a neuron can be recorded by inserting a microelectrode so that its conducting tip is very close to a specific neuron. When the neuron fires, the action potential is recorded as a brief (1/1,000 second) wave, a "spike," thus permitting examination of the pattern of activity of that neuron under various circumstances. The method has been applied in studies of the various neuronal sources of aspects of behavior, such as, sensory processes, motor systems, and learning, and will be discussed later in these contexts.

Microelectrodes inserted inside a single neuron allow measurement of the potential across the nerve cell membrane. The normal resting potential across the membrane is about 70 millivolts, with the inside negative. When synaptic excitation occurs, the transmembrane potential of the affected region of the postsynaptic membrane becomes less negative (depolarized) until it reaches the spike discharge threshold required to generate an action potential that sweeps down the axon. In synaptic inhibition the local potential becomes more negative than at rest (hyperpolarized). Hence it is possible to measure the synaptic effects of other neurons on a given neuron.

A major area of current research is concerned with the molecular basis for the events registered as the action potential. Much of current knowledge rests on studies of the giant axon of the squid. The immediate consequence of excitation is a sudden, brief opening of ion channels ("gating") in the membrane so that sodium ions (Na^+)—which otherwise are blocked from diffusing into the neuron—rush in; this is followed by the exit of a lesser number of potassium ions (K^+) through a separate set of channels. This is the mechanism of depolarization. An ion channel appears to be a micropassage through a specific protein molecule that traverses the membrane, opened or closed by a subtle change in the three-dimensional conformation of the protein. It is the transiently lowered potential across one region of membrane that causes opening of the "gates" in the adjacent region.

The toxic principles of the puffer fish (tetrodotoxin), of *Gonyaulax*, the dinoflagellate that causes "red tides" (saxitoxin), and of the scorpion, all of which affect the nervous system by binding to the gating protein on the outside surface of the membrane, have been invaluable in the study of this process. Shortly after the influx of Na^+, a special enzyme, close by in the membrane, utilizes the energy of ATP to pump these ions out, through channels in the pump itself, until the resting potential is restored. At each locus on the axon the entire cycle is complete in about two milliseconds. This pair of phenomena occurs consecutively down the full length of the axon, accounting for the sweep of the action potential. The ATP-dependent sodium pump-protein binds and is inhibited by cardiac glycosides, such as ouabain, accounting for their effects on the conducting system of the heart and elsewhere.

Synaptic Transmission

The immediate events leading to the release of chemical transmitter at the presynaptic axon terminals when the action potential arrives at the terminals are also becoming clearer, as are the events that occur when released transmitter arrives at the postsynaptic membrane receptors and excites or inhibits the next neuron.

The enormous growth of general biochemical understanding in the last generation has had a marked effect on neuroscience. Thirty years ago neurochemistry was largely preoccupied with the structures of unusual lipids that occur in the nervous system. Unraveling the detailed structure of the insulating myelin sheath wrapped around the larger, more rapidly conducting, nerves was a great triumph; multiple sclerosis is the consequence of demyelination of certain nerves. A number of disorders of the nervous system, such as Tay-Sachs disease, involve derangements in the metabolic pathways of these special compounds, and work in this area is continuing. For the past 10 years, however, the main preoccupations of many neurochemists have been the chemistry of synaptic transmission and nerve membranes (Figure 6). The rapid expansion of understanding of these areas has been rewarding; already this has led to major improvements in the treatment of some of the commonest of human ailments, and the prospects for rapid progress are bright.

By the 1950's only two chemical substances had been positively identified as neurotransmitters: acetylcholine and norepinephrine. It now appears that there are probably over 20 different compounds used as transmitters in various parts of the nervous system.

Since chemical transmission was first established by Loewi in 1921, there has been general awareness of a similarity between the actions of hormones and of nerve transmitters. Each system involves the secretion of a special chemical by a specific type of cell, and this compound acts by binding and modifying the properties of a special protein molecule, the receptor, lodged in the membrane of a second, recipient, cell. The main difference between hormone action and synaptic transmission lies in the directness of the action. Most hormones are released into the circulation and travel a great distance before finding a receptor, whereas only a miniscule fraction of a millimeter separates the release site of a transmitter from its receptor; the effect is virtually instantaneous. There is little chance, then, that the transmitter can affect receptors remote from the junction where it is released.

The most thoroughly studied synapses are those employing acetylcholine, thanks to their abundance and easy access in the electric organs of the sting ray and the electric eel. These have revealed that in the presynaptic terminal are collected miniature packets of transmitters bound to a protein and wrapped in a thin membrane (vesicles); entire vesicles are discharged into the synaptic space when the action potential arrives. The vesicles burst, acetylcholine binds to a receptor protein on the postsynaptic membrane and this, somehow, occasions the opening of Na^+ channels, exciting the postsynaptic cell. The effect is fleeting since, in the presynaptic membrane, there is an arrangement that rapidly withdraws acetylcholine back to be repackaged while in the postsynaptic membrane there is an enzyme (acetyl cholinesterase) that can degrade (hydrolyze) all of the acetylcholine in the synaptic space within one–two milliseconds. The Na^+ channels close, excess Na^+ is expelled by the "pump," and the way is then clear for another burst of transmitter.

The entire synaptic arrangement is accessible to chemical agents from the outside. Thus, botulinus toxin specifically prevents the release of transmitter vesicles to the synaptic space; nicotine mimics the action of acetylcholine on the receptor; the toxins of certain snakes such as the cobra and the krait (bungarotoxin) specifically bind to and inactivate the postsynaptic receptor protein; tubocurarine, the active ingredient of curare, acts somewhat like bungarotoxin but only on those acetylcholine receptor proteins where motor nerves synapse on muscles; the drug prostigmine reversibly inhibits acetyl cholinesterase whereas certain "nerve gases," for example, diisopropylfluorophosphate (DIFP), irreversibly inactivate this enzyme. This accessibility offers enormous potential for the development of drugs that can modify or control synaptic function. Ironically, one disease, myasthenia gravis, appears to be the consequence of

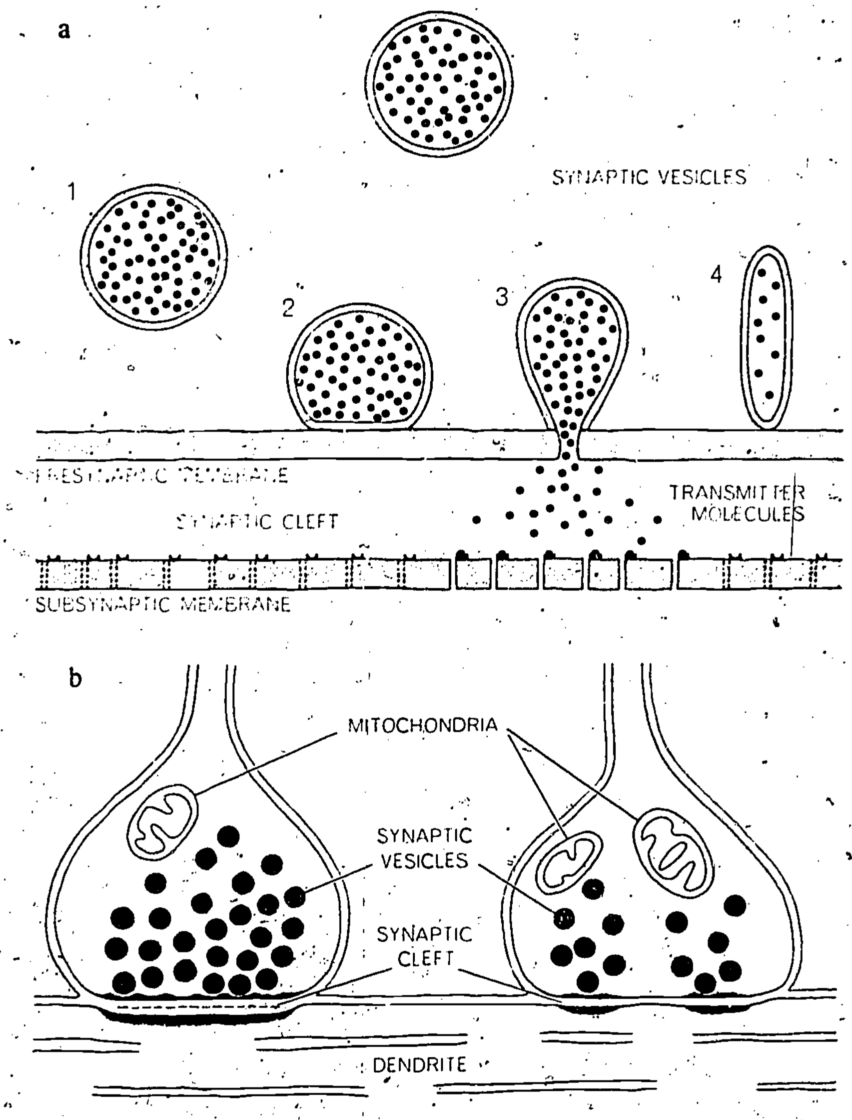


FIGURE 6 (a) Nerve impulses and chemical transmitters. Cued by a nerve impulse, synaptic knobs deliver short bursts of a chemical transmitter into the synaptic cleft. Mitochondria supply the cells with energy. (b) The transmitters, which are stored in synaptic vesicles, require only a few microseconds to diffuse across the synaptic cleft and to attach to specific receptor sites on the surface of the adjacent nerve cell. The probable pattern of movement, as shown here, is that synaptic vesicles move to the cleft, discharge their contents, and then return to the interior for recharging. (From "The Synapse," by Sir John Eccles, © January 1965, Scientific American, Inc.)

formation of antibodies against one's own acetylcholine receptor protein.

It is not clear what advantages accrue from the fact that there are more than one excitatory and one inhibitory transmitter substance with each synapse specific for only one transmitter. Perhaps it is but a relic of independent evolutionary history of different systems within the brain.

At least three transmitters—norepinephrine, dopamine, and serotonin—function at their synapses in a manner similar to the action of certain hormones elsewhere in the body, in that their binding to their postsynaptic receptors activates an enzyme that synthesizes a substance called cyclic adenylate (CAMP) and the latter, in some manner, initiates the membrane depolarization that starts conduction of the impulse down the axon. And again, at each such synapse there is a very efficient mechanism for rapid removal of the specific transmitter. Agents are known that specifically affect various components of these synaptic mechanisms, for instance, cocaine and reserpine block the mechanism that reabsorbs norepinephrine, theophylline prevents degradation of CAMP, and lysergic diethylamide (LSD) prevents binding of serotonin to its receptor.

Perhaps half of all synapses are inhibitory rather than excitatory; binding of their transmitters to the receptor appears to occasion inflow of chloride ion (Cl^-) rather than Na^+, thereby hyperpolarizing the postsynaptic membrane. γ-Aminobutyrate was the first such transmitter to be recognized; the simple amino acid glycine appears to so serve at most inhibitory synapses in the brain. The convulsion-inducing drugs picrotoxin and strychnine, as well as bee venom, block binding of γ-aminobutyrate and glycine to their respective receptors.

Polypeptides of the Brain

A huge new chapter in understanding the brain has barely been opened. It has been known for some years that a series of "releasing factors" flows from the hypothalamus to the pituitary, each specifically signalling the latter to release one of its specific hormones. The short connecting vein may be regarded, strategically, as an exaggerated synapse conveying transmitters. The latter, the releasing factors, have been identified as a set of small polypeptides. Surprisingly, these polypeptides, such as somatostatin and thyrotropin releasing factor, have been found widely distributed in the higher levels of the brain, including at certain synapses. So too is a family of other polypeptides, each of which was also first recognized in some other connection, for example, the hormones gastrin and cholecystokinin, and other less familiar peptides. In some areas, they seem to serve as synaptic transmitters, while in other regions, a given polypeptide may merely be released by one or more nerve cells, free to diffuse and slowly influence dozens or hundreds of cells in their surroundings—a mechanism appropriate to mood rather than perception or action. Thus, information transfer in the brain occurs only in part through tightly wired connections. It seems likely that there are many more functional polypeptides than have yet been recognized, evoking an image of the brain as a unique endocrine organ controlled by its own secretions.

Thus the still incomplete but greatly improved understanding of the chemistry of synapses has had profound effects on pharmacology. As we have seen, many drugs, old and new, have effects at one or another stage in these sequences of events involving nerve transmission. Some prevent release by the first nerve's terminals, some mimic the transmitter, some actually *are* transmitters, others prevent the removal or destruction of transmitter, still others are involved in the metabolism of cAMP reactions, or block receptor actions. Knowing roughly how nerve-to-nerve junctions work, even when the knowledge is incomplete, makes it much easier to invent new drugs and predict their actions, and to use older drugs more effectively. Even the unwanted side effects of new drugs have occasionally had a great impact on understanding of disease.

The current treatment for Parkinson's disease offers a good example. Parkinsonism, the cause of which is still in doubt, and which is characterized by tremors and stiffness of the limbs and difficulty in getting movements started, originates in the brain structures termed the basal ganglia. A few years ago, high concentrations of the transmitter dopamine were found in normal basal ganglia. But, in Parkinson's disease, in which these structures are visibly abnormal, dopamine levels were much below normal, suggesting that administration of dopamine to Parkinson patients might relieve their symptoms. Whereas dopamine does not cross certain barrier membranes that separate the brain from the blood stream, its metabolic precursor, dioxyphenylalanine (L-dopa), can cross these barriers. When L-dopa was given to Parkinsonian patients it was found to be amazingly effective in relieving the symptoms of many, an elegant example of the use of basic scientific knowledge in designing rational therapy.

NEURONAL BASES OF BEHAVIOR

The central concern of this branch of neuroscience is to understand what happens in the brain and spinal cord when an organism perceives, thinks, learns, remembers, feels emotion, speaks, is motivated, and acts. Most neuroscientists take it for granted that to understand the mind means to understand the brain, and that it is within the capabilities of humans to do so, one day. Progress in the field in the last 25 years has certainly not discouraged this view.

Work on the structure, functioning, and chemistry of neurons and synapses has provided a powerful armamentarium of methods and techniques to tackle the neuronal mechanisms underlying basic aspects of behavior and experience. To take only one example, by recording the action potentials of a neuron in the visual part of the brain, it is possible to determine the manner in which nerve cells represent or "code" various aspects of visual stimuli—the neuronal basis of sensation and perceptual experience. Recording the action potentials of individual neurons is being used successfully in the analysis of motor systems of the brain, and of neuronal activity associated with motivation, sleep, learning, and memory.

Genetic and Developmental Roots

The extent to which behavior and even brain structure is due to genes or to environment (the "nature–nurture" issue) is no longer viewed as either-or but rather as a continual process of interaction. To understand how the brain develops into its adult form is a stupendous problem. Brain development can usefully be separated into development before birth, which is largely programmed genetically, and after birth, when both genetic and environmental influences are at work. Regrowth of the nervous system after injury involves similar problems.

A central problem in development is the way in which the genome is translated into an organism. For the nervous system, this involves such matters as the cell-to-cell connections within various structures, mutual spatial relationships among cells and structures, and their connections with distant structures. For example, consider the connections between the eye and the brain. The optic nerve contains about 1 million fibers that originate from the retina; each fiber is connected to a tiny part of the retina and any one part of the retina is represented by a small number of fibers. These all connect to a platelike layer of brain cells in a systematic way. Each region of retina connects to a particular region on this plate; the retina can be regarded as being "mapped" onto the plate. During development the optic nerve fibers grow out of the retina, reach the plate, and distribute themselves with absolute topographic precision. How each fiber finds the right destination is not known.

Other specifically wired sets of cablelike connections between topographically mapped areas are common in the nervous system. How such precise wiring is laid down remains a great unsolved problem that is the focus of intense research activity, frequently involving experiments in which a nerve is cut and allowed to regrow. One can rotate the target tissue, such as an eye, remove half of it, or remove half the source of the growing nerve cable, and see whether the fibers regrow properly. In amphibia this occurs with high precision. But in a number of comparable clinical circumstances, nerve regrowth is chaotic. For example, if the nerve growing to a muscle in the hand is severed, regrown fibers rarely find their proper targets, and the result may be almost as bad as paralysis.

The prospects are reasonably good that in a decade or so we will have learned how nerves seek their targets. More pointedly, there might be gained the insight so badly needed to give direction to attempts to stimulate the regrowth and reconnection of a severed peripheral nerve trunk or spinal cord.

An important area of studies concerns the development of nerve–nerve and nerve–muscle junctions. When a developing nerve reaches a muscle, it induces changes in the muscle membrane it innervates. Synapses appear and the transmitter substance at these junctions sometimes changes as the structure matures.

A promising tool for studying development is the use of genetic mutations. Mutant mice have already improved understanding of a few human neurological diseases, including retinal dystrophy (a cause of blindness) and various cerebellar diseases. Work with mutant strains of fruit flies offers great promise of illuminating the genetic control of neural development, and studies of nerve tumor cells in culture may shed light on the basic aspects of neuron development.

The potential importance of developmental studies is great, not only for what they reveal of how the brain works, but because many neurological conditions seem likely to be developmental. These include most birth defects, Down's syndrome, muscular dystrophy, common epilepsies, and a number of rarer diseases.

Sensory Processes—The Neural Basis of Sensation and Perception

Recent progress in the analysis of neuronal aspects of sensation—vision, hearing, touch, pain, smell, and taste—has been rapid. Sensation is probably the most active and highly populated area of neuroscience. This is due, in part, to the fact that one can study sensory systems in anesthetized animals. The main strategy has been to record from single cells using microelectrodes in the appropriate part of the brain while stimulating with sound, light, or other stimuli. The specific problem with respect to the visual part of the brain is how the brain interprets the messages it receives from the light receptive cells of the retina. A fairly satisfactory knowledge has been gathered of the molecular events whereby absorption of light by the pigment in the retinal cells engenders an action potential or "message." These messages have been followed into the brain for some six or seven stages, including the part of the cerebral cortex concerned with vision. The retina is now one of the best understood structures in the nervous system. Incidentally, color blindness, an inherited condition that affects a significant percentage of the male population, results from absence of one of the three proteins used to make the light-sensitive pigments in the cells of the cones of the retina.

A typical cell in the visual cortex responds actively (generating action potentials in rapid succession) when light shines on the retina, but only when the pattern is exactly the right one for the cell. For example, a cell might respond to a tiny line shaped out of light shining in just the right part of the retina and tilted in just the appropriate way. Such a cell would not react to a line anywhere else in the retina or with a different tilt. Since there is a huge area of the visual world where a line may fall, and tilts include every angle around the circle, this obviously requires that there be a huge number of cells in the visual cortex.

This brings us only a little way toward knowing how we see, but it approaches that goal more closely than could have been imagined 20 years ago. At least one region of cerebral cortex—the size of a few postage stamps as compared with roughly 1 1/2 square feet of human cortex—can be understood and in rather simple terms. Of course there is far more to be learned about the visual cortex, and other areas may not be as easy to study. Nevertheless, in principle, the cerebral cortex is capable of being understood, and these results greatly encourage attempts to understand other areas.

An unexpected by-product of this new knowledge about the visual cortex was the possibility of investigating what goes wrong when a young child becomes cross-eyed or develops cataract. Experiments in animals showed that, in these cases, the visual cortex soon loses its correct connections. As a result, such children now have corrective surgery as early as possible.

The properties of the individual neural elements of the visual cortex have yet to explain sensation, and the higher-order influences of the visual neurons on "association" areas of the cortex, processes that may underlie perception, remain unknown. Hopefully, as understanding of sensory systems increases—and it seems likely that the auditory system will come to be understood as well in the next 5–10 years as the visual system is now—approaches to effective prostheses for the devastating disorders of blindness and deafness may emerge. Indeed, there are promising indications that, in certain forms of deafness, differential electrical stimulation of auditory nerve fibers may produce "hearing," even understanding of speech.

Motor Systems—The Neural Basis of Movement

That a particular region of the cerebral cortex produces movement when stimulated electrically was demonstrated just over a hundred years ago. Long before that, however, physicians knew that injury to the brain produced paralysis or difficulties in movement. In recent years, the neural correlates of movement have been widely studied, using a variety of anatomical, physiological, and chemical techniques.

From these investigations has emerged the concept of a motor system or systems that control an organism's movements. Among the structures of the motor system, the ventral region of the spinal cord, the cerebellum, basal ganglia, and motor regions of the cerebral cortex are the most important. Although precise knowledge of how specific components of the systems function together to produce a movement or series of movements is not yet available, new procedures for studying electrical and chemical activities of individual nerve cells in awake and moving trained animals show promise for solving this problem. As an example, it has been found that certain classes of neurons in the motor area of the cerebral cortex "code" the force of a movement rather than its extent.

In the next few years, understanding of how the different components of the motor systems function together to generate and control movements, as well as their failure in certain neurological diseases, should slowly develop.

Sleep

Why do we sleep? The most promising theories rest on new knowledge about brain structures and neurotransmitters. In the 1950's it was discovered that there are two distinct kinds of sleep: "slow wave" sleep and "rapid-eye-movement" (REM) sleep. A person (and all higher animals) goes from one to the other of these states of sleep four or five times a night, spending about 20 minutes to 1 hour in each. There is considerable, although disputed, evidence that REM sleep is associated with dreaming. Much recent research has sought to identify regions of the brain specifically involved in these two phases of sleep. Two small clusters of nerve cell bodies in the brain stem, the locus coeruleus (which utilizes norepinephrine) and the raphe nuclei (which utilizes serotonin), are the current focus of such studies.

Some years ago it seemed that why we sleep would be known by now. The simplest hypothesis—that a sleep "substance" exists—has been investigated intensively. None has yet been found. Discovery of a "sleep-inducing factor" might account for how the brain falls asleep or awakens; but it might still not clarify the physiological "purpose" of the sleeping state. Progress in this field is likely to be slow, while the many brain systems that show changes during the two phases of sleep are analyzed in detail.

Motivation and Emotion

Knowledge of the biological bases of such elemental motives as thirst, hunger, and sex has grown rapidly. In the early 1950's the hypothalamus was implicated as the most importantly involved brain structure. To take hunger as an example, damage to one region of the hypothalamus eliminates eating behavior in animals; they starve to death if not force-fed. Damage to an adjacent region produces a voracious animal that becomes obese. Such hypothalamically damaged obese animals, like obese humans, are much influenced by taste of food, and will not overeat if they have to work hard to get food.

It has been known for some time, of course, that sexual behavior is strongly dependent on the endocrine system. Hormones from the ovaries and testes are reciprocally controlled by hormones secreted by the pituitary gland, itself controlled in part by the hypothalamus. The gonadal hormones act directly on neurons in the hypothalamus and other brain regions that, in turn, control further the actions of the pituitary gland. It is difficult to overstate the importance of the various sex hormones in the biological development of organisms and in sexual behavior. Study of such neural–hormonal interactions is growing at a rapid pace.

Another example of hormonal influence on behavior is thirst. When cellular dehydration occurs, the kidney releases an enzyme, renin. Renin cleaves a piece from a blood protein, ultimately to yield the polypeptide angiotensin II, which is, in effect, the thirst hormone. It appears to have direct actions on the hypothalamus and other brain regions that result in an intense feeling of thirst.

The discovery in 1953 of reward centers in the brain was one of the most important findings yet made concerning behavior. When an electrode was implanted in a region of the hypothalamus and connected to a lever so that a rat could deliver weak electrical stimuli to its own hypothalamus, the animal would press endlessly. The same phenomenon occurs in other species, including humans. Some starving rats, given a choice between such electrical self-stimulation and food, will self-stimulate until they die of hunger. This behavior is reminiscent of severe drug addiction. The brain reward region overlaps the areas of the hypothalamus concerned with drinking, eating, and sex, but seems to be more general.

Other regions of the upper brain stem, some closely adjacent to the reward area, appear to yield pain when stimulated. Strongly addicting drugs like morphine, which seem to exert their actions primarily on these regions of the brain, both alleviate pain and produce strong feelings of pleasure. When it was found that many neurons in the pain (midbrain and thalamus) and pleasure (amygdala) regions of the brain have receptors that tightly bind morphine and related substances, the question arose as to the nature of the material that normally binds at these opiate receptors. The answer has proved to be that regions of the brain itself, particularly the hypothalamus, make and release peptides that bind to the opiate receptors even more effectively than morphine. These substances, called endorphins, very effectively relieve pain, induce a sensation of pleasure, and are strongly addicting. They are the brain's own natural opiates. The possible implications of this discovery, both for relieving pain and for understanding and combatting drug addiction, are considerable. Whether endorphins function at synapses or on other cell surface receptors is not clear.

This area, in which several different fields are just coming together, is one of the most rapidly moving and promising fields of science. It seems quite possible that the underlying causes of a range of human disorders from obesity to drug addiction will, at the least, be much better understood in the next 5–10 years.

Learning and Memory

Language, science, and society endure only because each individual learns anew. There is a general relationship between the evolutionary status of animals and their ability to learn. *Homo sapiens* differs from other primates because of what humans can learn. Apes learn much more than rats, which, in turn, learn a great deal more than flies.

The problem of analyzing the neuronal bases of learning and memory is formidable. It has been surmised that a well-educated adult human has more bits of learned information stored in memory than there are neurons in the brain. This fact alone argues that memory must somehow be coded in the synaptic interconnections among neurons, where the possible patterns are virtually limitless.

Fortunately, basic processes of learning seem to be quite similar over a wide range of animals. Habituation, a simple decrease in the strength of a reflex response as a result of repeated stimulation, has essentially the same properties in primitive invertebrates—reflexes of the neurally isolated spinal cord—and in intact humans. Pavlovian conditioning and simple instrumental learning seem to be similar in mammals from rat to human. Consequently, animal models may be used to study these simpler basic processes of learning.

Considerable progress has been made in the analysis of the neural mechanisms underlying habituation. In simple reflexes of certain invertebrates and in the isolated spinal cord, where it has been possible to obtain habituation across a single set of synapses, it seems due to a decrease in the probability of transmitter release in response to a given stimulus. While a very simple form of behavioral plasticity or learning, habituation has adaptive value. It has been suggested that habituation is the primary mechanism whereby humans adapt to unpleasant but unavoidable situations like urban stress.

More important in human behavior is associative learning, ranging from simple conditioning to language. Learning is thought to occur in two phases in both animals and humans. When information is first learned it is fragile; it can be disrupted, for example, by electroconvulsive shock. However, after it has been well-learned for a period of minutes to hours it is stable and not subject to disruption. Short disruptable neuronal processes are converted to more permanent, long-term storage mechanisms. In animals the two processes can be prevented by quite different drugs, one of which blocks protein synthesis in the ribosome, while the other prevents RNA synthesis on the DNA template.

The neural bases of associative learning are being sought, using animal models, by recording the action potentials of individual neurons in various brain structures during the learning and retrieval (i.e., memory) of information. This search for the "engram" (a term used to refer to the neuronal coding of memory) entails identification of the neurons or systems of neurons where activity changes as a result of learning, and the underlying synaptic mechanisms of storage.

A structure of the limbic system, the hippocampus, seems to play an important role in the memory system in both humans and infrahuman animals. Certain neurons in the hippocampus begin to increase their rates of activity from the very beginning of training in animals. Their activity continues to grow and is projected to other brain structures. The hippocampus is not, per se, the locus of long-term memory storage—humans with severe damage to the hippocampus have intact memory for information learned before damage but have great difficulty in placing new information into storage or retrieving it.

There are tantalizing hints from neuroanatomical studies that either new synapses may form or existing synapses become enlarged in the hippocampus and cerebral cortex as a result of learning experiences. One neurotransmitter, norepinephrine, seems most important in the learning process. Animals learn best when optimal levels of this neurotransmitter are in the brain. ACTH, the stress hormone, exerts a modulatory action on learning. Certain drugs can actually facilitate learning and memory storage. The relations of these effects to synaptic transmission is under current investigation.

Cognition—Language and Consciousness

The neuronal bases of higher mental processes in humans is a field for the future. However, some significant advances have been made in the past 15 years. Information has come primarily from study of patients with brain damage due either to injury or as a result of necessary neurosurgical procedures. It has been known for some time that language is represented in the left hemisphere of the brain in virtually all right-handed individuals. Two regions of the cerebral cortex seem to be specialized for language—an anterior area involved more in the expression of language and a posterior area involved in the understanding of language.

Studies in the past few years indicate that consciousness is also differentially represented in the two hemispheres. Because of the way the eyes project to the brain, visual information can be addressed separately to each of the two hemispheres in persons in whom the large bands of fibers interconnecting the two hemispheres had been surgically severed to prevent epileptic seizures. In essence, left hemisphere consciousness is verbal and right hemisphere consciousness is nonverbal. Both seem to exist more or less independently in these patients.

This is a very young field, and unless expected new technological developments occur, progress is likely to be slow. Thus, methods are needed to record the activity of single nerve cells and to safely stimulate small regions of the brain without drilling holes in the skull. Positron emission techniques can now give a picture of brain activity with about centimeter resolution. Much will be learned about human brain function as this method is applied, even though it will not permit recording the activity of single neurons.

Abnormalities of Behavior and Experience—The Psychoses

Extraordinary progress has been made in the past 30 years in treatment of the psychoses. Most neuroscientists and psychiatrists agree that schizophrenia, the most severe and widespread psychosis, reflects some fundamentally abnormal brain function. This is consonant with the fact that the closer a person's hereditary relationship to a schizophrenic, the greater the likelihood that person will develop the disorder. At this time, the primary defect in schizophrenia is not known.

Research on schizophrenia in the past few years has strongly implicated one neurotransmitter—dopamine. L-dopa, used for the treatment of Parkinsonism, elicits symptoms rather like schizophrenia; chlorpromazine and other antipsychotic drugs that relieve schizophrenia elicit the symptoms of Parkinsonism. These observations are explained by the fact that the antipsychotic drugs block the action of dopamine at dopamine receptors—suggesting that schizophrenia results from overactivity of some set of nerves that utilize dopamine as transmitter.

Whatever the ultimate cause of schizophrenia, it is, thus, unlikely that the schizophrenic brain is, in the usual sense, physically entirely normal. It is hard to exaggerate the importance of accepting that concept if there is ever to be a real cure or prevention.

Changes in scientists' attitudes toward the psychoses, partly as a result of genetic studies and partly because of the advances in neurochemistry and pharmacology, will have important effects on psychiatry and psychology, even on attitudes toward the neuroses, the lesser behavioral disorders.

BIOLOGY AND AGRICULTURE

In the past two decades advances in plant science have marched in step with the progress of the other sciences. Molecular genetics has conferred a more penetrating understanding of plant variation, and promises new approaches to plant breeding; cell biology has demonstrated how nutrients move across membranes; photochemistry has clarified the primary events of photosynthesis; biochemistry has elucidated the subsequent events of photosynthesis, the process of atmospheric nitrogen fixation and various factors that regulate plant growth. Application of this newly discovered fundamental knowledge holds exciting prospects for the future.

PHOTOSYNTHESIS

Life is sustained on this planet by photosynthesis, through which green plants capture the sun's radiant energy. In ordinary combustion, oxygen from the air reacts with organic compounds, e.g., a carbohydrate such as cellulose, to form carbon dioxide (CO_2) and water (H_2O)—releasing energy as heat, as in the flames of a wood fire.

$$\text{(A)}\quad C_6H_{12}O_6 + 6O_2 \longrightarrow 6\,CO_2 + 6\,H_2O + \text{energy}$$

In living cells the controlled oxidation of glucose by oxygen releases the same amount of energy, but in a chemical form suitable to power the processes of life.

Photosynthesis is, in overall effect, the reverse of this process. Plants convert $CO_2 + H_2O$ into carbohydrate, obtaining the necessary energy from sunlight, and liberating oxygen in the process. This occurs within a specialized subcellular organelle in leaf cells called the chloroplast.

In fact, there is no direct reaction between CO_2 and H_2O. Rather, the overall process can be viewed as a set of partial reactions. To begin with, in the specific process that utilizes solar energy, electrons are withdrawn from H_2O and transferred to a carrier substance, the special nucleotide denoted as NADP:

$$(B_1)\quad 2\,H_2O + 2\,\text{NADP} \longrightarrow O_2 + 2\,\text{NADPH}_2$$

Reaction B_1 should spontaneously flow from right to left; however, within a chloroplast light energy is used to drive the system to the right. In subsequent processes, the NADPH_2 is utilized to convert CO_2 to glucose:

$$(B_2)\quad 6\,CO_2 + 12\,\text{NADPH}_2 \longrightarrow C_6H_{12}O_6 + 12\,\text{NADP}$$

Let us first consider how light makes possible reaction B_1.

Chlorophyll Photocenters

Absorption of light energy by chlorophyll is the critical event in photosynthesis. But how is that energy transduced into a form of chemical energy that the plant cell can utilize? Energy from absorbed light migrates to two types of photochemical centers. At each photocenter the functional chlorophyll molecule exists as a sandwich between two other molecules. Chemically they are quite different, and current research is focused on identifying them and understanding the precise three-dimensional arrangements between them. For the present purpose, however, it suffices to denote them A and B, respectively, at one type of photocenter and C and D, respectively, at the second type.

When the chlorophyll absorbs a photon (usually of red light, which is why plants appear green), the electrons in its resonating structure are activated to such an extent that one is ejected, leaving behind a positively charged chlorophyll molecule (Chl^+, a free radical). The ejected energetic electron is instantly accepted by the adjacent molecule of $B°$. But A then donates an electron to the chlorophyll radical, which returns to its original state:

$$A°\text{Chl}°B° \longrightarrow A°\text{Chl}^+B^- \longrightarrow A^+\text{Chl}°B^-$$

If A^+ and B^- were in contact, they would react to reform $A°$ and $B°$. But the large, greasy molecule of chlorophyll keeps them apart; hence the process is termed "charge separation." This process is complete in a few picoseconds (10^{-12} seconds). Analogous events occur at the second photocenter.

There follows then the process, summarized above, wherein, by drawing electrons from H_2O, O_2 is formed and A^+ is restored to $A°$. At the same time the excess electron in B flows to the second photocenter and then on to NADP to form NADPH_2. Schematically the flow of electrons in the entire arrangement can be depicted as

$$H_2O \longrightarrow e \longrightarrow A\text{Chl}B \longrightarrow e \longrightarrow C\text{Chl}D$$
$$\searrow O_2$$
$$\longrightarrow e \longrightarrow \text{NADPH}_2$$
$$\text{NADP}$$

so that the tiny electrical current flows only when the chlorophyll at both photocenters is being illuminated.

The $NADPH_2$ that is produced is used in the conversion of CO_2 to carbohydrate, but the reducing power (chemical potential) of $NADPH_2$, alone, is insufficient to the task. A second form of chemical energy, that in the structure of adenosine triphospate (ATP), is also required. Like $NADPH_2$, ATP is made by the chloroplast photochemical apparatus, from adenosine diphosphate (ADP) and inorganic phosphate (P).

$$\text{(C)}\quad \text{ADP} + \text{P} \xrightarrow{\text{light energy}} \text{ATP}$$

Knowledge of how this happens has been hard gained. The photochemical center chlorophyll molecules and all of the other intermediate electron carriers are imbedded in the chloroplast membrane in fixed physical relationships. As electrons move via consecutive carriers between the two chlorophyll centers, protons (H^+) are discharged across a membrane to the interior, building up a strong electrochemical potential. This drives the protons back through a channel in a complex enzyme, lodged in the same membrane, providing the immediate energy for reaction (C).

Formation of $NADPH_2$ and of ATP is the sole role of the light-using apparatus. All other events in photosynthesis can happen in the dark, given a supply of $NADPH_2$ and ATP.

The overall accomplishment of the photochemical apparatus of the chloroplast may usefully be summarized as:

$$\text{(D)}\quad 12\ H_2O + 18\ \text{ADP} + 18\ \text{P} + 12\ \text{NADP} \xrightarrow{\text{light energy}} 6\ O_2 + 18\ \text{ATP} + 12\ \text{NADPH}_2$$

In the subsequent "dark reactions" the reducing power of $NADPH_2$, together with energy suplied by ATP, is used to reduce CO_2 to carbohydrate:

$$\text{(E)}\quad 6\ CO_2 + 18\ \text{ATP} + 12\ \text{NADPH}_2 \longrightarrow C_6H_{12}O_6\ (\text{sugar}) + 12\ \text{NADP} + 18\ \text{ADP} + 18\ \text{P} + 6\ H_2O$$

CO_2 Fixation

The conversion of CO_2 into carbohydrate (process E) is accomplished by a series of reactions requiring more than a dozen distinct enzymes. But CO_2 itself participates only in one chemical reaction, that in which a five-carbon compound, ribulose bisphosphate, reacts with CO_2 and then splits into two molecules of the three-carbon compound, 3-phosphoglyceric acid.

Using the supply of ATP and $NADPH_2$ from the photochemical process, the system operates in such fashion that, overall

$$6\ \text{Ribulose bisphosphate} + 6\ CO_2 \longrightarrow 12\ \text{3-phosphoglyceric acid} \longrightarrow 6\ \text{Ribulose bisphosphate} + C_6H_{12}O_6\ (\text{glucose})$$

Glucose thus accumulates, and ribulose bisphosphate is regenerated, in a combination of reactions sometimes referred to as the C_3 cycle (Figure 7a).

Improving the Efficiency of Photosynthesis

Considerable research is aimed at increasing the efficiency of photosynthesis in the intact plant.

The simplest approach involves altering plant structure. Usually full sunlight provides more light than necessary to saturate the photosynthetic systems in the upper leaves, but not those below. By selecting plant strains with a more upright alignment of the leaves, more light penetrates to the light-limited lower leaves. This approach has yielded a number of hybrid maize varieties that are widely used in the United States. A similar strategy should be applied in the future to other plants.

Another approach is to breed plants for photosynthetic units that operate more efficiently, that is, convert more of their incident light into chemical energy. The most efficient crop plants, such as maize, have an efficiency of only about 5 percent during their period of most rapid growth, when their leaf canopy is full and light capture is nearly complete. In contrast, some algae have an efficiency of 25 percent under optimal conditions in the laboratory. Although there have been some hints that the photosynthetic efficiency of plants can be improved, it should be appreciated that the light reactions in plant photosynthesis are already unusually efficient when compared with most other photochemical reactions.

Another research aim is to improve energy utilization in the dark reactions of photosynthesis. The key enzyme, ribulose bisphosphate carboxylase, is not very efficient at its job compared with other enzymes; it has both a low affinity for CO_2 and a slow turnover time. The plant manages to maintain a reasonable rate of photosynthesis only because it contains large quantities of the enzyme. (In fact, this is the most abundant enzyme known; it may constitute 50 percent of the soluble protein in chloroplasts.)

The main problem, however, is that this enzyme not only fixes CO_2 but is also quick to catalyze a wasteful reaction, called photorespiration, between ribulose bisphosphate and oxygen.

$$\text{Ribulose bisphosphate} + O_2 \longrightarrow \text{3-phosphoglyceric acid} + \text{phosphoglycolic acid}$$

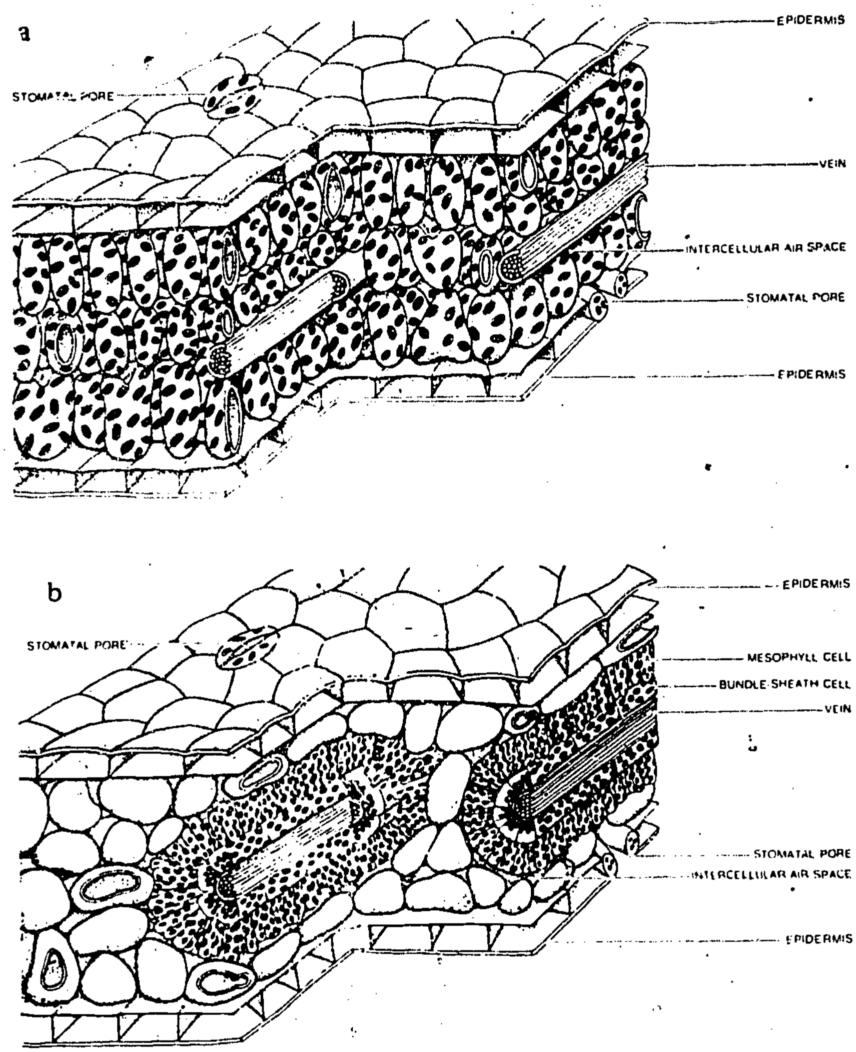


FIGURE 7 Different leaf structures for plants using different photosynthetic pathways. Both plants are varieties of the saltbush, but one uses a three-carbon pathway, the other a four-carbon one. In the three-carbon variety, *A triplex patula* (a), chlorophyll-containing chloroplasts, shown as black dots, are distributed throughout the plant: but in the four-carbon variety, *A triplex rosea* (b), they are concentrated in two types of cells that form concentric layers around the fine veins of the leaf. The structure of the latter plant is designed to provide a net transport of carbon dioxide from the outer cell array to the inner one, producing a higher concentration of carbon dioxide at the inner cells and enabling a faster diffusion of carbon dioxide into the leaf, and therefore a higher rate of photosynthesis. The result is that *A. rosea* grows better than *A. patula* in dry, hot weather. (From "High-Efficiency Photosynthesis," by Olle Björkman and Joseph Berry, © October 1973, Scientific American, Inc.)

Since the phosphoglycolic acid is then further oxidized through photorespiration to CO_2 and H_2O in an uncoupled reaction that generates no ATP, photorespiration dissipates energy and depletes the supply of ribulose bisphosphate needed for photosynthesis.

There is a small chance that one might be able to enhance the affinity of ribulose bisphosphate carboxylase for carbon dioxide, or otherwise avert the waste of photorespiration. To do so, one must better understand the enzyme's intimate three-dimensional structure. However, the protein is large and complex, presenting a formidable task for crystallographic analysis. The challenge remains, nonetheless, and undoubtedly there will be attempts to alter and improve this critical enzyme, however small the chances of success. The carboxylase from photosynthetic bacteria, which has a lower molecular weight, may be more amenable to X-ray analysis.

C_4 Cycle

To carry on photosynthesis, a plant needs CO_2 from the atmosphere. But the low concentration of CO_2 (about 0.03 percent in air) limits the rate of photosynthesis. A special arrangement that mitigates this problem is present in such plants as sugarcane, maize, sorghum, and crabgrass. In certain leaf chloroplasts of these plants, an appropriate set of enzymes catalyzes the addition of CO_2 to a three-carbon compound to form a four-carbon acid. This four-carbon acid then migrates to other chloroplasts deeper in the interior of the leaf, where it releases CO_2 for use in the C_3 cycle.

By furnishing the C_3 system with additional CO_2, the C_4 system favors CO_2 as it competes with O_2 for the enzyme ribulose bisphosphate carboxylase, and thus minimizes the waste of energy through photorespiration (Figure 7b). Although the C_4 + C_3 system requires more ATP than the C_3 system alone, the double advantages of operating effectively at a relatively low concentration of CO_2 and of decreasing the losses from photorespiration more than compensate the C_4 plants for this cost.

Accordingly, there have been attempts to convert other plants to C_4 type metabolism. This seems a reasonable goal, for both systems occur in nature, and intermediate forms between the pure C_3 plants and C_4 + C_3 plants have been discovered.

In addition, specific inhibitors that reduce the rate of photorespiration could, potentially, increase photosynthetic efficiency and utilize available energy more efficiently. To date, such attempts have not been successful, but they will be continued because of the tremendous importance of operating the photosynthetic process at the highest possible efficiency.

NITROGEN FIXATION

The importance of biological nitrogen fixation to contemporary agriculture may be second only to photosynthesis. All forms of life need nitrogen for proteins and nucleic acid, and the nitrogen supply is probably the commonest factor limiting plant growth. Although nitrogen gas (N_2) is abundant in the atmosphere, it can be used by biological systems only if it has been "fixed": the bond between the two atoms of nitrogen is broken, and each of the two atoms binds three hydrogen atoms to form ammonia (NH_3), which is then used for diverse metabolic purposes.

Nitrogen can be fixed either chemically or biologically. The industrial Haber-Bosch process utilizes high temperatures and high pressures to make ammonia from gaseous hydrogen and nitrogen. Chemical nitrogen fixation is a major industry, annually producing nearly 40 million metric tons of fixed nitrogen to be used as fertilizer. Unfortunately, this process requires a great amount of energy, most of which now comes from natural gas. It is possible, though perhaps not now economically feasible, to substitute coal or other fuels for the natural gas. The Tennessee Valley Authority is installing a coal–gas pilot plant that will begin to test this approach in the 1980's. As supplies of fossil fuels dwindle, the price of chemically fixed nitrogen will rise. These concerns have spurred a recent surge of interest in biological nitrogen fixation.

In nature, N_2 fixation can be accomplished only by microorganisms: certain bacteria that live freely in the soil, blue-green algae, and the rhizobia—bacteria that infiltrate, and form nodules on, the roots of leguminous crops. Here, too, the product is ammonia, and here, too, the energy cost is great. In contrast to the industrial process, however, the biological system operates through the nitrogenase enzyme system, at normal temperatures and with the pressure of nitrogen less than one atmosphere.

The nitrogenase system contains two proteins, dinitrogenase (a molybdenum–iron protein) and dinitrogenase reductase (an iron protein). These proteins are readily inactivated by oxygen, from which they are protected by special mechanisms in N_2-fixing microorganisms.

Driving the nitrogenase reaction requires much ATP (about 12 ATP per N_2) as well as a strong reducing agent such as ferredoxin (an iron–protein complex).

$$N_2 + 6 \text{ electrons} + 12\ ATP + 6\ H^+ \longrightarrow 2\ NH_3 + 12\ ADP + 12\ P$$

The ferredoxin, in turn, must be reduced by the metabolism of the cell. Reduced ferredoxin transfers electrons to the dinitrogenase reductase. The latter uses energy from ATP to make it a strong enough reductant to reduce the dinitrogenase which, in turn, reduces N_2 to NH_3. Even in the presence of N_2, however, the system wastefully reduces H^+ to H_2 gas.

Nitrogenase can reduce not only nitrogen, but also protons, acetylene, and several other substances. The reduction of acetylene to ethylene furnishes an easy and sensitive test for nitrogenase, widely used in the laboratory and in field testing nitrogen fixation.

Structure of Nitrogenase

There is keen interest in the structures of the complex components of the nitrogenase system. A promising new development is the isolation from dinitrogenase of a substance referred to as FeMo-co (Fe = iron, Mo = molybdenum, co = cofactor). When this compound is added to a nitrogenase-deficient mutant of the bacterium *Azotobacter vinelandii*, it restores nitrogenase activity. Because this relatively simple compound carries the active site of the enzyme, it becomes particularly attractive to attempt to crystallize it and to obtain its detailed three-dimensional structure by X-ray crystallography. It might provide a model for developing improved catalysts for chemical nitrogen fixation.

Improving Nitrogen Fixation

Because improved systems for biological nitrogen fixation could translate into better nutrition for millions of people with less drain on energy reserves, scientists are working to develop more effective bacteria, more responsive plants, and more productive associations between the two.

Genetic and Biological Aspects of Nitrogen

Fixation Several types of free-living nitrogen-fixing bacteria already exist in nature; the task that confronts scientists is to improve their effectiveness. In the past, scientists working to improve nitrogen fixation have simply selected the most effective naturally occurring strains. Today, a more fruitful approach may be genetic manipulation.

The gene that controls the nitrogenase system is referred to as the *nif* gene. In a recent major advance, the *nif* gene has been transferred from nitrogen-fixing *Kle lla pneumoniae* to the non-nitrogen-fixing organism, *E. coli*,[4] prompting speculation that it might be possible to transfer the gene to non-nitrogen-fixing higher plants, such as corn or wheat. However, merely transferring the *nif* gene does not establish a useful nitrogen-fixing system. Unless the recipient—*E. coli*, corn, wheat, or whatever—can be protected from oxygen, no nitrogen will be fixed. Moreover, as we have noted, nitrogen fixation requires a large supply of energy in the form of ATP, as well as an ample supply of ferredoxin. Because this energy would come from energy normally used for other plant processes, the crop yield would be decreased unless overall photosynthesis could be increased. Developing auxiliary systems for supplying more energy and for protecting against oxygen may well prove more formidable than actually transferring the *nif* gene. Despite these complications, genetic techniques, including those employing recombinant DNA, may open ways to develop effective nitrogenase systems; and gene-transfer studies should be pursued vigorously.

Nitrogen-Fixing Legumes Another approach to improved agricultural production resides in more widespread use of nitrogen-fixing legumes—peas, beans, clover, alfalfa, and soybeans. (There are over 10,000 known species of leguminous plants.) Various species of *Rhizobium* invade the roots of various legumes where they produce nodules that consist of plant cells packed with nitrogen-fixing bacteria (Figure 8). In this symbiotic relationship the plant supplies carbohydrate to the bacteria, and the bacteria supply the plant with fixed nitrogen. Commercial preparations (inoculants) of the rhizobia, added to the soil, assure an adequate infection of legume roots. Farmers have long exploited the ability of such leguminous crops to revitalize the soil, by rotating them with crops that do not develop such nodules.

There is an extensive program for selecting native bacterial strains for improved nitrogen fixation; this program could be broadened to include testing of artificially produced mutants. Studies must include tests of the leguminous plant as well as bacterial strains, because effective N_2 fixation depends upon proper interaction between the specific plants and bacteria. Moreover, added strains of the root-nodule bacteria must compete effectively with native strains that are present in the soil: It is of little use to select a strain that is effective under greenhouse conditions if it does not compete well in the field. Hence, there is great interest in establishing the basis of this characteristic so that it can be incorporated in inoculants for leguminous seeds. Unfortunately, testing is slow and laborious, because the index of success is the increment in crop production.

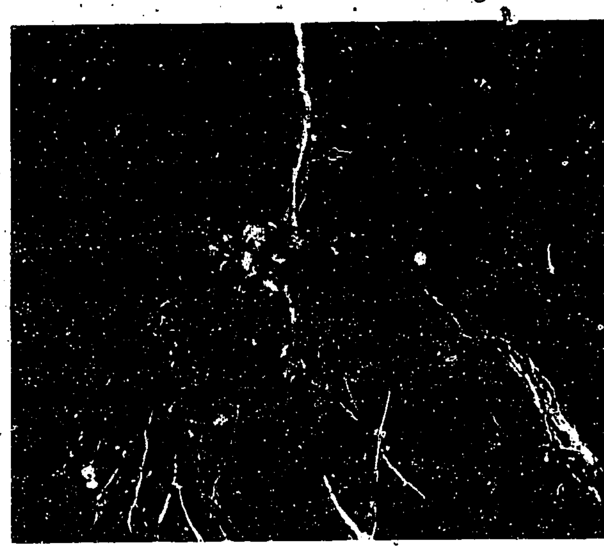


FIGURE 8 Nodules formed by the nitrogen-fixing bacterium *Rhizobium japonicum* on a young soybean root. (Harold J. Evans, Oregon State University)

Recycling Hydrogen To the extent that the nitrogenase system readily diverts electrons to H^+ rather than to N_2, the system is wasteful. And there is no known way to limit H_2 production by nitrogenase without also impairing nitrogen fixation. But certain root-nodule bacteria contain a hydrogenase, an enzyme that activates H_2 and permits its reoxidation; this yields ATP that can be used for nitrogen fixation. Field studies support the idea that rhizobia selected for their hydrogenase activity fix nitrogen more efficiently than bacteria lacking hydrogenase. Although this approach has been limited to soybeans, it should also be applicable to other species of legumes.

Delaying Senescence In the legume–rhizobia symbiosis it would be useful if plant aging could be delayed, for during the most active period the rates of growth and nitrogen fixation of the plant are essentially exponential. Extending the growing season by even a week could thus markedly increase crop yield.

The leguminous plant, however, operates in a self-destructive manner: At the time it sets seed, it senesces, utilizing all of its nitrogen resources, including nodules, to form seeds. By selecting plants that age more slowly or by using chemical agents, it may be possible to control senescence. Such leads need to be followed vigorously to convert them to practical agricultural techniques.

Other Approaches Legumes, as C_3 plants, are less efficient photosynthetically than C_4 plants. There would be great reward if legumes could be converted to a $C_4 + C_3$-type metabolism. This has not yet been accomplished, but efforts are continuing.

Increasing the concentration of CO_2 available to plants dramatically increases their nitrogen fixation. It is seldom practical to increase CO_2 concentrations under field conditions; but this phenomenon suggests that the full potential of the nitrogen-fixing system has not yet been reached. It may be possible to alter control mechanisms so that organisms will fix nitrogen more effectively at normal levels of CO_2.

Associative Symbioses In the associative symbioses, which are much less structured than the rigid symbiosis of leguminous plants and the rhizobia, bacteria capable of fixing nitrogen by themselves grow on or inside the roots of a host plant without producing a structured nodule. The bacteria obtain photosynthetic products from the roots and use them as a source of energy for fixing nitrogen, which is then available to the plant.

Under tropical conditions, *Azospirillum* associated with the roots of grasses actively fixes nitrogen. Tests have demonstrated that nitrogen can be fixed in such associations with maize, but the fixation characteristically is only 1 to 10 percent of that in better associations of leguminous plants and their root-nodule bacteria.

Because so much of the world's population depends primarily upon cereals for food, it would be of great importance if associative symbioses could be developed to supply a substantial percentage of the nitrogen needed for cereal crops. The possibilities in this field have been examined only superficially; an exhaustive examination of both bacterial species and plant cultivars may reveal associations that will fix a high percentage of the needed nitrogen. Field trials to date have not been encouraging, but the great potential of effective associative symbioses for cereal culture makes it imperative that the studies continue.

Blue-Green Algae The numerous blue-green algae that can fix nitrogen offer a fine example of nature's ingenuity. These organisms grow in chains of cells; about every twentieth cell differentiates into a special thick-walled cell known as a heterocyst. Unlike the normal vegetative cells in the chain, which photosynthesize and produce oxygen, the heterocysts lack the oxygen-producing phase of photosynthesis. The vegetative cells produce sugars photosynthetically and transfer them to the heterocysts. The heterocysts, in turn, fix nitrogen in their oxygen-deficient environment—an arrangement that keeps the heterocyst nitrogenase from being destroyed by oxygen—and, in exchange, they transfer fixed nitrogen compounds to the vegetative cells. These self-sufficient organisms are widespread, from deserts to lakes.

Nitrogen fixation by blue-green algae is crucial to maintaining rice production in Asia. The algae grow in abundance in paddies during flooding; as the water retreats, they decompose at the soil surface, leaving their fixed nitrogen to nourish the developing rice plants. This explains why rice can be cultivated year after year in paddies without adding nitrogenous fertilizer.

It is important that modern scientific methods focus on the nitrogenase system of blue-green algae to improve its effectiveness in rice culture and in other applications as well. Strain selection, genetic manipulation, control mechanisms, nutrient requirements, and conditions of culture all offer avenues toward improving fixation by these organisms.

CONTROL OF GROWTH AND PLANT STRESS

Plants show tremendous diversity in size, morphology, metabolism, composition, and response to environmental conditions. Nevertheless, they are much alike in their cell structures, ability to photosynthesize, need for water and mineral nutrients, and responses to growth substances. In nature, plants are subjected to extremes of temperature, water and nutrient supply, solar radiation, wind velocity, disease, and insect attack. Surviving plants have adapted to these stresses, but crop plants bred for high productivity may lose some of their resistance to stress. Fortunately man can control some stresses and minimize the effects of others.

Plant Growth Substances

Plant growth substances are products that spur or stunt plant growth. The first such substance to be isolated and characterized, about 50 years ago, was the auxin, indoleacetic acid. During World War II a variety of compounds with profound growth effects were synthesized, and some were adopted as selective weed control agents shortly thereafter. Other agents, developed to provide broad spectrum and selective weed control, were eagerly accepted in agricultural practice.

Growth substances control a variety of plant responses. The auxins and gibberellins control cell enlargement, whereas the cytokinins control cell proliferation and leaf expansion. Other agents control responses to light, cell differentiation, internodal elongation, germination, rooting, bud development, and so on.

These substances have found important practical applications. An early use was to stimulate the rooting of cuttings. Another was to delay development of buds on fruit trees until danger of frost damage had passed. Growth substances can minimize the sprouting of potatoes in storage, break the dormancy of seeds, and control the senescence of maturing plants.

The precise biochemical locus of action of herbicides is well defined in only a few cases. Most herbicides have been developed empirically; when an active compound has been synthesized or isolated from a natural source, a family of similar compounds has been synthesized and tested for activity A more rational approach, based on knowledge of plant metabolism, may lead to new herbicides with high specificity and low toxicity to animals.

The largest application of growth substances, used in combination or succession, is as selective herbicides in field crops. Pre-emergence herbicides can be applied at the time of planting, effecting early weed control with a saving in labor. Moreover, herbicides are important for minimum tillage, a procedure that saves energy, reduces compaction of the soil, and minimizes erosion and runoff. Perhaps 10 percent of the cropland in the United States is now being worked on this basis, and there is pressure to increase this percentage. Development of herbicides specifically designed for minimum tillage in different areas and for different crops is a challenge to agricultural industry.

Continuing research on plant growth substances will surely uncover new and improved agents and will suggest new applications.

Although the use of herbicides has engendered concern, it is doubtful that there has been a well-documented case of serious injury in humans from the proper use of herbicides on crop plants in the United States. Nevertheless, in the development of new agents for the protection or control of growth in plants, it will be prudent to place emphasis on the production of nonpersistent compounds that are not toxic to animals.

Plant Tissue and Cell Culture

Just as with animals, it has become easier to study many plant responses under highly controlled conditions with tissue cultures instead of intact plants. Growth control processes, for example, can be studied very effectively with plant tissue cultured in an appropriate medium. Small pieces of tissue or even single cells from certain plants are totipotent, i.e., with proper manipulation a single cell can be grown into an entire plant.

In the usual reproduction of plants, by seeds or vegetatively, viruses and other pathogens often are carried along. By manipulating plant tissue cultures, it is possible to develop and maintain virus-free and bacteria-free plants. Culturing specialized plant cells by fermentation techniques may prove more effective for generating useful products than growing whole plants. Genetically altering cells in tissue cultures may also speed the search for plants resistant to disease and other stresses. The culture of cells and meristems has become a common horticultural practice with prospects for wide agricultural application.

Effects of Light and Photoperiod

Among the most spectacular responses of plants to light and photoperiod (regularly recurring changes of light and darkness) are those generated through the action of phytochrome, a blue-green protein that serves as a light-activated biological switch. The form of the phytochrome molecule can be changed in different ways by different wave-lengths of light, and the response of the plant depends on the light to which the phytochrome was last exposed. Phytochrome exposed to far red light, for example, elicits changes in the bioelectrical potential, root tip adhesion, coleoptile elongation, leaflet closure, plastid movement, ATP levels, promotion of seed germination, and flowering of plants. Practical work on the phytochromes could be emphasized with benefit.

Plant growth and development is generally dependent upon the length of the light and dark periods. Certain plants will stay in the vegetative stage during long days and enter a reproductive stage and flower only when the photoperiod is shortened. In contrast, some familiar garden plants stay vegetative during the short spring days but exhibit reproductive growth and set seed when the days lengthen. Most photoperiodic responses can be effected at low light intensity. However, the control of growth may be influenced by the intensity as well as the quality of the light, and also by plant growth substances. Although control over light is seldom practical in the field (shade-grown tobacco being one exception), it is used in greenhouse operations to bring plants into flower at desired times. Chemical modification of light responses offers potential for field applications.

Water Supply

Frequently, the water supply limits plant growth. By controlling the opening or closing of the stomata—the pores on leaves—it is possible to control the rate at which water is lost. Plants in dry habitats frequently restrict water loss by closing their stomata during the day and opening them at night. However, chemical control of stomatal opening will have rather limited practical application until agents are developed that cause little damage to the plant while they restrict its water loss. A natural growth factor that regulates stomata, abscisic acid, may prove useful. Attempts also have been made to breed plants that assert greater stomatal control, or that have a reduced number of stomata. However, these approaches are limited by the need for open pores for uptake of CO_2 commensurate with the growth rate.

The method of supplying water also warrants attention. Overhead sprinkling distributes water uniformly, but it is subject to heavy losses through evaporation on hot and windy days. Open trench irrigation incurs much less evaporation, but it can leach nutrients from the root zone and transfer them to areas not penetrated by the roots. Trickle irrigation from porous or perforated tubing is attractive, because the water can be placed directly in the area of use, supplied under close control, and losses avoided of nutrients from the root zone. Improved porous tubing should increase the popularity of this practice for specialty crops.

Frost Resistance

Some plants show remarkable winter hardiness, whereas others show damage even above the freezing point. The speed of freezing and thawing, the physiological condition of the plant at the time of exposure, the adjustment of the plant to gradually lowered versus suddenly lowered temperature, and the state of water in the plant all influence the plant's response. Certain diseased plants exhibit an increased cold hardiness, suggesting that subtle changes in a plant may influence its response to cold. Breeding has improved the cold hardiness of a number of crop plants, but more knowledge of the mechanisms of cold resistance and susceptibility would be helpful.

Plant Stress from Diseases and Pests

A wide spectrum of chemical agents has been developed to control plant diseases and pests. Some are effective but costly; some are only slowly degradable. Agents that are not biodegradable can build up in the soil; if assimilated into the plants, they can be carried into the human or animal food supply with opportunity for toxicity. Breeding plants for disease or pest resistance avoids this problem, and indeed it is one of the more spectacular successes of agricultural research. Research costs are modest relative to benefits; inbred resistance eliminates investment in special field labor, and it avoids use of toxic compounds. However, breeding for resistance is a never-ending task, because as resistant plants are developed, new strains of pathogens may appear.

Crown gall is a classical bacterial disease that attacks a variety of plants. The growth of a tumorlike gall does not require the continued presence of the intact bacterium, but only of a large plasmid (an autonomously replicating piece of DNA) carried in the bacterium. Bacteria lose their virulence if these plasmids are removed, and regain it if the plasmids are restored. This development suggests possibilities for establishing a more rational method for control of the crown gall.

Insect pests present a different problem. Although there has been some success in building insect resistance into plants, this has proved more difficult than incorporating resistance to microbial infection. The usual rationale is to breed plants with an increased level of compounds that are repellant, distasteful, or toxic to the insect. Whether plant- or man-made, agents toxic to insects must be monitored with great caution to assure that they are innocuous in the food supply. It may be that agents that disturb the life cycle of the insect (e.g., juvenile hormone) or which counteract pheromones (insect attractants) may be incorporated into plants for their protection.

Mineral Nutrition of Plants

The availability of an essential mineral from the soil depends upon the acidity of the medium, the presence or absence of other compounds or mineral ions, the binding properties of the soil, and the feeding effectiveness of the plant. Plants may be bred for high feeding capacity. Nutrient elements not used by the plants may be leached from the soil, or the soil itself may be eroded and carry away nutrients on the soil particles.

The usual procedure for improving crop yields in nutrient-deficient soils has been to adjust nutrient levels through fertilizer and lime applications. Recent studies indicate that by applying insights from plant physiology and plant breeding it will be possible to develop economic crops much more efficient in nutrient utilization. Adaptation of the plants to specific nutrient deficiencies is a promising alternative to the traditional fertilization and liming of nutrient-deficient soils. Maize strains that feed more efficiently on both added fertilizer and soil nutrients, particularly potassium, have recently been developed. Continued research on this problem promises excellent returns.

NUTRITIONAL QUALITY OF PLANTS

Alteration of plant cultural practices has little influence on the nutritive quality of plants. However, it is possible to breed plants with increased abundance of vitamins and altered balance among their carbohydrates, proteins, and lipids. After the biochemical nature of a genetic alteration has been established, breeding may proceed on a more rational basis. Plant breeders try to achieve high nutritional quality together with high yield, disease resistance, and other desirable qualities. A striking case in point is the development of a strain of maize relatively abundant in lysine, an amino acid essential in the human diet; classically the low lysine content of maize has limited the nutritional value of its protein. As yet, the yield of this strain is low, but this limitation will surely be overcome by further breeding.

In addition to improving traditional crops, new crops such as the winged bean are being examined for their nutritional quality, in the hope of finding highly productive plants that will be useful in producing more balanced diets.

Leaves, which contain a limited amount of high-quality protein, are usually discarded or returned to the soil in agricultural practice. Leaf protein can be separated from structural materials by simple techniques. Protein from alfalfa leaves has been used successfully in poultry feed, but it has been used for human consumption only on an experimental basis. Nothing precludes such use, and the yield of leaf protein per acre is impressive. Residues left after extruding soluble proteins from leaves can be converted to silage for dairy cattle.

Nowhere in this presentation have we forecast an early, major breakthrough. None is in sight. Rather, we observe the need to pursue fundamental studies of plant physiology and the applied research necessary to take advantage of myriad opportunities to achieve modest increases in productivity, reduction in cost, and conservation of resources. Their sum could well be of huge economic benefit and provide greater assurance of a continuing food supply. Emphasis has been given to those opportunities relevant to domestic agriculture. Research opportunities for agriculture in developing nations, and the arrangements that might facilitate such research, have been described in greater detail in the recent NRC report *World Food and Nutrition*.[5]

CONCLUSIONS

UNITY AND DIVERSITY

The selections in this chapter have illustrated the dynamism of current molecular and cellular studies in various areas of biology, and the remarkable unity of the living world at this level of organization. The source of this unity is the evolutionary continuity of the living world: Organisms today perpetuate many successful biochemical structures and mechanisms that emerged in the earliest living cells 3 billion years ago.

However, our survey of biology would be unbalanced if we did not also note the other side of evolution: diversity. For evolution arises from the continual production of novel combinations of genes, together with the preferential survival, by natural selection, of those organisms that are better adapted to their environment. This process has yielded literally millions of species, of gradually increasing complexity and adaptability; and it has culminated in our species—one whose unique capacity for abstract thought, deep feeling, communication of complex information, construction of efficient tools, and creativity have added a new chapter: cultural evolution.

For some years the dramatic advances in molecular studies rather eclipsed studies concerned with the properties of whole organisms and populations. However, concern for conservation of the environment, and of the rich legacy of evolution, has aroused wide interest in ecology. Furthermore, a bridge now links evolutionary and molecular studies, as the latter have made it possible to quantify precisely the amount of genetic difference between species (and between individuals within a species), by measuring the degree of similarity in the sequences of their DNA or of its protein products.

Nevertheless, evolutionary studies have a unique quality: The physical sciences, and their applications to biology, deal primarily with uniform entities (i.e., the identical molecules of a pure substance); but population biology deals with groups in which almost every individual is genetically unique. So the population distribution of genetically determined traits (such as blood groups), and of genetically determined potential (physical and behavioral), must be characterized in statistical terms. The result has been the replacement of stereotypes with a true appreciation, in general terms, of the scope of human diversity: Within every group individuals differ genetically over a broad range.

An important recent development has been the emergence of the discipline of sociobiology, which aims at understanding the biological foundations of social behavior in animals, and in man in particular. This approach aims to throw light on the biological constraints that evolution has imposed on our species. Within these constraints cultural forces produce an enormous, but not limitless, range of social behavior.

CONCERNS WITH FUTURE IMPACTS OF BIOLOGY

With the increasing depth of our advances in biology some have come to fear that certain areas of research, especially in molecular and human genetics, threaten public welfare and thus need to be regulated. The recent intense debate over recombinant DNA research is a case in point. The potential dangers perceived involved both those inherent in the research itself (Could a novel virulent organism escape from the laboratory to annihilate whole populations?) and those latent in new knowledge that could confer potentially dangerous powers (Might genetic engineering be used to manipulate our personalities?).

The anxiety over recombinant DNA research has abated considerably, for several reasons. Among these is the failure of five years of experience to produce evidence of any illness or other harm. In addition, sober professional analyses have gradually displaced earlier unrealistic demands for absolute protection against hypothetical risks. Accordingly, the NIH guidelines are gradually being relaxed; and the Congress has determined that the guidelines are adequate for handling the problem without legislation. Nevertheless, the costs of both the debate and the resulting restrictions, in money, time, and morale, have been large.

When the hazards of research are well defined, as with toxic or radioactive substances, or when experimentation on human subjects involves real risk, the problem is straightforward, though sometimes complex. However, when available facts do not clearly define or quantify the risk, it is more difficult for society to reach the best judgment. The experience with recombinant DNA suggests that in similar discussions in the future it is important to try to separate the technical phase of assessing the risks and benefits, on the basis of expert and informed judgment, from the political phase of legislative and public participation in formulating the required value judgments and reaching policy decisions.

In contrast to risks inherent in carrying out research, there is the risk that the knowledge produced by research may be used for harm as well as good. Since virtually any knowledge is double-edged, and since its consequences cannot be predicted in detail, our society has proceeded wisely, in our view, in the belief that *on balance* knowledge is less likely to be dangerous than is ignorance. Our society is trying to assess and to regulate harmful technological applications earlier before damage occurs, but has declined to regulate research on the basis of speculations about possibly harmful applications.

Still other types of new biological understanding raise fears that deeper insights into human nature might imperil a just and decent society. The example most often cited is study of human behavioral genetics, which would sharpen our perception of inborn individual differences in various intellectual capacities, talents, drives, and learning patterns. Here, too, knowledge is two-edged, with possible misuses and valuable uses.

The problem will not remain confined to genetics. Eventually, advances in sociobiology, neurobiology, and the behavioral sciences are also likely to conflict with treasured preconceptions, widely held to be indispensable foundations for public morality. But human curiosity cannot be permanently extinguished, nor can the scientific method be unlearned. Someone will learn, somewhere, sometime. Moreover, the realities will be there, whether or not scientists are permitted to find them; and if we build social policies on false assumptions, which contradict reality, we will be building on a crumbling foundation. A democratic and open society, therefore, has no choice but to defend freedom of inquiry, just as it defends freedom of expression.

OUTLOOK

The following outlook section on the living state is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

A remarkable burst of advances is propelling biology into a new era. Powerful new tools are generating profound insights into the secrets of life, and scientists are now able to move with assurance among realities that they could only contemplate 10 years ago. There is no reason for this momentum to falter. Rather, biologists look for a flowering of their science in the years ahead. Extrapolating new knowledge and exploiting new technologies, they are eager to discern ever more clearly the molecular organization and functioning of living cells, tissues, and organisms.

MOLECULAR GENETICS

In molecular genetics, basic mechanisms that have been elucidated in bacteria and viruses are now being studied in higher organisms. Since major health problems almost invariably involve cellular malfunction, frequently reflecting intrinsic or imposed defects in gene structure or regulation, molecular genetics is in a position to exert a major impact.

The potential contribution of recombinant DNA technology is hard to exaggerate. By splicing fragments of DNA from higher organisms into simple bacteria, one can purify the DNA segments, obtain them in quantity, determine their base sequences, and study the behavior of their genes. Early experiments using recombinant DNA methodology in higher cells are showing the types of surprises that can be in store: The structural genes of higher organisms, which code for the synthesis of specific proteins, are not simple continuous sequences of DNA, as in bacteria and viruses. Instead, each is a set of DNA base sequences interrupted on the DNA helix by several long, unexpressed sections. Scientists are investigating the hypothesis that the intervening, nonactive sequences may somehow function in determining which genes in a cell are expressed at a given time. Other studies are exploring the precise mechanisms by which DNA is duplicated when cells reproduce. Still other studies are investigating the mysteries of the structure of chromosomes, where DNA resides.

Molecular biology is also facilitating studies of mutations in genes, both those in germ cells, which may cause inherited disorders, and those in body cells, which may lead to the development of diseases such as cancer. It may take 10–20 more years before we adequately understand diseases of cellular malfunction such as cancer or autoimmune disorders. Other diseases, however, such as defective hemoglobin synthesis, will probably be understood sooner. In the years just ahead we can also expect improved ability to detect, via amniocentesis, various genetic defects in fetuses. There will be improved procedures to assay the mutagenic properties of chemical compounds and, hence, their potential for carcinogenesis.

Advances in understanding the subcellular structure called the ribosome have revealed how a number of antibiotics work, and how cells develop resistance to them.

Another prospect involves using recombinant DNA methods to manipulate microorganisms so that they will synthesize specific gene products. Human hormones will become more readily available. It may become possible to design and synthesize new enzymes or to manufacture known enzymes cheaply enough to substitute them in certain processes. Microorganisms may yield useful sources of energy, or food plants may be freed of their dependence on com fertilizer as a source of nitrogen.

A relatively distant advance is the possibility of treatin diseases by replacing defective genes with normal ones. The such gene therapy to become a reality will probably be replacement of defective blood cells, which are produced from a stem cell line in the bone marrow in a way that allows them to be replaced from the blood stream. With more highly organized organs, however, the problem seems insuperable. It is even less likely that a gene replacement can be carried out in sex cells, as would be required to eliminate further inheritance of a genetic disorder.

CELL BIOLOGY

With knowledge derived from simpler systems, and using DNA technology, the study of the cell biology of higher organisms will move toward a new level of sophistication. Research efforts will be trained on the intricate mechanisms that integrate the complex interactions of genes and enzymes. Such knowledge is pivotal for understanding not only normal cell reproduction and differentiation, and presumably the conversion of a normal cell into a cancerous one, but also the process by which cells age.

The fascinating structure and workings of cell membranes, which maintain the vital balance between fluids inside and outside a cell, and which are the site of interactions of a cell with hormones, viruses, bacterial toxins, and drugs, are the object of intense study. Such investigations will have importance for a great variety of health-related concerns. Critical to new understanding are studies of the newly discovered specific receptor proteins in cell membranes, which are the specific sites of attachment of hormones, neurotransmitters, etc.

Studies are probing the molecular basis for the cytoskeleton's dynamic nature, seeking to understand the motility that allows white blood cells to mobilize to fight infection, or cancer cells to invade nearby organs, or, within the cell itself, for subcellular components to move about with precision and efficiency.

IMMUNOLOGY

Enriched by the new tools and understanding of biochemistry, molecular genetics, and cell biology, immunological research has gained new sophistication and become a powerful source of new insights. Accordingly, one may confidently expect a marked increase in the capability to manipulate both desirable and untoward immune reactions.

New vaccines are in the offing for several infectious diseases, including serum hepatitis and malaria. Interferon, a powerful antiviral protein manufactured by the body, is promising to be therapeutically effective; its current scarcity may be remedied through the use of recombinant DNA technology in a few years. Kidney transplants can expect to enjoy much better survival rates in the years ahead, mostly as the result of improved matching of donors and recipients. Graft survival will also benefit from new therapies with drugs, irradiation, and antibodies against destructive white cells.

In allergy research, new discoveries about the types of white blood cells that influence allergic reactions promise to lead to a rational, guided approach to desensitization therapy. Moreover, the recent discovery of cell receptors for the antibodies that provoke allergic symptoms may permit the design of chemicals to block these receptors and thus abort the symptoms. Chemical studies on the purified components of the allergens will bring improvements to hyposensitization therapy.

Individuals with defects in their immune systems that destroy their resistance to infection may soon benefit from transplants of appropriate tissue such as bone marrow or thymus. New discoveries regarding mechanisms of tolerance promise to lead, relatively soon, to improved methods for controlling several autoimmune diseases in which the immune system attacks a normal tissue; already patients with myasthenia gravis are showing improvement after surgical or drug therapy to alter immune function.

There is hope that advances in tumor immunology will eventually make it possible to eliminate cancer cells that remain after chemotherapy. So far, however, efforts at cancer immunotherapy have been disappointing.

NEUROSCIENCE

New insights will continue to illuminate the remarkable workings of the human brain. Recent research has revealed the nature of the action potential and of synaptic transmission; the synapse, the basic mode of communication in the brain, has become the unifying concept in neuroscience.

Thanks to the development of new techniques, analysis of the structural organization of the nervous system—the wiring diagram—is proceeding at a rapid rate, and a detailed topography of the functional anatomy of the brain is imminent. Spectacular progress in the identification and study of the modes of action of synaptic neurotransmitters promises to enhance markedly the understanding of the biological basis of behavior ranging from neurological disorders to motivation, sleep, motor systems, memory, and psychoses. Most importantly, to the extent that behavioral disorders have their bases in altered neurochemical functioning, they are subject to pharmacologic remedies.

The study of hormones and their actions, both those produced in the brain and elsewhere, is also advancing rapidly. Hormones, now believed to act on the brain in a manner analogous to synaptic transmitters, underlie such important phenomena as sexual development and basic motivations like thirst and response to stress, and may modulate higher order processes such as learning and memory.

That the brain also produces its own hormones, most of them polypeptides, has just been recognized. Some were first known to exist as the releasing substances of the hypothalamus that occasion synthesis and release of pituitary hormones; others as hormones functioning in the gastrointestinal tract. Most recent are the endorphins, the brain's own opiates, with high affinity for opiate receptors. How many such brain hormones there are, what are the physiological and psychological consequences of their activity, and whether they act at synapses or other receptors, constitute a chapter that has just opened and will be pursued apace.

It seems likely that in a decade or so we will know how developing nerves seek their targets and we may even be able to improve regrowth of damaged neural tissue. In view of the severity of derangements of neurological development—birth defects, muscular dystrophy, epilepsies—these advances cannot fail to make a tremendous clinical impact. However, to understand how the gene program is translated into a nervous system remains one of the most gigantic challenges that neuroscience offers.

Much new information will be gained in the next 5–10 years about how the brain codes and integrates sensory information. New treatments for sensory disorders will emerge and prostheses for certain sensory defects may be developed. The most rapid advance in analysis of motor systems will probably relate to certain of the synaptic chemical transmitter systems and may result in improved treatment of certain abnormalities of movement.

In the next decade or so, much will be learned about the neuronal substrates of basic motivations, and more effective treatments may become available for such important and intractable behavioral disorders as drug addiction and obesity. Systematic research will probably elucidate the circuits in the brain that underlie simpler forms of learning and memory. Since memory probably involves persistent physical changes of some kind at nerve-to-nerve junctions, insights into the basic mechanisms for learning and memory are surely forthcoming, although precise localization of the memory trace for a specific bit of information may remain an elusive goal. It is possible that treatments may be developed to improve certain forms of learning disabilities. Finally, with rapid growth in our knowledge of synaptic transmissions, we seem on the threshold of major improvements in the understanding and treatment of schizophrenia, the most severe and widespread form of mental illness.

In most neurological conditions that lead to substantial brain destruction, the prospects for a real cure are indeed poor, for such a cure would involve the removal of scars and debris and the regrowth and rewiring of countless nerves. However, one can look for significant advances in mitigating the effects of such conditions. Better drugs will be developed against epilepsy, psychoses, Parkinson's disease, spasticity, etc. The outlook for devastating diseases such as multiple sclerosis and Alzheimer's disease is somewhat different. The hope for a cure seems to be reasonably good—especially if damage to the nervous system has not progressed very far. However, the cures will probably come from immunology or virology, not neurobiology.

While neuroscience is not promising immediate cures for the wide range of neural and behavioral disorders that afflict humanity, the current rapid growth of knowledge about the brain and nervous system makes it seem highly probable that improved treatments for many disorders will become available in the next few years. Perhaps even more importantly, as we learn more about the brain we will approach closer to a genuine understanding of the human condition.

BIOLOGY AND AGRICULTURE

Biological research in agriculture, stimulated by advances in so many areas, is striving to increase productivity while reducing dependence on chemical fertilizers and pesticides.

More detailed understanding of the remarkable process of photosynthesis is engendering efforts designed to enhance its efficiency. Research will be directed to develop crop plants that utilize light more efficiently, while other work will focus on altering key enzymes so that they are less easily diverted to unproductive activity during the dark reactions of the photosynthetic process. Other research will attempt to convert the ordinary C_3 plant to the more efficient type of metabolism seen in such C_4 plants as maize.

Because improvements in biological nitrogen fixation could translate into better nutrition for millions, with fewer demands on energy supplies, scientists will be working to develop more effective bacteria, more responsive plants, and more productive associations between the two. Free-living nitrogen-fixing bacteria already exist in nature; scientists will be employing genetic techniques, including those of recombinant DNA, to make them more effective. Cereals feed much of the world's population, and there will be continuing efforts to develop plant-bacteria symbioses that could fix nitrogen for cereal plants. Similarly, a variety of approaches is being explored to improve the natural fixation that blue-green algae provide in rice cultivation.

In the area of plant growth and stress, there will be continuing efforts to discover new and improved plant growth substances, and to foster new applications. Plant tissue cultures offer many advantages. Not only do they make it possible to develop and maintain plants free of infection with viruses or bacteria, they may provide ready access to useful plant products, and they may speed the search for plants resistant to disease and other stresses.

Chemical substances, too, open many avenues to progress. In addition to their role in controlling diseases and pests, they will be useful for manipulating a plant's responses to periods of light and dark, or for regulating the opening and closing of its stomata.

Breeding plants for specific characteristics is yet another biological approach to improving agricultural productivity. The breeding of plants for resistance to disease and pests has been one of the more spectacular successes of agricultural research. Current work indicates that there are excellent prospects for developing economic crops that utilize nutrients much more efficiently. At the same time, plants are being bred for better nutritional value, with an increase in vitamins and an altered balance among carbohydrates, proteins, and lipids. New, highly nutritious crops will be developed, and previously untapped sources of nutrition such as leaves, which are usually discarded but which contain a limited amount of high quality protein, will be exploited.

Although no major breakthrough is in sight, basic studies of plant physiology and applied research promise to yield many modest advances. Taken together, these promise high economic benefit as well as a greater assurance of continuing food supply.

REFERENCES

1. Chan, H.W., *et al.* Molecular Cloning of Polyoma Virus DNA in *Escherichia coli*: Lambda Phage Vector System. *Science* 203(4383):887–892, 1979.

2. Israel, M.A., *et al.* Molecular Cloning of Polyoma Virus DNA in *Escherichia coli*: Plasmid Vector Systems. *Science* 203(4383):883–887, 1979.

3. Benditt, E.P. Implications of the Monoclonal Character of Human Atherosclerotic Plaques. *Annals of the N.Y. Academy of Sciences* 275:96–100, 1976.

4. Dixon, R.A., and J.R. Postgate. Genetic Transfer of Nitrogen Fixation from *Klebsiella pneumoniae* to *Escherichia coli*. Nature 237:102–103, 1972.

5. *World Food and Nutrition Study: The Potential Contributions of Research* (NRC Steering Committee Study on World Food and Nutrition). Washington, D.C.: National Academy of Sciences, 1977.

BIBLIOGRAPHY

Benditt, E.P. Origin of Atherosclerosis. *Scientific American* 236(2):74–85, 1977.

Brill, W.J. Biological Nitrogen Fixation. *Scientific American* 236(3):68–81, 1977.

Cairns, J. *Cancer: Science and Society.* San Francisco: W. H. Freeman, 1978.

Cooper, J.R., F.E. Bloom, and R.H. Roth (eds.) *The Biochemical Basis of Neuropharmacology,* 3rd ed. Oxford University Press: New York, 1978.

Davis, B.D. The Recombinant DNA Scenarios: Andromeda Strain, Chimera, and Golem. *American Scientist* 65(5):547–555, 1977.

Eccles, J.C. *The Understanding of the Brain,* 2nd ed. McGraw Hill: New York, 1977.

Eisen, H.N. *Immunology.* Harper and Row: New York, 1974.

Kuffler, S.W., and J.G. Nicholls. *From Neuron to Brain: A Cellular Approach to the Function of the Nervous System.* Sinauer Assoc.: Sunderland, Mass., 1976.

Mazia, D. The Cell Cycle. *Scientific American* 230(16):54–64, 1974.

Watson, J.D. *Molecular Biology of the Gene,* 3rd ed. Benjamin-Cummings: Menlo Park, Calif., 1976.

3 The Structure of Matter

INTRODUCTION

The effort to understand matter has challenged the human mind throughout recorded history. It has been motivated both by the desire to manipulate and control matter for human uses and by simple curiosity. As our understanding of matter has deepened, so has the range of study to include the less tangible phenomena of visible light and other forms of electromagnetic radiation.

Understanding matter includes knowledge of its forms and behavior under all imaginable conditions, the transformations that can occur among various forms of matter, and the natural laws that underlie these phenomena. Such knowledge makes it possible to predict the behavior of matter under new circumstances, to establish how it is fashioned and how new forms may be created, and to control its forms and uses.

In view of the great diversity and complexity of matter, it must be an act of faith to search for an underlying simplicity. While the modern physical scientist continues the pursuit of a goal that began in ancient Greece, he is under the severe constraint that his explanations must account consistently for reproducible phenomena and must unfailingly predict the behavior of matter. There is no greater miracle than the success of this enterprise. Indeed, the successes of the past justify faith in the future.

The structure of matter may be viewed on a scale that ranges from the very large to the very small, from the macroscopic to the microscopic (see Figure 9). Cosmology is concerned with the structure of matter on the largest scale, with the grand structure of the universe: How did it originate; what is its age; of what is it made? It involves the study of such fascinating phenomena as black holes, gravitational collapse and neutron stars, the "Big Bang," and so on.

Then there is the more human scale of the laboratory, one which permits experiments and studies of matter under carefully controlled conditions. Laws governing macroscopic matter are well developed and understood. Among the most elegant laws of nature are the laws of thermodynamics, which describe the thermal behavior of matter in terms of a few parameters characterizing each substance. Another example includes the detailed laws of statistical mechanics governing particulate structure of matter. Our general understanding of the mechanical, electrical, magnetic, and optical properties of most normal forms of bulk matter is also well established in terms of a few parameters for each material. However, there are exotic materials such as superconductors and superfluids for which the classical theories have not been

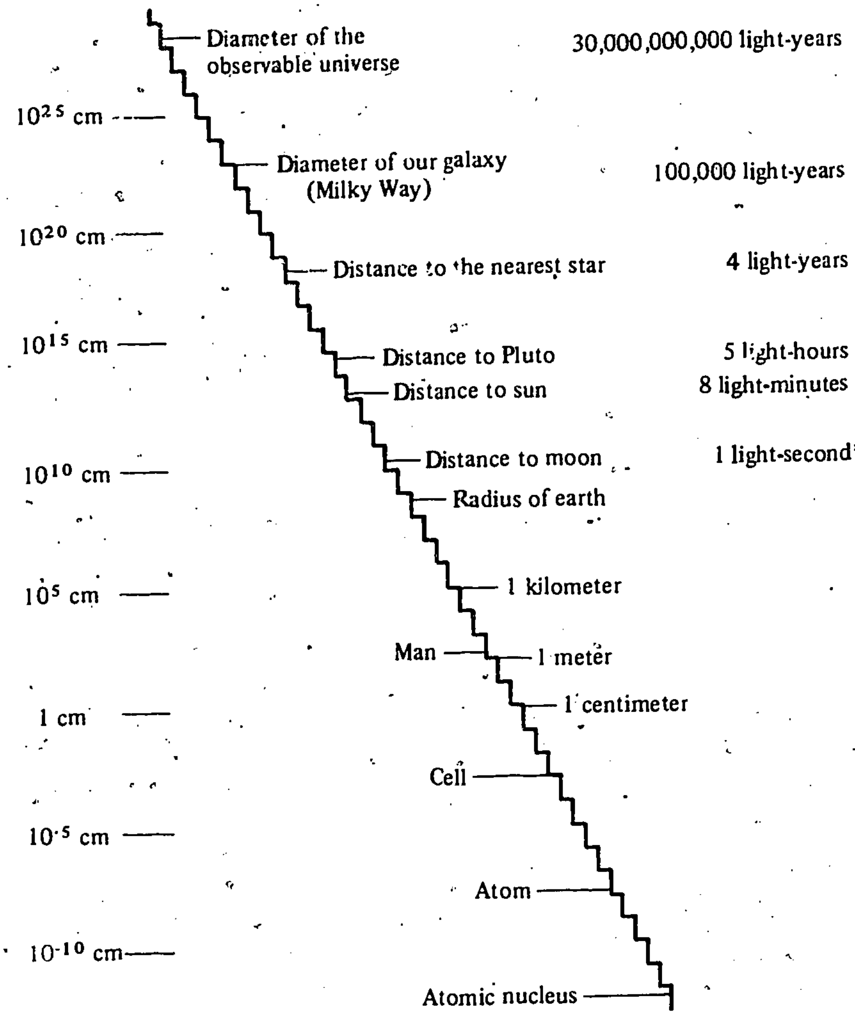


FIGURE 9 Range of sizes in physics (each step corresponds to a factor of 10 increase in size. *1 light-second = distance that light travels in 1 second = 3×10^{10} cm or 186,000 mi. (National Academy of Sciences, *Physics in Perspective*, Vol. I, p. 84. Washington, D.C., 1972)

adequate. They have turned out to be examples of macroscopic quantum fluids, a realization of the quantum theory associated primarily with the microscopic behavior of matter.

Indeed, in recent years, dazzling variations of the more familiar forms of materials have appeared, presenting challenges to established theories and many opportunities for application. Included are new types of semiconductors, amorphous materials with a great variety of tailored properties, polymeric and organic complexes, and two-dimensional as well as one-dimensional materials providing special opportunities for fundamental studies.

Finally there is the microscopic domain in which one attempts to identify the building blocks out of which matter is composed. In this domain, matter is reduced to collections of molecules and these to atoms. Atoms are reducible to electrons and a nucleus, the latter composed of neutrons and protons. Neutrons and protons now appear to be made up of "ultimate" building blocks, the quarks, a concept that arose out of considerations of internal symmetry introduced to account for the unexpected behavior of nuclear particles.

Although the general laws describing the behavior of matter at the atomic and molecular scale are well established, much remains to be done to arrive at a detailed description of the properties of matter on the basis of fundamental physical principles, quantum and statistical mechanics, except for very simple systems. High speed computers and new mathematical techniques are now advancing our knowledge of matter at the atomic and molecular scale. However, it will be a long time before our ability to understand matter at the molecular level will have advanced to where we can understand the building blocks of living systems in terms of basic physical laws.

ASTROPHYSICS AND COSMOLOGY

In earlier times, our knowledge of the universe beyond the earth depended entirely upon the perception of visible light, which represents only one very small portion of the broad spectrum of electromagnetic radiation that is emitted by matter under various conditions. The earth's atmosphere is an effective absorber of much of the radiation that falls upon it, and for this reason there are broad portions of the spectrum that can be detected only by specially designed instruments located above much of the atmosphere. During the past 25 years or so, however, remarkable instruments have been devised which allow us to see the cosmos as it appears in radio waves, in infrared and ultraviolet radiation, and in X-rays and gamma rays (Figure 10).

The sources of electromagnetic radiation are molecules, atoms, electrons, and the nuclei of atoms. Each of these forms of matter can exist in a number of different energy levels, or states, which correspond to different vibrational patterns, or different rates of rotation, or different configurations of the electron cloud surrounding atoms and molecules. The absorption of radiation raises the object to a higher energy "excited" state, and radiation is emitted when the object undergoes a change or transition to a lower energy state. The particular wavelengths or frequencies of the absorbed or emitted radiation are characteristic not only of the kind of transition involved but also of the specific kind of molecule, atom, or nucleus undergoing the transition. For example, the pattern of radiation emitted by iron atoms is different from that emitted by carbon atoms, and the same is true for different kinds of molecules and different atomic nuclei. The radiation pattern also depends on the temperature, pressure, and other conditions of the matter of which the atoms or molecules are a part. Thus the observation of electromagnetic radiation over a broad region of the spectrum provides a number of different and complementary windows through which the matter in the stars and galaxies of the universe can be studied.

In addition to the electromagnetic radiations, other radiations provide an increasing amount of information concerning the cosmos. For example, the charged nuclear particles in the cosmic rays are a direct source of information about galactic and extragalactic events. Also the ability to detect neutrinos offers a promising possibility for an entirely new means of astronomical observation. Furthermore, gravitational waves could be used as signals of catastrophic events if their detection could be successfully demonstrated.

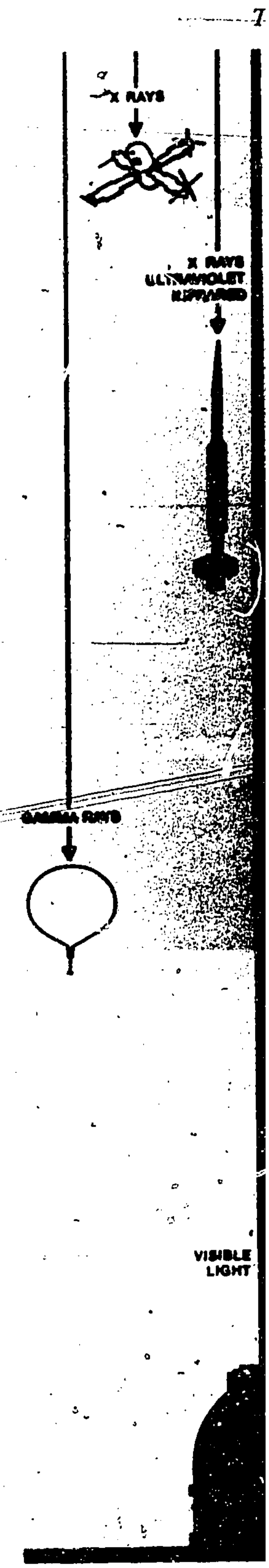
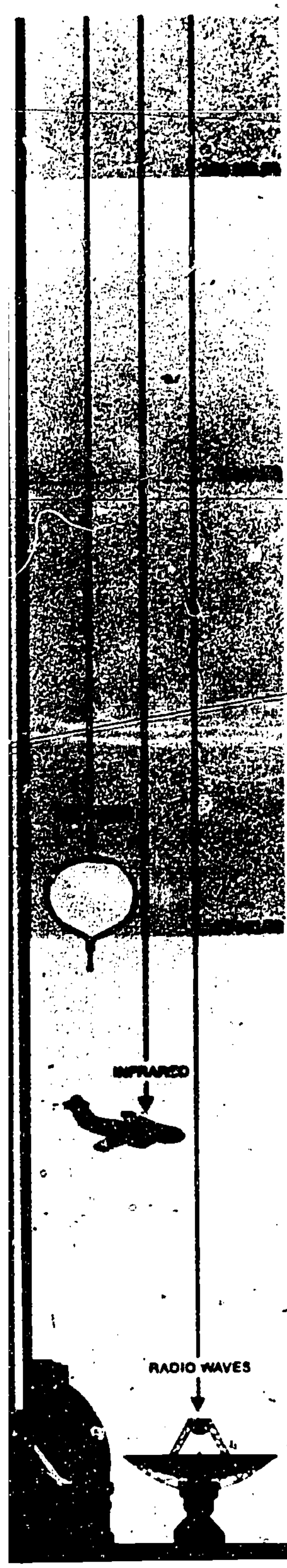


FIGURE 10 Electromagneti waves penetrate the earth's at ground-based instruments. Fo other parts of the electromagi be carried high up in, or abov

COSMOLOGY

During the early years of this century, most astronomers believed that all of the visible universe, both stars and nebulae, were contained within one great system,



radiation. Visible light and radio
iosphere and can be observed by
unobscured observation of most
etic spectrum, instruments must
the atmosphere.

the Milky Way galaxy. By 1929, however, Edwin Hubble had established the fact that many of the observed nebulae are actually "island universes" or independent galaxies much like our own, which are vastly more distant than had previously been thought, and that they appear to be flying apart from each other like pieces of shrapnel from a great explosion.

Our present picture is that the universe contains many billions of galaxies that are distributed throughout space rather uniformly in all directions out to the limits of our ability to observe them; and that the cosmic expansion first noticed by Hubble had its origin in a primordial explosion, the Big Bang, which occurred some 10–20 billion years ago.

Will the universe keep on expanding forever, or will the present expansion gradually slow down and perhaps eventually reverse into a collapse? If it does collapse back into a giant fireball, will there then be another Big Bang? Is ours a "one-shot" universe, or does it oscillate endlessly between expansion and collapse?[1] These questions are central issues in cosmology, and although they may seem mind-boggling, they do lend themselves to scientific analysis. Much of the excitement that pervades contemporary astrophysics arises from the power of the new observational astronomy and of modern theory to search out the cosmological evidence.

The Age of the Universe

Hubble's key observation was that the observed galaxies all appeared to be flying away from each other at speeds that were proportional to the distances between them. That is, galaxies twice as distant were receding from each other at twice the speed. By running this mutual expansion of all the galaxies backward, it is possible to arrive at the "Hubble time" when all of the galaxies would have been packed together in the primordial fireball of the Big Bang. Hubble's original calculations indicated a time of about 2 billion years since the Big Bang, but subsequent studies have steadily lengthened this estimate of the Hubble time until it is now placed between 10 and 20 billion years. The large uncertainty in this estimate still remains to be resolved, but it is noteworthy that the same range of values has been found for the ages of the oldest observed stars, and also for the ages of the chemical elements, which are determined by measurements of radioactivity.

The Hubble time would be expected to give the correct age for the universe only if it had been expanding at a constant rate since its beginning in the Big Bang. If the expansion has gradually been slowing down, however, then the actual age of the universe must be less than that indicated by the Hubble time. It is reasonable to expect that the mutual gravitational attraction of all the matter in the universe must work against the universal expansion and thus tend to slow it down, but by how much?

Density of Matter in the Universe

A second approach to the age-problem is to relate the expansion of the universe to the total amount of matter and energy that it contains. Gravity acts upon all forms of matter and radiation (including light), and mutual gravitational attraction will eventually stop the universal expansion if the average density of such mass-energy throughout the universe is great enough. Conversely, if the average density is less than a certain critical value, then the universe will continue to expand forever—an open universe. At present, the estimates of mass-energy in the *observed* forms of stars, gas, and dust fall far short of the critical density needed to halt the expansion and produce a closed universe.

If the universe is in fact closed, then there must be a great deal of matter present that has not yet been observed. Some preliminary evidence from the new X-ray astronomy points toward the presence of substantial amounts of matter in very hot, dilute intergalactic gas. There are also several other ways, not yet confirmed, in which large mass may be detected if present—in black holes, for example. However, arguments based on calculations of the abundances of elements produced by a Big Bang tend to favor an open universe of low density and these arguments include all matter whether in galaxies or intergalactic gases.

Ages of Stars

Many facets of astrophysics also bear on cosmological questions. The ages of stars are an important clue. For example, our present model of stellar evolution indicates that a massive star will consume only about 10 to 15 percent of its hydrogen before it expands into its "red giant" phase. This standard model also predicts that a certain number of neutrinos should be observed on the surface of the earth as a result of the thermonuclear reactions occurring in the core of the sun, but the most persistent efforts to detect these neutrinos have yielded only one-third or so of the predicted number. Some possible explanations for the solar neutrino puzzle would yield a model for the sun which might burn hydrogen for 20 billion years instead of only the 10 billion years predicted at present. Estimates of the age of the galaxy based on the ages of its oldest stars would then have to be revised correspondingly upward.

This uncertainty emphasizes the fundamental importance of understanding the sun, since we can hardly trust our modeling of other stars if the solar model is in doubt.

Three-Degree Background Radiation

During the first few minutes of the Big Bang, temperatures were billions of degrees, and energy appeared in the form of intense, indistinguishable fields of matter and radiation.[2] As the universe expanded, the temperature decreased and the wavelength of the radiation increased—both in direct proportion to the expanding scale of the universe. After some 100,000 years, the temperature had fallen to about 4,000°K (degrees Kelvin = degrees Celsius + 273.15), and the intensity of the radiation was no longer great enough to prevent atomic nuclei from capturing electrons and thereby becoming neutral atoms. At the time of this "decoupling" of matter and radiation, the radiation was no longer scattered by free electrons, and the universe became transparent. As though a fog had been dispelled, the universe was filled with blinding light.

Since the time of decoupling, the universe has expanded by about 3,000 times. As a consequence, the wavelength of the radiation present at that early time should now be increased by the same factor of 3,000 to a present wavelength of about one millimeter. The spectrum of such radiation—a cool, fading afterglow of the primordial fireball—would have a characteristic temperature of about 3° above absolute zero (3°K). This relic radiation was actually discovered in 1964–65, and its existence is perhaps the strongest evidence we have for the Big Bang picture of the universe. The 3°K background radiation has since been observed with remarkable precision in many different experiments: from the ground, from balloons, and even from a U-2 aircraft. Even though the intensity of this background radiation is only about a hundred billionth of the thermal radiation from this sheet of paper, it accounts for most of the radiant energy in the universe, with an energy density exceeding that of all starlight and cosmic rays.

The universal background radiation appears to be remarkably uniform in all directions, with the exception of a slight systematic variation that can be attributed to the motion of the solar system through the universe. Measurements indicate that the sun and its planets are moving toward a point in the constellation of Leo at a speed of about 390 kilometers per second relative to the background radiation.

A satellite mission, the Cosmic Background Explorer (COBE), is now being prepared to make refined measurements of all aspects of the background radiation. Results of this mission should provide important new knowledge of the nature of the early universe.

HIGH-ENERGY ASTRONOMY

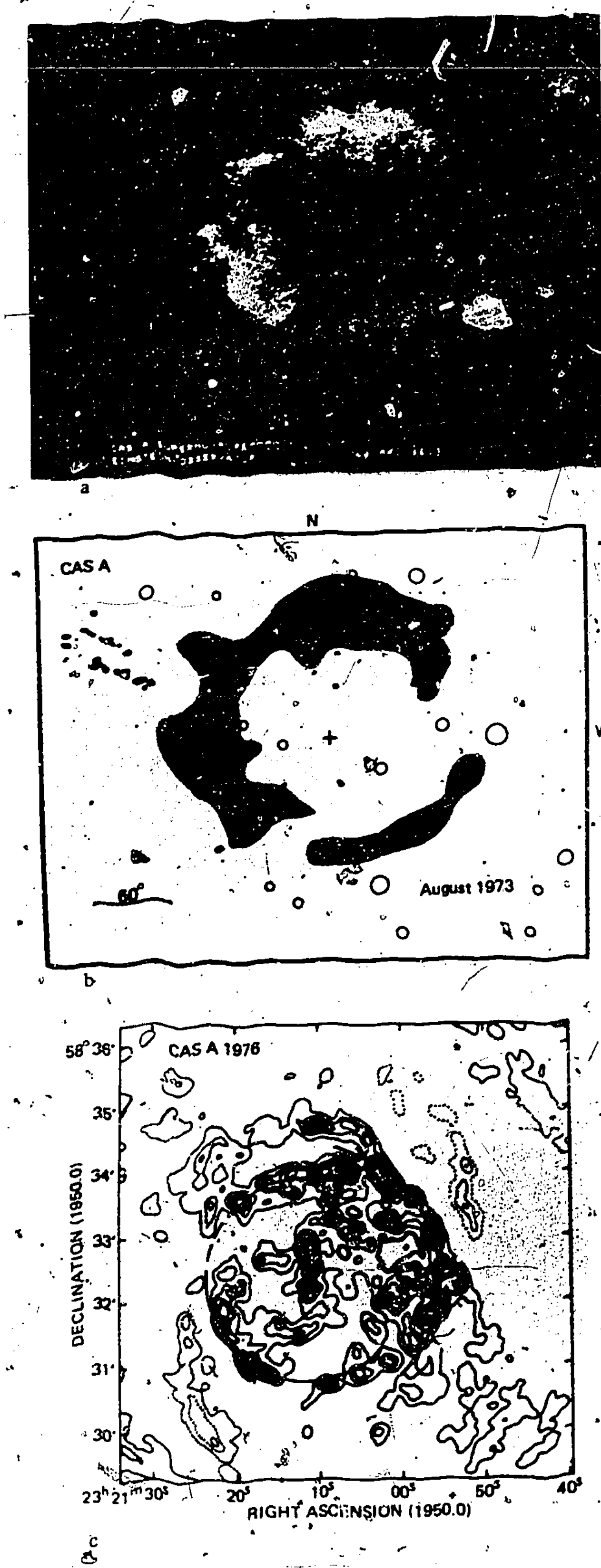
The branches of astronomy that are based on space technology are still in their pioneering stages; therefore, great improvements in observational capabilities can be expected within the near future. This is particularly true of high-energy astronomy, now the most rapidly developing of the new astronomies. High-energy astronomy's principal window to the cosmos, the X-ray region of the electromagnetic spectrum, provides a view of many of the most bizarre and violent phenomena on the celestial scene (Figure 11). Spectacular variability is almost universally characteristic of X-ray sources, with time scales ranging from milliseconds to years. As an example, there are now more than 30 known sources called X-ray bursters that typically flash at a power level of a million suns for about 1 second, but are quiescent 10 seconds later. One such source, known as the rapid burster, repeats its explosions about a thousand times a day, going off like a string of cosmic firecrackers.

X-Ray Stars

The X-ray sources within our own galaxy include white-dwarf stars, neutron stars, and (theoretically) black holes. Each of these three kinds of curious objects is believed to be an alternative end-point to the process of stellar evolution. When a star has exhausted its nuclear fuel and can no longer generate the central gas pressure that had previously sustained the crushing burden of its overlying mass, the star begins to shrink inward under the force of its own gravitation. For stars of relatively small mass, such as our sun, the shrinkage ends with the formation of a white dwarf, which is no larger than the earth but is so dense that a spoonful of its matter would weigh about a ton.

Neutron Stars

The end point for stars of medium or large mass can be either a neutron star or a black hole. In the first case, the inner core of the star is crushed down in a fraction of a second to form a neutron star, and simultaneously the star's outer layers are blown off in a gigantic explosion called a supernova. A neutron star is essentially a giant nucleus of about 10^{57} tightly packed neutrons in a volume only 10–20 kilometers in radius and with a density of about a billion tons per cubic inch. In its collapse, the star's original magnetic field can be concentrated to perhaps a million million times its initial strength and its rotation speeded up to



a few or even tens of revolutions per second. Electrons locked in the star's magnetic field can be whipped around the star at nearly the speed of light, with the result that they beam out radiation of every wavelength from radio waves to gamma rays, which is called "synchrotron radiation" for reasons given later.

Pulsars

It is believed that radiation is more intense from some regions on the surface of a neutron star than from others. Thus, as the star rotates, the radiation appears to vary in intensity, or pulse, depending upon whether these regions or "hot spots" face the observer on earth or not. Such pulsating sources of radiation, or pulsars, were first discovered in 1967 and now number several hundred. The first X-ray pulsar detected was the neutron star at the center of a supernova remnant, the Crab Nebula. The Crab pulsar emits strongly in every region of the electromagnetic spectrum. No other pulsar has been observed in both the radio and X-ray regions, although a pulsar in Vela emits radio waves and gamma rays quite strongly. Two recent discoveries of pure gamma-ray pulsars add to a diversity of pulsars that we cannot yet explain. The pulsar process must be more complex than the simple description given above.

Binary Systems

When a neutron star is coupled closely to a normal star to form a binary pair, it can accrete gas by intercepting the "stellar wind" of particles thrown off by its larger companion or by direct gravitational attraction from the companion's surface. The powerful gravity of the compact star attracts the gas toward its surface with great kinetic energy, and the hot infalling gas is funneled by the star's magnetic field onto an area of only about one square kilometer at each magnetic pole, which thus becomes a source of intense X-rays. If the magnetic axis of the star is not the same as its axis of rotation, the "hot spots" at the magnetic poles will sweep around the sky and thus beam X-rays in pulsar fashion as the star spins. At present, some 14

FIGURE 11 Supernova remnant Cassiopia A as seen at different wavelengths. (a) X-ray picture from Einstein Observatory by S. Murray (Center for Astrophysics, Cambridge, Mass.). (b) Drawing of regions of optical emission as photographed by Kamper and Van den Bergh (David Dunlap Observatory, University of Toronto). (c) Radio map at 4,995 MHz as measured by A. R. Bell (Mullard Radio Astronomy Observatory, Cavendish Laboratory, Cambridge, England).

X-ray pulsars are known, with rotation periods ranging from less than 0.1 second to 16 minutes.

Black Holes

When the nuclear burning cycle ends, the gravitational forces for stars of sufficient mass become so strong that no equilibrium condition such as that of a white dwarf or a neutron star is possible. In such cases, the final collapse is thought to occur in less than a second and to result in the formation of a black hole—a region of space in which matter, figuratively speaking, has been crushed completely out of existence. The name was chosen to express the fact that theoretically nothing, not even light, can escape from the surface of this hole in space because of its immense gravitational field. Black holes resulting from the collapse of stars would typically have masses 3–50 times that of the sun, and diameters 18–300 kilometers.

The gravitational force is believed by most scientists to be correctly described by Einstein's general theory of relativity, and the idea of black holes is so fundamental a consequence of this theory that astronomers look for every possible clue to confirm their actual existence.[3] Since no radiation can escape from the black hole itself, we must search for effects produced outside the hole by its powerful gravitation. A black hole creates a kind of gravitational whirlpool in space that draws any nearby matter toward it. The combined centrifugal and gravitational forces cause the swirling particles to form a flat, gaseous accretion disk that can be millions of kilometers in diameter. Within the disk, frictional forces cause the individual bits of matter to spiral gradually inward until they are finally swallowed up by the black hole. The frictional forces at the inner edge of the disk heat the swirling gas to such high temperatures that as much as 80 percent of their thermal energy is radiated away as X-rays.

Astronomers have so far identified four X-ray sources which are likely candidates for black holes. The X-ray emission from these sources has the noisy, flaring character that would be expected from the intensely turbulent inner region of the accretion disk. In this inner region, any hot spots that develop will beam out exceptionally intense X-rays as they orbit around the hole; and since the orbital period, or time per revolution, is in the range of only a few thousandths of a second, X-ray astronomers are searching for trains of pulses that recur with such short periods. Observation of such X-ray pulses would be strong evidence for the existence of black holes and would reveal important properties of the gravitational "warping" of space close to the hole.

HIGH-ENERGY EXTRAGALACTIC SOURCES

The sources of X-rays, gamma rays, and neutrinos beyond our own galaxy include the hot gas that forms a diffuse medium in clusters of galaxies, and also several kinds of discrete sources such as the nuclei of unusually active galaxies.

Cluster Gas

Most of the observed galaxies are found grouped with others in clusters of various sizes (Figure 12). There is also evidence for a clustering of clusters. Such superclusters or groupings represent the largest aggregations of matter in the universe. Our own Milky Way is a member of a local group consisting of 21 galaxies, while the largest galactic clusters, such as those in Virgo and in Coma, contain thousands of galaxies, all gravitationally bound together and traveling at high speeds within a roughly spherical volume that is not much larger than that of the local group.

The space between galaxies in such clusters appears to be filled with a diffuse gas that can be heated to very high temperatures in several different ways including gravitational infall. The X-rays emitted by this hot gas carry information about its composition, density, and temperature, and thus about the dynamics and evolution of the cluster as a whole. On the largest cosmic scale, there is some preliminary evidence that superclusters may also be enveloped in enormous clouds of hot gases emitting X-rays. As stated previously, early studies have tended to indicate that the average density of matter does not appear to be great enough to halt the universal expansion and thus to form a closed universe. However, the matter present in the diffuse gas clouds in clusters and superclusters may account for a great deal of mass that was previously invisible but that now can be seen in the light of its X-radiation.

Discrete Sources

A number of localized sources rival the richest galactic clusters in the intensity of their X-ray emission. These include the central regions, or nuclei, of certain galaxies in which cataclysmic events appear to be occurring—Seyfert galaxies and others which radiate strongly at radio frequencies—and the still more powerful starlike quasars and related BL-Lacertae objects. These sources were identified during the first major X-ray astronomy mission, HEAO-1, but their detailed study and the discovery of other such sources must await more sensitive instrumentation of the kind described in the last part of this section.

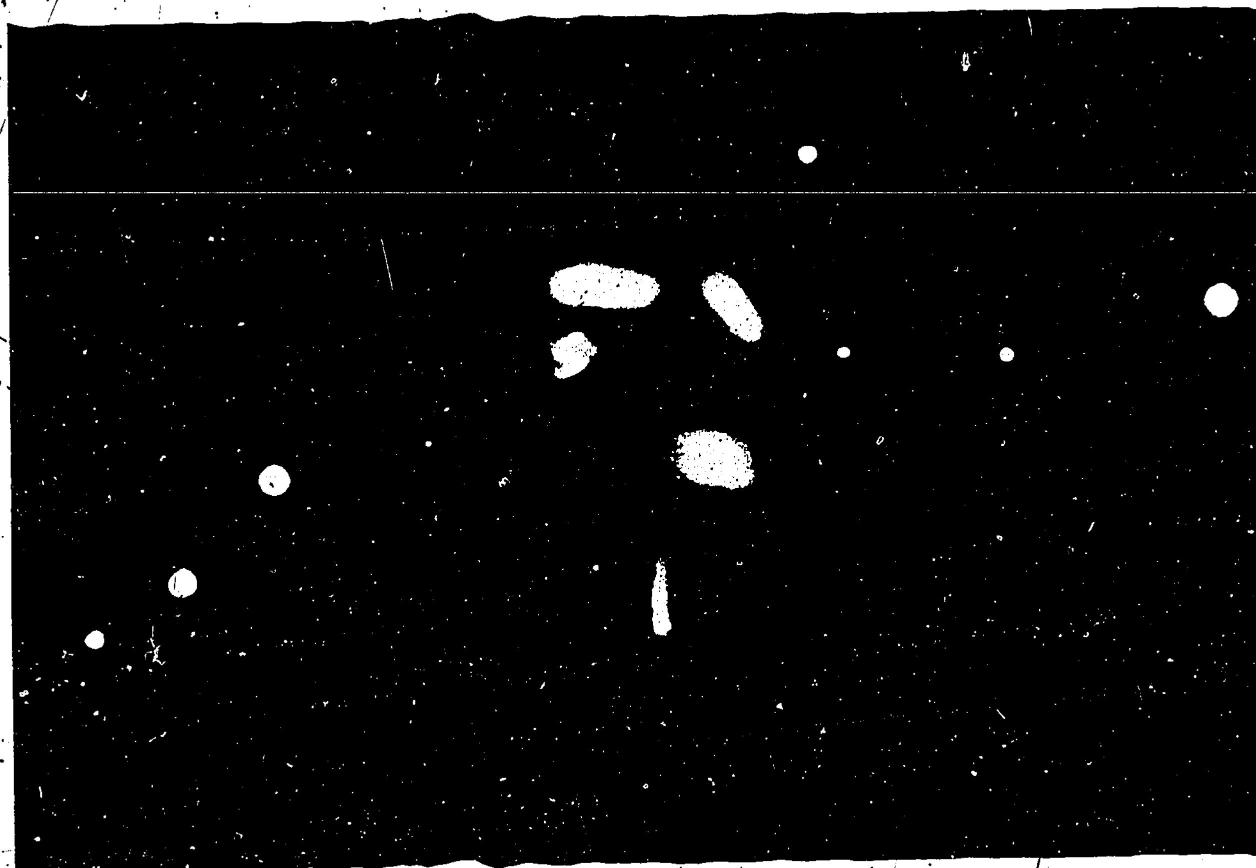


FIGURE 12 Group of five galaxies in Serpens with unusual connecting clouds. (Hale Observatories)

GAMMA-RAY AND NEUTRINO ASTRONOMY

Gamma-ray astronomy has developed more slowly than X-ray astronomy because the required instruments are intrinsically much larger and heavier. Even though the instruments that have been used so far in balloons and in small satellites have had only limited sensitivity, important results have already been achieved. Gamma-ray emission is generated primarily within the disk of our own galaxy, and it reveals the structure of the disk and the interactions of cosmic rays with the ambient interstellar medium. Some two dozen point-sources have been found so far, but few of these correspond to sources observed at other wavelengths. Exceptions are the pulsars in the Crab and Vela, plus several other pulsars. It comes as a great surprise that these sources generate so much of their energy as gamma-ray pulses.

Since many processes that generate gamma radiation will also generate neutrinos, these new observations suggest that neutrino astronomy holds great promise for further exciting developments. Neutrino astronomy divides itself into three energy regions: low energy neutrinos from stars (about 1 MeV or million electron volts); intermediate energy neutrinos (about 10 MeV) from gravitational collapse; and ultra high-energy neutrinos (greater than 1 GeV or billion electron volts). In the Homestake gold mine experiment to detect low-energy solar neutrinos not as many were counted as were expected, as mentioned before. Future experiments are being planned to resolve this discrepancy.[4]

It is expected that in the final stellar collapse to form a neutron star or a black hole more energy will be radiated in the form of neutrinos than anything else. Currently, experiments are operating in the Homestake gold mine and in the Soviet Caucasus that should see the intermediate-energy neutrinos from a collapse anywhere in our Galaxy. The ultra-high-energy neutrinos are produced whenever high-energy protons collide anywhere in the universe. There are both discrete sources such as quasars and pulsars and diffuse sources such as the collisions between cosmic rays and the interstellar gas. Limits on these sources are beginning to be obtained by the detectors used in the intermediate energy range; however, definitive measurements with angular resolution capable of picking up point neutrino sources will probably have to wait for *dumand*, the *D*eep *U*nderwater *M*uon *a*nd *N*eutrino *D*etector, for which detailed design studies are just beginning. *dumand* would consist of a cubic kilometer array of detectors located approximately six kilometers underwater.

One interesting spinoff of neutrino astronomy is that these large arrays of detectors may provide the best means to determine limits on proton lifetimes. This is a crucial question in the grand unification of theories to be mentioned later. Also, proton decay may even be related to the origin of matter in the Big Bang.

HYPERACTIVE NUCLEI OF GALAXIES

Quasars were recognized in 1963 as possibly the most energetic objects in the universe. So great are their luminosities, or total energy emission, that theorists speculated about the need for new physics to explain them. The hundreds of quasars and other quasarlike objects that have been discovered since 1963 appear to form a general class of violently active extragalactic objects in which the light of stars is overwhelmed a hundredfold or more by nonstellar light. As the catalog of quasars and other active galaxies (Seyfert galaxies, radio galaxies, etc.) has expanded, a sense of evolutionary continuity relating all of these objects has begun to emerge. The common link is believed to be a gravitational concentration of enormous masses of stars and stellar debris in the central regions, possibly into black holes having masses millions or even billions of times that of the sun.

Quasars are optically brilliant objects, and many are also powerful radio, infrared, and X-ray sources. Even though the energy source in such objects may be concentrated in a region only a light day across, the associated radio structure sometimes extends over several million light years. The power output of the energy source is as startling as the rapidity with which it can vary. As an example, in 1975 a certain quasar flared to unprecedented brilliance in both the optical and radio bands. Within a few weeks its brightness had risen to 10,000 times that of the entire Milky Way galaxy and just as quickly had subsided.

Since a black hole is intrinsically an absorber rather than an emitter of matter and energy, how can it produce the enormous power of a quasar? The mechanism is thought to be the capture by a massive black hole of the gaseous debris of stars that have collided in the densely packed central core of a galaxy, or of stars that have come so close to the black hole that gravitational tidal forces have ripped them apart. It would require the accretion of only one or two star masses per year for a black hole of a hundred million solar masses to power a quasar, and this process might continue for several million years before the reservoir of stellar material in the nucleus of the galaxy was exhausted. Most of the observed quasars and other powerful radio sources are very far away in distance and thus in time, and this fact can be interpreted as evidence that these immensely luminous objects were much more common when the universe was young, dying out within perhaps a billion years after they were originally formed.

Although the black-hole model of active galactic nuclei has the virtue of being a powerful energy-generating mechanism, this idea is not accepted by all theorists. Perhaps the best that can be said of it at present is that it is the least untenable theory yet offered. The next decade must bring far more sensitive and detailed observations in every part of the spectrum before we can hope to solve the quasar mystery. As an example of the problems involved, recent studies tentatively suggest that the quasar 3C-273, one of the most powerful radiation sources in the entire universe, produces a very large fraction of its energy in the form of energetic gamma rays. This is an astounding and baffling observation.

RADIO ASTRONOMY

The ability of the mirror of an optical telescope to separate two close images, or its resolving power, increases in proportion to its diameter and decreases in proportion to the wavelength of the radiation it is used to detect. Since the wavelength of celestial radio waves is thousands or more times longer than that of optical radiation, it would be totally impractical to build a radio antenna that could match the resolution of even a modest optical mirror. The radio astronomer's solution to this problem is an instrument called an interferometer, which consists of two or more radio telescopes that are separated from each other by a certain distance. When the signals from the two telescopes viewing the same source are combined, the resolution obtained matches that of a single telescope whose diameter equals the distance of separation. The separation does not, of course, affect the instrument's signal-gathering power, which depends only on the combined surface areas of the individual radio dishes.

Early interferometry was limited to telescopes located close enough together to be linked by cable or microwave communication. With the development of very accurate clocks, however, it has become possible to transfer time precisely and to combine, in a computer, signals recorded on tape at widely separated telescopes; a technique called "aperture synthesis." With only the size of the earth to limit telescope separation, very long baseline interferometry (VLBI) has pushed resolution in the radio spectrum to better than a thousandth of a second of arc, far beyond anything achieved optically (one second of arc subtends an angle that is 1/3,600 of a degree). Any problems in matching atomic clocks can be eliminated by transferring phase information between telescopes via real-time satellite communication, a method whose feasibility has been clearly demonstrated with a Canadian satellite.

Plans are now being developed for a steerable three-meter radio dish antenna to be carried aboard the space shuttle and to be used in combination with one or more ground-based telescopes such as the dish located in Arecibo, Puerto Rico. It will provide continuously variable baselines, a great advantage for aperture synthesis.

The taped-signal technique has allowed astronomers to combine VLBI arrays of several existing telescopes to gain more detail in the synthesized data. The gain is related to the possible number of pairings of telescopes: two provide a single pair, three make three pairs, four make six, and five make ten combinations. A more widespread geographical distribution permits a fuller map of the sky, and operations are planned for transcontinental and eventually transoceanic combinations stretching from Massachusetts to Hawaii and south into Texas and Mexico.

INFRARED ASTRONOMY

The infrared astronomer is interested in objects with temperatures less than about 3,000°K, the sun, by example, has a surface temperature of 5,700°K. In the near-infrared region of the spectrum we see very cool stars and the "tail" of the emission from ordinary stars. The stellar emission decreases toward longer wavelengths, and the infrared contribution comes mainly from dust particles heated by nearby stars.

Within the galaxy, the major sources of infrared radiation are regions of ionized hydrogen of great extent and optical brightness: dense molecular clouds, dark nebulae where starlight is obscured by dust that reradiates infrared, and young stars surrounded by dust shells. All of these objects are associated with stellar birth. At the other extreme of stellar evolution, dying stars may eject large amounts of dust and become very bright in the infrared. The mass loss may take the form of a slowly growing dust shell around a cool star or of an explosive ejection accompanying the outburst of a nova or supernova. Such reprocessing of material to the interstellar medium has an important role in the evolution of the galaxy.

The center of our galaxy is especially bright in the infrared, comparable, in fact, to the visible luminosity of the entire galaxy. Thermal emission by dust must be the source of this infrared radiation, but the heating process is not understood. Many distant galaxies radiate infrared from their nuclear regions at 10,000 times the luminosity of our galactic center. The mechanism involved must be closely connected with the total process of hyperactivity in galactic nuclei, but it is again not clearly understood. In fact most quasars, Seyferts, and other extragalactic objects radiate more energy in the infrared than in any other part of the electromagnetic spectrum. This means that infrared astronomy may eventually be one of the most effective ways to study these exciting objects.

NEW INSTRUMENTS

Space Telescope

Scheduled for launch on the Space Shuttle in 1983, the space telescope will offer such fundamental advantages over ground-based instruments that it will become the main tool for the deep-space studies that are essential to cosmology. The power of a ground-based telescope to detect and resolve very faint objects is ultimately limited, not by the quality of its optics, but rather by the shimmering of the earth's atmosphere and the background light of the atmosphere's airglow. A location in space makes possible the sharpest imagery and also extends the range of possible observations into the ultraviolet and infrared regions of the spectrum. The space telescope will provide a resolution of about 0.1 to 0.05 of an arc-second, and will detect objects as faint as the twenty-eighth or twenty-ninth magnitude—at least 100 times dimmer than can be observed from the earth's surface. The high cost of placing the 2.4-meter mirror and all of its accessory equipment in space requires that the space telescope be serviced and its instruments updated over a period of 15 years, and perhaps even a decade. This concept has become feasible only with the advent of the space shuttle.

Ground-Based Telescopes

The largest ground-based telescopes now operating are the five-meter reflector on Mount Palomar and the six-meter reflector on Mt. Semirodnika in the Soviet Union. It is doubtful that larger mirrors could be successfully cast or that such mirrors could be mounted to operate without distortion. Instruments consisting of multiple mirrors are not limited by these constraints. Now nearing completion atop Arizona's Mount Hopkins is the first large Multiple Mirror Telescope, the prototype of the next generation of very large optical telescopes. This instrument will use six 1.8-meter mirrors to match the light-gathering power of a 4.4-meter telescope at less than one-third the cost of a conventional mirror and its dome housing. Tests have demonstrated that a laser sensing system can be used to bring all six mirrors into a common focus. Designs are also being studied for multiple mirror telescopes with equivalent apertures up to 25 m. A direct scale-up of a monolithic mirror of such size, even if it were technically feasible, would cost perhaps $2 billion and require some 50 years to complete.

Radio Telescopes

A very large array (VLA) is now nearing completion on the plains of San Augustin near Soccoro, New Mexico (Figure 13). It will be the world's largest aperture-synthesis radio interferometer. With 19 of the 27 antennae now complete, along with about one-third of the trackage and electronics, the VLA already surpasses in sensitivity any radio astronomy facility in the world. The full array of 27 antennae will be ready for operation in 1981. Its agenda will include quasars, radio galaxies, black holes, and stellar and cosmic evolution.

It is also important to move toward shorter radio wavelengths to gain resolving power. A very successful millimeter-wave antenna serves the Five Colleges Radio Astronomy Observatory at Amherst, Massachusetts. Its efficiency will be further increased by the use of cryogenically cooled detectors. At the Owens Valley Radio Observatory of the California Institute of Technology, a mirror 10 meters in diameter has attained a resolution of 27 arc seconds at a wavelength of 1.3 millimeters. This excellent performance was achieved through the development of some remarkable mirror-fabrication techniques.

Infrared Telescopes

As noted earlier, most of the infrared spectrum must be observed from aircraft, balloons, rockets, or satellites. The technology of cryogenically cooled detectors and mirrors is very difficult to perfect, and small cooled telescopes have been used in only a few rocket flights. In 1981, a small helium-cooled telescope will be launched aboard Spacelab 2. Also planned for the same year is the Infrared Astronomical Satellite (IRAS) a joint project of the Netherlands, the United Kingdom, and the United States. It will carry a larger (60 centimeters) cooled telescope and detectors to cover a large fraction of the infrared spectrum. It should be able to detect a photostellar region like the molecular cloud in Orion out to distances of about 10,000 light years and to provide a catalog of perhaps a million infrared sources.

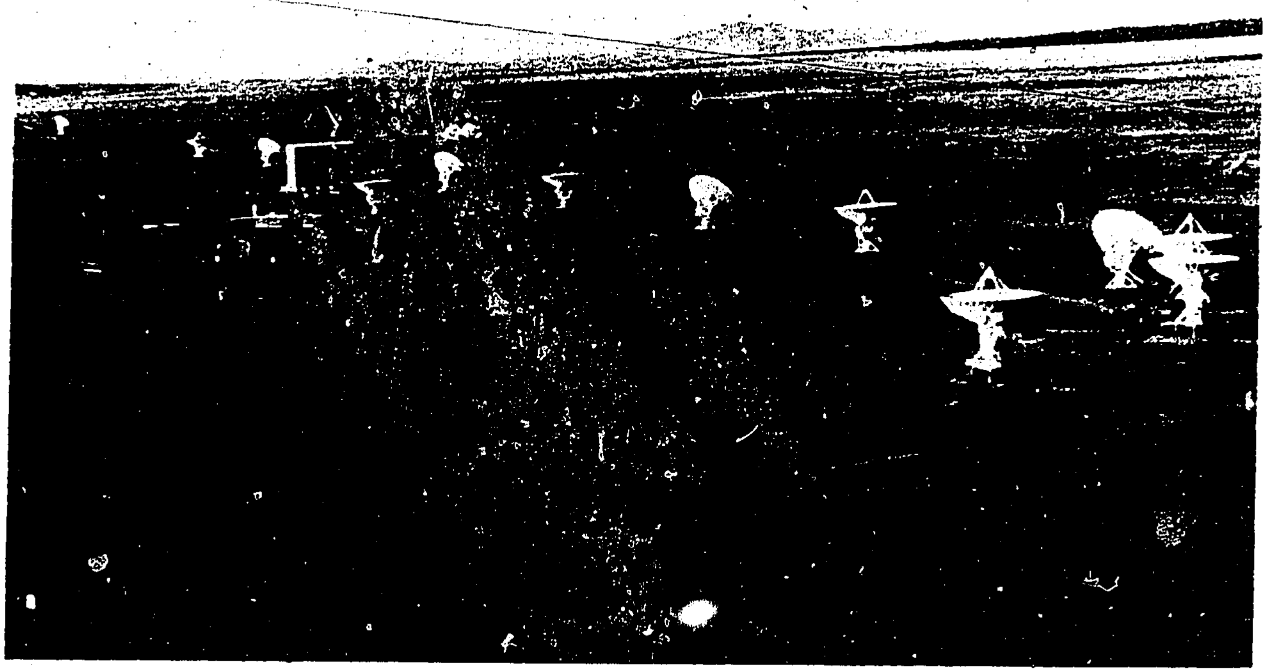


FIGURE 13 Very Large Array radio telescope being built near Socorro, N. Mex., by the National Radio Astronomy Observatory under contract with the National Science Foundation.

X-Ray and Gamma-Ray Instruments

The first High Energy Astronomical Observatory (HEAO-1) has been in orbit since mid-1977 with an array of about 80 square feet of X-ray detectors. The catalog of X-ray sources compiled through HEAO-1 will exceed 1,000, and the new observatories now being planned may extend this catalog to 100,000 or more. The recently launched HEAO-2 (Einstein Observatory) satellite is now carrying the first true X-ray telescope aboard a satellite. Its 60-centimeter diameter nested mirror has special instrumentation for refined wavelength measurements. Since it will be possible to perform only a small fraction of the high-priority observations suited to this satellite during its expected life of two years, X-ray astronomers are already well into design studies of an Advanced X-ray Astronomy Facility (AXAF) for the mid-1980's. AXAF will have a 1.2-meter X-ray telescope launched as a free-flyer from the space shuttle, and its lifetime is expected to be 10–20 years.

A group of instruments has been selected to fly on a Gamma Ray Observatory (GRO) in 1984. The payload of about 15,000 pounds will permit gamma-ray astronomers to develop instruments that are 10 times as sensitive as those of the first generation. The prospect is that GRO will make gamma-ray astronomy a full-fledged partner of the other new astronomies.

CONDENSED MATTER

Condensed matter science is concerned with the electronic and atomic properties of the solid and liquid phases of matter and with the ways in which these substances respond to mechanical forces, to heat, to electric and magnetic fields, and to radiation. The subject area is extremely broad, ranging from questions of a chemical nature concerning natural or synthetic structures, to such topics as the distortions of electronic and atomic symmetry that result in the properties of magnetism and superconductivity. Previous fundamental advances in the field have led to the discovery of the transistor, solid-state masers and lasers, high-temperature superconductors, solar cells, superconducting junction technology for ultrafast computers, and many other important solid-state devices.

The accumulated knowledge of past years has resulted in our present very detailed understanding of simple crystalline solids and of liquids. Indeed, the electronic properties of many crystalline solids are now so well understood that the properties themselves can be used to elucidate the nature of the bonding between the atoms in the material.

All solids and liquids can exist in different phases, that is, different structural arrangements of their atoms and electrons. The most significant conceptual breakthrough in this field during the past 10 years has undoubtedly been the vastly improved understanding of phase transitions. Examples of such transitions are the change from a magnetic to a nonmagnetic phase in solids, and the evolution of a gas phase into a liquid. The key to this achievement has been precise description of the relationships between the scales of length, energy, and time that are characteristic of the particular change in phase. The basic ideas developed in this work are also applicable to other phenomena, for example, to the electrical conductivity of noncrystalline materials.

The past 10 years have also seen many important experimental discoveries in condensed matter science. Among the most significant of these have been the discovery of the superfluid (frictionless-flow) phase of the light isotope of helium, ^3He, and the discovery of charge-density waves in solids. A charge-density wave is a periodic spatial variation of the electron density in a crystal that is different from the spatial structure of the crystal lattice itself—a sort of electron-supercrystal. The possible uses of this new phase of matter are now being studied.

Present work in the field is proceeding on a broad front. There is renewed interest in the study of disordered or amorphous materials and glasses, an area where many fundamental questions concerning electronic and atomic structure remain to be answered. The study of surfaces has also drawn intense interest, in part because of the new experimental techniques that ultrahigh vacuum technology has recently made possible. A good deal of progress has been made in understanding the properties of defects and of grain boundaries in crystalline solids. In addition, new methods of growing crystals have been developed that allow a solid to be built up essentially one atomic or molecular layer at a time. Studies of novel materials have resulted in the observation of complex and fascinating behavior, and some of these new materials have already begun to find practical applications.

In all of this work, progress has been linked closely with the development of more powerful radiation sources and more sensitive detection instruments. These include sources of synchrotron radiation in the ultraviolet and X-ray regions of the electromagnetic spectrum, high-resolution electron microscopes, and intense sources of neutrons and of energetic ions.

AMORPHOUS SOLIDS

The remarkable progress that has occurred during the past 50 years in the understanding and practical application of crystalline materials has been based on the special properties of their regularly repeated (periodic) structure. Although our understanding of the far more complex amorphous materials has been much slower to develop, the past decade has seen an accelerating pace of discovery and a growing awareness of the exceptional importance of amorphous materials for both scientific and technological purposes. The significance of contributions in this field was recognized by the 1977 Nobel Prize in physics.

The arrangement of the nearest neighbors of any particular atom in an amorphous solid is often quite similar to that in the corresponding crystalline material. This regularity does not extend out to more than a few atomic separations, and thus overall periodicity is lost. In contrast to crystals, which have both short- and long-range order, amorphous solids have only short-range order. One consequence of this difference is the fact that many of the limitations on composition and on atomic arrangement that hold for crystals do not apply to amorphous materials, with the result that an enormous range of such materials can be prepared. Thus new material properties become accessible that are uniquely characteristic of amorphous materials, new kinds of chemistry occur, and an entirely different theoretical approach is required for understanding these materials.

The next decade may well see a revolution in materials science and its applications, as work with amorphous solids begins to come to fruition. Three examples may help to illustrate some of these possibilities: Metallic glasses now commercially available have magnetic properties that make them very attractive as transformer-core materials; they also have unusually high resistance to corrosion and radiation damage, or unusually high ductility (see p. 173). Very low loss glass light-guides are now being manufactured that are suitable for communicating by light waves. Amorphous semiconducting films are now being prepared on an experimental basis, which may lead to economically feasible generation of electricity from sunlight.

SURFACES

The study of surfaces involves determination of their atomic and electronic structure and energy levels, the strengths of their chemical bonds, and the interrelationships among these. The goals of such studies include an understanding of surface chemical reactions, of crystal growth, and of the interface (surface-to-surface) phenomena that occur in electron devices. The progress made in this field during the past decade has largely depended upon advances in ultrahigh vacuum technology. This is because clean surfaces are very reactive chemically and thus require a high-vacuum environment if rapid contamination is to be avoided.

Surface experiments typically consist of directing a beam of electrons, ultraviolet radiation, X-rays, atoms, or ions at the surface; then measuring the energies and angular distributions of the particles that are reflected (scattered) back from the surface or knocked out of it. The sample under study can usually be moved to different locations within the vacuum chamber so that it can be prepared, tested, and then studied in an efficient sequence of operations that minimizes contamination.

Theoretical modeling of surfaces has advanced in concert with the development of more sensitive experimental techniques. Recent successes include the deduction of detailed atomic configurations from the scattering of low-energy electrons, and insight into chemical bonding and atomic geometry from the radiation-induced emission of surface electrons. These and several other new experimental techniques are still in their early stages but are now developing rapidly. It is possible to foresee major advances in the understanding of surface processes as the accumulating experimental results begin to reveal the systematic underlying trends. The knowledge gained in this work will clearly have direct application to such complex surface phenomena as catalysis, corrosion, lubrication, and crystal growth, as well as to the interface phenomena in electronic devices and in electrochemical processes.

NOVEL MATERIALS

The study of such relatively simple materials as metals, semiconductors, and magnets was the central focus of condensed matter research during its early years. With the present understanding of these materials, there has been increasing emphasis during the past decade on the search for novel materials that exhibit properties of unusual scientific or practical interest. Examples of such materials are those in which the structure is built up of successive layers or of linear chains of atoms. There are also new materials that consist of rigid clusters of atoms, or of two different groups of atoms in which one of the groups can move relatively freely through the stable lattice formed by the second group. A brief outline of some of the more promising of these new materials is given in the following paragraphs.

Layered Materials

The class of materials known as graphite intercalation compounds consists of layers of carbon atoms that have sandwiched (intercalated) between them alternating layers of other kinds of atoms or molecules. Interest in these compounds was greatly increased by the recent discovery of certain forms that rival copper in their ability to conduct electricity. With a better understanding of the manner in which the intercalated atoms are taken up and ordered between the graphite layers, it may be possible to develop economical new materials of light weight that can perform some of the same functions as metals.

Earlier work based on the use of alkali atoms as the alternate layers in an intercalated structure has resulted in the development of a new class of batteries that will soon be available commercially. In addition to this practical application, these materials were found to have remarkable properties even when the intercalated atoms were not present in the structure. For example, at low temperatures the mutual interaction of the electrons within the material resulted in the formation of spatially periodic arrays of electrons called charge-density waves. In a sense, the electrons formed a crystal structure of their own that was not related in any simple way to the underlying crystal structure of the material itself. In fact, there is some preliminary evidence that this electron crystal can move as a whole within the material under the influence of an electric field.

Another example of a layered material is a compound in which layers of alumina (aluminum oxide) are interleaved with layers of alkali ions that occupy only a fraction of the available sites within the layer. As a consequence of this fractional occupancy, the ions are relatively free to move within the alkali layers, and the ionic motion is rapid enough to allow these materials to be used as solid electrolytes (charge carriers). The study of such materials is a necessary step toward the eventual goal of lightweight and reliable batteries made up completely of solids.

Chains and Clusters

A different approach to the search for new metal-like materials consists of forming organic solids in which chains of molecules that accept electrons are alternated with chains that donate electrons. Work in this area has led to the synthesis of organic solids that are good conductors of electricity. The chain character of these compounds gives rise to a number of different forms of electronic instability, the study of which has resulted in important theoretical insights into the nature of such materials.

Work during the past several years has resulted in the synthesis of a very interesting series of cluster compounds, so called because the electrons within the material are confined to certain clusters of atoms. Such materials often exhibit superconductivity (the flow of electric current without resistance) at higher than customary temperatures. However, when magnetic ions are interspersed within the materials, the incompatibility of magnetism and superconductivity becomes graphically apparent, with the material alternating between superconducting and magnetic phases as each group of electrons in turn comes to dominate the behavior. Because of their potential importance for electrical power transmission and magnetic confinement in fusion reactors, study of these materials will be valuable in the continuing quest to raise the temperature at which superconductivity begins.

There are many other examples of important new materials that might be mentioned here, including the development of metals that can absorb and evolve hydrogen efficiently. In any case, the search for new materials with novel properties is certain to continue during the coming years, perhaps with special emphasis upon organic compounds and upon mixtures of organic and inorganic materials.

ATOMIC AND MOLECULAR ENGINEERING

Traditional methods of crystal growth have primarily involved melting and recrystallizing solids at high temperature or growth of crystals from solution. At present, several new techniques are being developed in which thin films are grown from a controllable beam of atoms, ions, or molecules so that new materials can essentially be engineered a layer at a time. Because of the major advances in analytical tools, primarily electron microscopes, it is now possible to study defects, diffusion, and other microscopic phenomena in much greater detail.

One of the more promising methods for growth of new materials and structures is molecular beam epitaxy. With this vapor-growth technique, single-crystal films can be grown, layer by layer, on an underlying crystalline substrate. Controlled deposition makes it possible to vary the chemical composition in predetermined ways and thus to fabricate unique structures with electrical and optical properties not normally found in nature. To date this technique has been applied mainly to the controlled modification of layered semiconductor structures that have important technological applications. However, one can envision superconducting, metallic, and magnetic structures grown in this fashion. For example, when superconductivity occurs at an elevated temperature in a material, the superconducting phase is usually less stable than other phases, and the material is more difficult to grow. However, it has been demonstrated that epitaxy facilitates the growth of the desired phase so that it appears possible that such unique metastable compounds may be fabricated using this new growth technique.

Along somewhat different lines, the study of diffusion phenomena at surface interfaces has led to the discovery of solid phase compound formation. The diffusion of metals into semiconductors, for example, very often results in compound growth proceeding into the semiconductor. Solid-state epitaxy is thus a new method for producing desired structures at interfaces. Interfacial interactions and diffusion are also important in the oxidation process. Greater understanding of these processes is of considerable technological interest for a variety of metal-oxide-semiconductor devices.

The properties of amorphous semiconductors have also received a good deal of attention during the past decade. Recently, vapor-growth techniques for amorphous silicon and germanium have been developed that allow the controlled modification of these materials in a manner somewhat analogous to their crystalline counterparts. The detailed physical mechanisms responsible for the properties of the amorphous materials are not yet known. As noted above, these materials are of much scientific and technological interest.

Another new technique of atomic engineering is ion implantation. This method involves implanting foreign ions into a host material to a predetermined depth, which depends on the kind of ion used and its energy. This process also leads to structural damage and hence to a disordered layer in the material. This technique has already found uses in the "doping" of semiconductors with foreign atoms and in the production of metastable and disordered structures and of new alloys.

Very thin films, of a thickness comparable to atomic dimensions, exhibit phenomena that are unique to these two-dimensional systems. As an example, ultra-thin metal films serve as an important testing ground for our understanding of metallic conductivity. In addition, thin (two molecules thick) liquid-crystal films are being used to study the phenomenon of melting in systems of two dimensions, while incomplete single layers of inert gases on crystalline substances are also being used to probe these phenomena. Ultra-thin magnetic films also have unique properties. Better control of the growth processes of these thin-film structures promises further advances in these areas.

The work described above represents only a few examples of this rapidly expanding field, which is important to such areas of technology as microfabrication of integrated circuits, development of high-strength materials, and exploration of new techniques for the generation of energy.

SYSTEMS FAR FROM EQUILIBRIUM

The different phases of matter—for example, liquid, crystalline, magnetic, superconducting—are usually studied in systems that are in or near states of thermal equilibrium with their surroundings. It is also interesting to study the behavior of systems in states far from thermal equilibrium, since such situations arise quite commonly under natural conditions. Examples are the response of a gas to a sudden increase in pressure that will eventually cause it to condense to the liquid state, and the evolution of convective-flow patterns in a liquid layer strongly heated from one side. The latter system will eventually reach a state of turbulent flow, the detailed characterization of which remains an important unsolved problem in hydrodynamics. In many cases, the behavior of a system can be drastically altered by making only a small change in some external condition. The subsequent evolution of the system is usually quite complex and has been difficult to predict.

Examples of such instabilities occur not only in phase transitions and hydrodynamic flow but also in the dynamics of chemical reactions and in the high-temperature plasmas used in laboratory thermonuclear fusion devices. The present studies of instabilities in systems far from equilibrium bring together the disciplines of condensed matter physics, chemistry, nonlinear mechanics, and hydrodynamics in quite novel ways. The next five years will see a major new effort at understanding these problems, aided by the theoretical methods developed in the study of phase transitions.

FUTURE DIRECTIONS

The studies of amorphous materials, surfaces, novel materials, atomic engineering techniques, and systems far from equilibrium described in the previous sections are among those that are easily identified as ripe for progress during the next five years. However, these selected topics provide only a sampling of the much broader range of past achievements and future research opportunities that exist in condensed matter science. The exact directions that the field will take during the next five years are of course impossible to predict, but rapid progress can reasonably be expected in both scientific understanding and technological application. The field is marked by increasing cross-fertilization of ideas among the conventional disciplines of chemistry, solid-state physics, materials science, and metallurgy, and this trend can be expected to intensify as the work evolves toward studies of more exotic materials and more complex phenomena.

Condensed matter research shares with the other fields of study on the structure of matter a continuing development toward instrumentation of greater power and sensitivity, and thus of greater cost. An inevitable consequence of this development is the trend toward centralization of major research facilities at fewer laboratories, and this in turn has led to important changes in the ways in which the work of the field is carried out.

MOLECULAR AND ATOMIC STRUCTURE

Since the familiar macroscopic forms of matter are made up of molecules, and molecules are made up of atoms, studies of these fundamental building blocks of matter are among the most important in all of physical science. The entire field of chemistry is concerned in one way or another with the structure of molecules. Chemical studies provided the earliest information about molecular form, and the continuing evidence provided by such studies remains very important. Subsequent to the classic experiments of Rutherford establishing the model of the nuclear atom, much of the information about the structure of atoms was derived from observations of the characteristic patterns, or spectra, of visible light that are emitted by different species of atoms. Such observations marked the beginnings of spectroscopy, which continues to this day as the principal experimental technique for studying the smallest particles of matter.

Since spectroscopic studies will form an important part of the content of this section, it will be useful to recapitulate briefly a few fundamental facts of spectroscopy:

- Atoms and molecules can exist in any of a number of discrete (quantized) states, each of which represents a different energy level of the particular system in question.
- These states correspond either to rotational motion of the system, or to vibrational motion, or to different configurations of the electron cloud around an atom or a molecule.
- Changes from one state to another are called transitions, and each transition is accompanied by either the absorption or emission of electromagnetic radiation of a precise wavelength (or frequency).
- The absorption and emission of radiation occurs in distinctive patterns (spectra) that can often serve to identify both the specific kinds of atoms or molecules that are involved and the particular kinds of transitions that have occurred.

MOLECULAR SPECTROSCOPY WITH SYNCHROTRON RADIATION

The most important recent advances in the spectroscopy of molecules in the ultraviolet and X-ray regions have resulted from the use of new synchrotron radiation sources. Electron synchrotrons and storage rings provide a variable-frequency X-ray source whose intensity is about a million times that of a conventional X-ray tube.

Fluorescence Spectroscopy

The variable-frequency feature of synchrotron radiation is particularly helpful in fluorescence spectroscopy, which makes use of the radiation that is re-emitted after absorption of X-rays by heavy elements in the molecule. This technique can distinguish the fluorescent radiation from different elements, thereby greatly increasing the sensitivity and reliability of detecting elements that are present in extremely small proportions. Also, since the synchrotron radiation is produced in precisely repeated, very sharp pulses, it can be used to measure accurately the lifetimes of excited states as short as a nanosecond (one billionth of a second).

EXAFS *Studies*

The technique known as EXAFS (extended X-ray absorption fine structure) has been developed rapidly into a useful technique for molecular-structure research because of the availability of synchrotron radiation. EXAFS studies involve the absorption of an X-ray photon that ejects an electron from the absorbing atom. The details of this process depend on the local environment of the absorbing atom in a molecule and thus provide information about short-range molecular structure, even when the structure is not highly ordered. An example is the identification of the elements surrounding a metal atom in a noncrystalline biological system. Very small amounts of metal play a central role in the function of important biomolecules (for example, iron in hemoglobin). The capabilities of EXAFS also hold out great promise for a deeper understanding of catalysts, those substances that greatly accelerate chemical and biological functions without appearing to react themselves. The role of catalysts in chemical reactions is of course vitally important for industrial chemistry as well as biochemistry.

MOLECULAR SPECTROSCOPY BY MAGNETIC RESONANCE METHODS

Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) methods make use of the fact that the nuclei of many kinds of atoms are small magnets that can take on any of several different orientations in the presence of an applied magnetic field. Transitions from one orientation to another can be induced by radio waves, and the particular frequencies at which such resonance transitions occur serve to identify important elements of molecular structure.

As an example, hydrogen atoms are a significant structural element in organic substances, and the frequency at which these atoms reorient or flip is quite sensitive to the arrangment of their neighboring atoms in a molecule. After some years of experimental study, an enormous body of NMR information has been collected that correlates specific NMR frequencies with particular structural units of organic molecules. As a result, it is now possible to identify these units quickly in other molecules simply by seeing their characteristic frequency "signatures" in an NMR spectrum.

Carbon is also an essential element in organic substances, but its common isotope, carbon-12, has no magnetic moment and thus does not resonate. There is, however, a rare isotope, carbon-13, which does have a magnetic moment and can therefore be used to identify structural elements by the NMR technique. Similar methods can be used to examine the surroundings of other kinds of atoms through the use of suitable isotopes of fluorine, phosphorus, nitrogen, and about 100 other isotopes. Structural problems that once required years of study by chemists can now be disposed of in an afternoon with the aid of an NMR spectrometer.

Because of its particular value for organic chemistry, NMR plays an important role in illuminating biological function. It has been used, for example, to elicit the role played by magnesium atoms in chlorophyll and has also provided valuable insights into the behavior of enzymes, the structure of DNA, RNA and hemoglobin, and so on. NMR is also a powerful tool for monitoring the synthesis of organic compounds; this may in fact be its largest single use.

Electron Paramagnetic Resonance

Electron paramagnetic resonance (EPR) is a spectroscopic technique that closely parallels NMR because electrons are also tiny magnets that can take on either of two orientations in the presence of an applied magnetic field. Molecules that contain an extra electron (i.e., not paired with another electron) can therefore exist in either of two energy states in such a field. The energy difference between the two states is much greater than that of atomic nuclei because of the much stronger magnetic moment of the electron, which means that the resonance frequency is correspondingly higher (in the frequency range of microwave radar).

The EPR method has been widely applied to problems in physics, chemistry, and biology. One of the most important areas of application is called spin-labeling. Although molecules containing an unpaired electron are usually very reactive, those known as free radicals are stable. When a stable free radical is attached to another molecule, the result is a spin-labelled system that is detectable by EPR. A spin-label attached to a component of a cell membrane, for example, provides a direct probe of the membrane's structure. From the characteristics of the EPR signal, it is possible to deduce where in the membrane the spin-label is located, its rate of migration, and much additional information. The use of EPR with spin-labels has probably been the most informative nondestructive technique applied in recent times to the study of the structure and function of membranes in living organisms.

There are many important biological and biochemical processes during which free radicals are formed. An example is green plant photosynthesis, the process by which the energy of sunlight is used to convert carbon dioxide and water into the carbohydrates, proteins, and fats upon which all animal life, including that of man, depends. EPR studies have been used to demonstrate that the primary light-conversion step in photosynthesis involves a special pair of chlorophyll molecules. As a result, there has been considerable success in replicating natural photoreactive chlorophyll, and synthetic chlorophyll special pairs are now being examined for their possible application to solar energy conversion.

LASERS AND MOLECULAR STRUCTURE

A laser is a very intense source of light of a single wavelength.[5] This is made possible by the fact that atoms in the same excited state can be stimulated by light to emit light of the same wavelength. Laser light makes it possible to distinguish effects associated with very slight differences between molecules; for example, the effects of replacing one atom in a molecule with a different isotope of the same element. This is not only a powerful tool for high-precision structural studies but also a possible practical technique for separating the different isotopes of a single element.

Picosecond Spectroscopy

Many events of great scientific importance take place in an incredibly short period of time. The fastest ordinary chemical reactions occur in about 10^{-12} seconds, or a picosecond. The lifetimes of many intermediates in chemical reactions are also of this magnitude, and the rate at which energy is transferred from one part of a molecule to another is again on the same time scale or even faster. Thus the picosecond spectroscopy made possible by the laser has become an extremely important experimental technique. Experiments can be conducted with intense laser pulses only a few tenths of a picosecond long, so that one can record a flash spectrum of the molecule, and from that spectrum construct a picture of the structural changes taking place during these very brief periods in its life. This gives entirely new insights into chemical processes.

Perhaps the most interesting work of this kind concerns molecules of biological interest. For example, it has long been a mystery how the energy is transferred within molecules of chlorophyll to bring about the process of photosynthesis. When light is absorbed by chlorophyll, a small fraction of it is re-emitted (fluorescence) as a result of the energy-transfer mechanism, and this makes it possible to study the mechanism by means of picosecond spectroscopy. Although the interpretation of the results in hand is not yet unambiguous, the data have helped to define the problems of energy transfer more clearly, and the field is now being vigorously explored.

LASERS AND ATOMIC STRUCTURE

As previously stated, lasers can produce light of unprecedented intensity, directionality, spectral purity, and coherence; and certain kinds of lasers can be tuned over a broad range of wavelengths. These properties make lasers very powerful tools for probing the intricate structure of atoms. The extreme accuracy attainable in such studies also serves to test the validity of fundamental physical principles. In the following paragraphs we give only a few examples of the achievements of recent laser experiments.

Giant Atoms

Lasers are now being used to study atoms in which one of the electrons has been raised to an extremely high excited state and thus moves in a very large orbit. Because the volume occupied by these giant atoms can be a million or more times that of the unexcited atom, they exhibit vastly different properties. Such giant atoms occur naturally in regions of space where stars are believed to be forming, and also in such important laboratory environments as those associated with fusion experiments. Knowledge of the structure of giant atoms is thus expected to have both basic and applied significance.

Tests of Physical Principles

The precision of laser experimentation has already resulted in considerably more accurate determinations of two fundamental constants of nature: the speed of light, and the characteristic energy with which electrons are bound to atomic nuclei (the Rydberg constant). There are also in progress stringent tests of the time-dilatation effect (the slowing down of clocks) of the special theory of relativity, and of the basic electromagnetic theory called quantum electrodynamics (QED). In addition, a particularly interesting synthesis of atomic physics and elementary particle physics is now occurring in experiments at several laboratories, which are designed to test the landmark theory that posits the fundamental unity of the weak and electromagnetic interactions. The evidence from high-energy physics experiments is strongly supportive of this unified theory, but so far the results from the "table-top" atomic physics experiments now in progress have been indecisive.

TRAPPED IONS AND ELECTRONS

Since the early recognition of the atomic nature of matter, scientists have sought ways to isolate individual atomic and subatomic particles for detailed study. The problems involved in implementing such confinement schemes are formidable, but techniques that have been evolved during the past two decades have now led to practical realization of the goal. One of the new techniques makes use of a combination of electric and magnetic fields to confine either electrons or ions for periods as long as several days.

By studying the response of a single electron successfully confined for many hours to changes in the confining electromagnetic field, it has been possible to carry out the most precise measurement ever made of an intrinsic property of an elementary particle (its magnetic moment), and thereby to provide the most rigorous testing yet achieved of the theory that describes the electromagnetic interactions of charged particles—quantum electrodynamics.

The recent development of a novel method of radiative cooling holds promise of extending such measurements to even greater precision. One can also foresee the use of such trapped, low-temperature ions in practical applications such as frequency standards, where the frequency stability could be expected to exceed that of present devices by a factor of 10 or more to perhaps 1 part in 10^{17}.

EXOTIC ATOMS

Exotic atoms are atoms in which one, or more, of the usual constituents has been replaced by a particle not ordinarily found in matter because it is unstable. Since the lifetimes of such particles are at most only a few millionths of a second, they must be created by particle accelerators, then quickly trapped to form the desired exotic atoms. In this work the techniques of high-energy physics are combined with the precision spectroscopy of atomic physics to yield important information about the basic forces of nature.

Several kinds of exotic atoms have already been studied, and other kinds are possible. In what follows, however, we limit ourselves to the description of only one example.

Muonic Atoms

The particle called the muon appears to be exactly like an electron in every respect except its mass, which is about 200 times greater. When a muon replaces an electron to form a muonic atom, the greater mass forces it to take up an orbit in the atom that is about 200 times smaller. This has two important consequences: (1) Since the muon is moving in a region where it is affected almost solely by the positive electric field of the nucleus, its observed behavior can be compared with the very precise calculations that are possible as another important means of testing the validity of basic electromagnetic theory. (2) In heavier atoms, the muon's orbit becomes so small that it actually spends a large fraction of its time within the nucleus itself, and in so doing it probes both the size of the nucleus and the distribution of its electric charge.

FUTURE DIRECTIONS

It was not so long ago that many scientists considered atomic and molecular spectroscopy to be a closed subject, still valuable for its many practical applications, but no longer of direct scientific interest in itself. That this judgment was premature should be abundantly evident from the foregoing descriptions. In fact, the important developments in technique and instrumentation in recent times are only beginning to be exploited; and if the new facilities that will be needed to capitalize upon the exciting research opportunities can be developed, the prospect seems excellent for major advances in our understanding of atomic and molecular structure, and for important new practical applications. (The plans and need for new facilities are described in the Research Instruments section of this chapter.)

To conclude this section on molecular and atomic structure, we speculate briefly about two general lines of research that represent goals to be aimed for or limits on what may ultimately be achievable in this field of study.

Imitations of Life

It is instructive to compare some of the physical and chemical processes carried out by living organisms with those of modern technology. As an example, living cells are able to synthesize incredibly complex molecular structures at ordinary temperatures and pressures, and starting with very simple reagents. In contrast, chemical technology typically employs high temperatures and pressures, and violently reactive agents. As illustrations, compare the manufacture of rubber by a rubber tree with the corresponding brute-force industrial processes, or the heroic commercial methods required to fix atmospheric nitrogen with those of nitrogen-fixing bacteria. One can speculate that a more complete understanding of these biological processes at the molecular level may eventually enable us to mimic chemical processes in living organisms and thus to reap the benefits of their high efficiency and great synthetic power.

Molecular Electronics

In an even more speculative vein, there are fascinating long-range possibilities in what has been called molecular electronics. The ultimate goal here would be to synthesize molecules that would act as individual circuit elements, such as conductors, resistors, capacitors, etc.; and then to combine these elements into amplifiers, memory devices, and so on. Such devices would serve the same purposes as present-day electronic devices but would be very much smaller in size and in power consumption, and very much faster in operation.

The example of electrical conductors can serve to illustrate these ideas. The chemical processes that constitute the metabolic activity of living cells involve the transport or movement of electrons as an essential feature. However, nature does not use small bits of metal wire to conduct electrons from one place to another. Instead, there is a carefully graded series of electron-transport proteins within the cell to serve this purpose. (Two such proteins are cytochrome and ferredoxin.) With a more complete understanding of how these proteins carry out their electron-transport tasks, it may become possible to develop practical electron conductors on the molecular level.

RESEARCH INSTRUMENTS FOR CONDENSED MATTER AND MOLECULAR AND ATOMIC STRUCTURE

An important theme that has run through this discussion of the sciences of condensed matter and molecular and atomic structure is the way in which research has been furthered by the development of new and more advanced instruments. The use of more powerful radiation sources and more sensitive detection instruments leads to increased understanding of the structure and atomic dynamics of molecules and condensed matter.

X-rays have provided the classical technique for studying the structure of systems that have atomic spacings characteristic of solids, liquids, and molecules, because these spacings are of a size comparable to X-ray wavelengths. X-ray measurements detect the positions of heavy atoms and also, along with measurements using ultraviolet radiation, yield valuable information on electron structure. In more recent years, neutrons have been used for similar purposes because they, too, have a comparable wavelength and provide a particularly sensitive means for determining the positions and momenta of the lightest atoms, such as those found in organic materials.

Circular electron accelerators (synchrotrons) and storage rings, such as those designed for high-energy physics research, are the most intense sources of radiation in the ultraviolet and X-ray regions because electrons traveling at high speeds in a curved path emit a broad spectrum of electromagnetic radiation.

Proton accelerators may be used to produce an intense source of neutrons since the collision of energetic protons with a target of heavy material such as uranium produces a burst of neutrons through a process called "spallation." A pulsed spallation source of this kind is now under construction in England, and a more modest version is being constructed in the United States.

On a smaller instrumental scale, studies of condensed matter and of molecular and atomic structure will continue to depend upon the techniques provided by electron microscopes, ion sources, lasers, and magnetic resonance devices. As before, each of these techniques will be used in structural or spectroscopic experiments to disclose different aspects of matter, but now with the greater sensitivity and resolving power that precision measurements of more complex materials will require.

As an example of this trend toward increasingly elaborate and costly instrumentation, the rather simple and inexpensive nuclear magnetic resonance (NMR) equipment that was commercially available some 10–15 years ago made the NMR technique readily accessible to the individual scientist. In contrast, a complete NMR facility suited to present-day studies may cost as much as half a million dollars and may include components that are not commercially available and that must be developed by the research group for its own specific purposes.

Similar remarks are applicable to what might be described as integrated facilities, that is, facilities in which no single large item of equipment is involved but in which a number of smaller instruments collectively constitute a major facility. For example, some of the most important experiments in surface physics require the use of an array of several different types of surface probes (electron and ion probes, electron microscopes) used under such high-vacuum conditions that not only the measurements themselves but also the preparation of the sample must be carried out in the final vacuum environment. A second example occurs in the field of very small-scale electronics, where a number of different techniques must be brought to bear on the fabrication and study of structures with dimensions on a scale of millionth of a meter or less. The trend toward more costly instrumentation and toward centralization of facilities has important consequences for the style of research in these fields of study.

NUCLEAR STRUCTURE

THE NUCLEUS AND ITS CONSTITUENTS

Most of the apparent structure and behavior of atoms is determined by the diffuse cloud of negatively charged electrons that surrounds the atom's other constituent, the nucleus. However, the mass of the atom is carried almost entirely by the nucleus, since even the lightest nucleus, a single proton, has a mass more than 1,800 times greater than that of an electron. Because the atom as a whole is electrically neutral under normal conditions, its positive electric charge, which is determined by the number of positively charged protons within its nucleus, also determines the matching number of electrons that the atom will have. Thus the specific molecular and chemical properties of matter are ultimately attributable to the large relative mass of each nucleus and to its total positive electric charge.

The chemical elements found in nature exist because the nuclei of their atoms are either completely stable or else so nearly stable that their lifetimes are longer than the time that has elapsed since they were first created. Most of the possible combinations of protons and neutrons that form nuclei are in fact unstable, and one of the important problems in nuclear research is that of understanding the connection between stability and nuclear structure. In a sense it appears to be purely accidental that of the two nuclear particles it is the proton that is nearly stable, while the neutron is not. Had the opposite been true, the universe as we know it could not exist because ordinary atoms and molecules of hydrogen would not exist as a stable substance.

Nuclear Forces

Since protons all have positive electric charge, they must exert upon each other powerful forces of electrical repulsion that would tend to blow them apart. But since this does not happen, what holds the neutrons and protons (nucleons) in the nucleus together? The answer is that the repulsive electrical force is overpowered by a much stronger force called the nuclear force. This force has associated with it an interaction energy, called "strong interaction," which acts to bind the neutrons and protons together in the nucleus. The strong interaction has the unusual property that it acts only over a distance that is comparable to the sizes of the smaller nuclei, about 10^{-13} cm, but within this range its strength is about one hundred times greater than that of electrical interactions.

The second unusual property of the nuclear force is that a nucleon is attracted only to a few of its neighbors even when there are more nucleons present within the range of the force. Because of this saturation property and the short range of the forces, nucleons are not bound into a large nucleus any more strongly than they are into a small nucleus. In contrast, the repulsive electric forces act over long distances and are additive: the more positive charges, the greater the force. Thus as the number of protons increases, the repulsive electrical force tending to push a proton out of the nucleus eventually becomes as large as, or larger than, the nuclear force tending to hold it in, with the result that very heavy nuclei containing many protons are unstable. This instability is counteracted to some extent by the addition of disproportionate numbers of neutrons to heavy nuclei to act as glue, but eventually the repulsive electrical forces become decisive.

THE VALLEY OF STABILITY

In principle, any combination of neutrons and protons would form a nucleus if they were brought together within the range of their mutual strong interactions. As we have noted, however, most such nuclei are unstable. The kinds of instability that occur can best be understood by starting from the naturally occurring stable nuclei.

If the positions of the various possible nuclei are plotted on a contour map showing contours of their masses in terms of the number of protons and the number of neutrons, as shown in Figure 14, the result is a valley, the Valley of Stability. The strongest mutual binding (lowest energy in accordance with $E = mc^2$) and thus the most stable nuclei are found at the bottom of the Valley. Those nuclei on the slopes of the Valley are unstable if they can drop to a lower position by radioactive decay. There are essentially two different kinds of instability of the lowest energy (ground) state of a nucleus: instability by beta decay, and instability by nuclear-particle decay.

Beta decay occurs when an unstable nucleus can change into a more stable form by converting one of its neutrons into a proton, or vice versa. In either case, the excess electric charge is carried off by the emission from the nucleus of a beta particle, with excess energy carried off by the beta particle and a neutrino. Beta particle is another name for either a negative or a positive electron (a positron), while the neutrino is a particle having no mass and no charge. Thus the total number of nucleons in the nucleus remains fixed, but the number of protons either increases or decreases by one to form a nucleus of a different element.

The second kind of decay process usually involves the emission from an unstable nucleus of a composite particle consisting of two protons and two neutrons, which is known as an alpha particle (and is also the nucleus of a helium atom).

Alpha-particle decay rates can be very slow, with lifetimes up to billions of years. This is why elements heavier than lead (82 protons = element 82) are still found in nature; all such elements are unstable because of the mutual electrical repulsion of the protons in their nuclei, but some are so long-lived that they have not yet had time to decay since they were originally formed.

Artificially Created Nuclei

Although the naturally occurring nuclear species are confined to the lower regions of the Valley of Stability, it is possible by bombarding nuclei with charged particles such as protons, deuterons (the nuclei of heavy hydrogen), and alpha particles of high enough energy to overcome the electrical repulsion forces to create new nuclei that lie beyond the end and higher on the slopes of the Valley. The availability of intense sources of neutrons (nuclear reactors) has also been an important source of such exploration. Since the neutron carries no electric charge, it is not subject to the repulsive electrical forces and can thus be introduced into nuclei at low energy. In this way, many unstable isotopes have been produced, including those

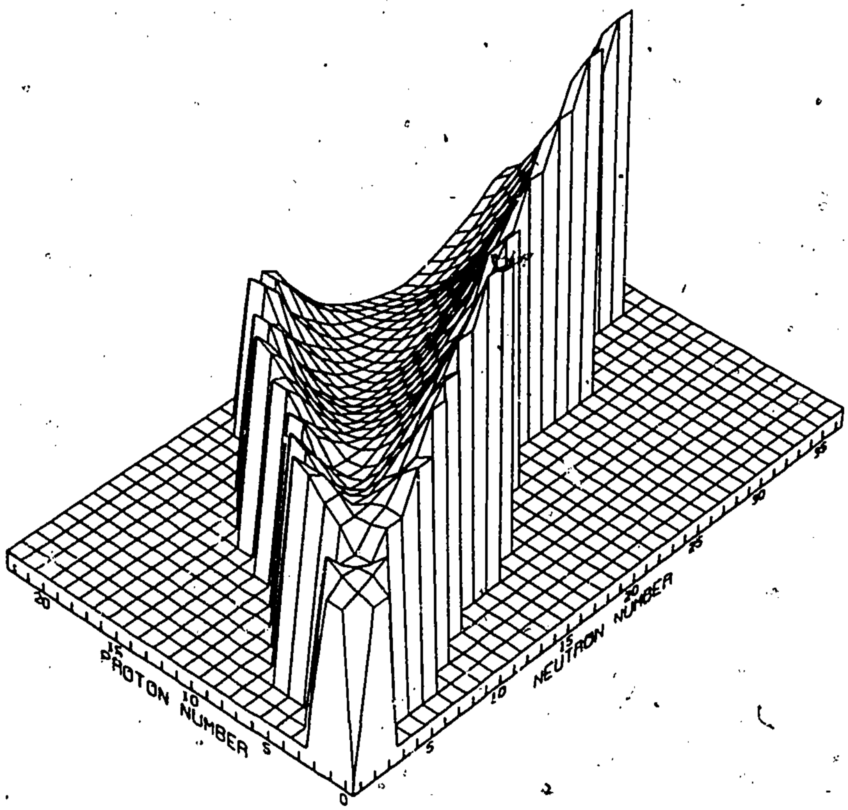


FIGURE 14 Valley of Stability. The bottom of the Valley of Stability is occupied by the stable nuclei. This figure shows the shape of the valley through element number 22. Nuclei on the slopes of the valley are radioactive, decaying to stable nuclei through any of the several decay processes. The highest slopes of the valley consist of the 6,000 or so nuclei predicted to exist but not yet discovered. These would be the least stable nuclei and would thus generally undergo the most rapid decays. (Computer construction courtesy of Jef Poskanzer. From "Exotic Light Nuclei," by Joseph Cerny and Arthur M. Poskanzer, *Scientific American*, June 1978.)

that decay into the transuranium elements neptunium (element 93) and plutonium (element 94). The recent development of powerful heavy-ion accelerators has made it possible to produce short-lived nuclei with considerably greater numbers of protons. The nucleus with the largest proton number observed to date is element 106.

Accelerators are now available that yield high enough energies and a wide enough variety of particle beams to explore the slopes of the Valley of Stability over wide ranges. For example, the normal isotope of sodium (11 protons) has a nucleus with 12 neutrons, but isotopes of sodium having as many as 24 neutrons have been produced.

The exploration of the rugged terrain beyond the Valley of Stability is also of great importance. There may be other "gorges" of stability far removed from the Valley we now know about. It may be possible to form super-heavy nuclei, having masses far beyond those observed. There may even be nuclei with peculiar elongated shapes, with densities twice that of normal nuclear matter, or with other unusual properties that would represent entirely new forms of nuclear matter. Exploration of this unfamiliar nuclear terrain has now begun in experiments at existing heavy-ion accelerators, and the future development of even more powerful machines can be expected to extend this promising field of study.

NUCLEAR SIZES, SHAPES, AND DENSITY DISTRIBUTIONS

Certain gross features of nuclei can be described in terms of the intuitive concepts of size and shape. These properties may be characterized by either the spatial distribution of electric charges and electric currents (charges in motion) in nuclei, or by the distribution of the nuclear matter itself. For a nucleus having rotational angular momentum, or spin, the electric currents produce a static magnetic dipole moment, like a simple bar magnet with a north and south pole. Magnetic moments can be precisely measured by several methods, using techniques that also have had important applications in many areas of material and biological science.

The scattering of high-energy electrons from nuclei is a powerful means for studying the distribution of electric charge in nuclei, since electrons interact only through the electromagnetic field. Further, since the electromagnetic interactions are relatively weak compared to the binding forces of nuclear particles, electron-scattering experiments can measure the structure of the undisturbed nucleus.

To measure the distribution of nuclear matter (rather than electric charge) requires the use of nuclear particles that interact strongly with both neutrons and protons. Since this strong interaction usually causes the probing particles to disrupt the nucleus, this method yields only a limited amount of information about the undisturbed nuclear state. The earliest information about the sizes of heavy nuclei was obtained from measurements on alpha-particle decay, and it has also been possible to obtain information on nuclear sizes from the scattering of neutrons and protons.

THEORIES OF NUCLEAR STRUCTURE

Nuclei present a wide range of empirical properties against which theoretical notions of their structure can be tested. Each nuclear state is characterized by its energy, by its rotational angular momentum (or spin), and by a number of other distinctive properties (quantum numbers). In addition, there is a wealth of information on decay and scattering processes, on electromagnetic transitions, and on nuclear reactions between different nuclear states.

The effort to extract order from this abundance of information proceeds in two related steps. First, qualitative patterns or systematic trends in the data are sought. Second, the particular empirical properties (such as spin, etc.) that underlie the systematic features of the data are identified and built into a simplified description or model of the nucleus. A theoretical model of the nucleus can be thought of as a kind of caricature wherein a few selected aspects of nuclear structure are given particular emphasis.

The formulation of models of nuclear structure in terms of individual nucleons is not expected to be as straightforward as is the description of atoms in terms of electrons, because there is no massive center of force to play the role of the nucleus in an atom. Nevertheless, a model of the nucleus that is a direct lineal descendant of the atomic shell model plays a major role in the theory of nuclei.

Nuclear Shell Model

The basic assumption of the many-particle nuclear shell model is that most of the nucleons combine to form a relatively inert inner core, a nucleus within the nucleus. The significant properties are then those of the remaining "valence" nucleons—those that occupy the lowest-energy orbits or shells outside the core and are free to interact and scatter among the orbits accessible to them. This simple picture has long been known to account for many qualitative features of nuclei, but a thorough test of the quantitative power of the model has taken much longer to achieve because of the enormously complex calculations that many-body problems require. As a result of advances in the technology of computers and in shell-model computations over the past decade, the interacting many-particle shell model has now been applied in detail to nuclei containing up to 40 nucleons and also to heavier nuclei with no more than four or five valence particles. In addition to fitting some of the energy levels and a variety of other properties and transitions with remarkable fidelity, the shell-model calculations have succeeded in reproducing some of the systematic trends of the collective models discussed below.

Collective Models

When many nucleons accumulate outside the central core, the nucleus may assume a permanently deformed shape roughly like that of a football. This deformed nucleus can rotate in space in different ways that correspond to different levels of excitation. In addition, the shape of the deformation can also change periodically in time, leading to different vibrational states of excitation. Rotation and vibration are the typical collective properties of nuclei in the sense that they involve the cooperative movement of many nucleons. Collective models can account for some of the characteristics of a large class of nuclei.

Further tasks of nuclear theory are to relate the collective models back to the shell model and also to relate the basic elements of the shell model (the binding energy and density of the core and the orbits and interactions of the valence nucleons) to the forces that act between free nucleons. Although progress has been made in both of these undertakings, much work remains. In particular, the many-body foundations of the shell model are poorly understood, and it has not yet been possible to use the qualitative insights gained from the success of collective models to reduce the complexity of shell-model calculations.

NUCLEAR REACTIONS

So far we have concentrated on the individual properties of nuclear states. A second major area of study concerns the various processes that occur when nuclei collide with each other or are struck by other kinds of particles that are capable of disintegrating them. When such nuclear reactions are induced by energetic beams of light ions (those consisting of up to four nucleons), two general kinds of processes can occur: direct reactions, in which projectile and target nucleus interact peripherally and the reaction products emerge almost instantaneously; and compound-nucleus reactions, in which projectile and target nucleus coalesce to form a larger nucleus that does not release the final reaction products until a very much longer time has passed. The study of light-ion induced reactions has provided a great deal of information about the relations between similar states in nuclei that lie close to each other in the periodic table. However, such studies reveal little of the response of nuclei to larger disruptions or excitations.

In order to explore these more complex and radical excitations, nuclei must be bombarded with heavier ions so that many nucleons can participate in the collisions. The following examples indicate some of the important new processes that have been observed in such work:

- Nuclear states are formed in which the nuclear rotation corresponds to very high values of angular momentum. Such states are significantly different from those previously studied in detail and should offer new insights into nuclear structure.
- In other collisions, the nucleons settle down into long-lived configurations that are simple but very unlike those found in the normal ground states. A striking example of such configurations called fission isomers has already been studied in heavy-ion experiments.
- New species of unstable nuclei are formed, particularly proton-rich isotopes, which extend the range of proton numbers over which nuclear properties can be explored.
- Heavy-ion studies have revealed a qualitatively new kind of reaction process that is intermediate between the direct and compound-nucleus reactions. Although these deep inelastic processes occur on the very brief time scale of direct reactions, they are not peripheral collisions because there is a large conversion of kinetic energy into internal heating or excitation of the colliding nuclei. These new processes can hardly fail to shed important new light on the nature of nuclei.

Under suitable conditions, nuclei can be excited by the beam particles into collective vibrational motion. The best-known example is the giant dipole excitation in which the neutrons in a nucleus oscillate as a whole against the protons. Recently, several more complex giant vibrational excitations have been discovered in studies of the scattering of electrons, protons, and alpha particles from target nuclei. It is clear that such vibrational excitations must play a significant role in the deep-inelastic processes described above, but the precise nature of that role remains to be elicited.

NUCLEAR FORCES AND NUCLEAR STRUCTURE

Viewing the nucleus simply as a collection of mutually interacting protons and neutrons confers fundamental importance on the strong or nuclear force that acts to bind these nucleons together. Scattering experiments over a wide range of energies have provided a detailed empirical description of the proton–proton interaction. The neutron–proton scattering data is much poorer and so, correspondingly, is our understanding of this interaction. However, indirect information obtained from the structure of light nuclei has led to the tentative conclusion that the nuclear forces obey the symmetry principle of charge independence, which means that they do not distinguish between protons and neutrons and thus affect both kinds of nucleons equally.

Pions

The electromagnetic interaction between two electrically charged particles is described in terms of a force-carrying particle (or field quantum), the photon, which is said to be exchanged between the two interacting particles. In a similar way, the strong interaction between two nucleons is believed to be carried or mediated by the exchange of a field quantum of a different kind. In contrast to the massless photon, the field quantum of the strong interactions must have an appreciable mass in order to account for the short-range nature of this nuclear force. This quantum has turned out to be the particle known as the pi-meson or pion (however, see p. 214).

The existence of pions as the intermediaries in nucleon interactions introduces an additional factor into the description of nuclear structure. One can no longer hope to obtain a complete description in terms of the motions of protons and neutrons alone but must also include the contribution of pions to the structure. For example, since some pions are electrically charged, their motions create electric currents within the nucleus which affect its electromagnetic structure. For many years there has been unambiguous evidence that some such mechanism was needed to account for the magnetic properties (moments) of certain light nuclei. Only recently, however, have precise experimental measurements been found in close agreement with the predicted effects of pion currents.

THE NUCLEUS AS A MICROSCOPIC LABORATORY

The wide variety of states in which nuclei can exist provides unique opportunities for exploring the fundamental interactions. Of the four basic forces in nature (strong, electromagnetic, weak, gravitational), all but gravity play a measurable role in nuclear structure.

Strong Interactions

The first evidence of strong interactions arose from studies of nuclei, and fundamental questions about these interactions continue to be investigated by the methods of nuclear physics. Precise quantitative measurements of nuclear forces provide an essential test of any theory of the strong interactions. Furthermore, the important internal symmetry principle of the charge independence of nuclear forces mentioned earlier has direct consequences for nuclear structure, and through study of these consequences the principle itself is being tested to increasing accuracy.

Weak Interactions

The earliest evidence of the weak interactions came from observations of beta decay of nuclei, and the study of nuclear physics has also provided much of the subsequent information about this basic force. It was an experiment with radioactive nuclei that first verified the remarkable fact that the weak interactions govern the only processes in nature that distinguish absolutely between left- and right-handedness (parity violation). Nuclear studies have been central to the efforts to unify the weak and electromagnetic interactions, and nuclear physics experiments are stringently testing the validity of this proposed unification.

FUTURE DIRECTIONS

In the years ahead, nuclear physics research appears likely to develop in both breadth and depth. Because the great variety of nuclei provide a very complex system requiring much further study, work will continue in determining nuclear properties with higher precision. In addition the availability of new techniques will open up many new research opportunities that are likely to provide the major impetus for progress in this field.

Nuclear Spectroscopy

Studies will continue in many areas of nuclear spectroscopy, which deals with the simple modes of excitation of nuclei, their properties and mutual relationships, and the nuclear models that embody their salient features. The questions here are primarily those that have motivated previous work, but that can now be studied at much greater levels of precision and sophistication. How are the various nuclear models related to each other and to the force between free nucleons? Will the underlying assumptions of these models allow their application to states of greater energies and angular momenta or of larger deformation? Are the currently recognized symmetries of nuclei all that are required? These are typical of the many questions that await sustained theoretical study and experimental work with accelerators of high precision and intensity but relatively low energy.

Extreme States of Nuclei

Our understanding of the structure of the lower energy states of stable nuclei is based on such extensive work that drastic changes in the overall picture are not likely. Instead, new phenomena are much more likely to occur in nuclei that are in some extreme condition, and the study of such extreme states is expected to be a central focus of future work. The advent of accelerators of higher energy and greater beam diversity will be an important stimulant in producing and examining such states. Some of the foreseeable systems to be studied are listed below.

- Nuclei on the edge of stability, with a large imbalance between neutrons and protons. These neutron-rich or neutron-deficient nuclei can be produced in several ways and studied on-line with isotope separators. The shapes, moments, and energy levels of some extraordinary isotopes of sodium and argon have already been studied in this way.
- Nuclei with extremely high masses, particularly superheavy nuclei that are stable or quasi-stable.
- Nuclei under extremely rapid rotation. Nuclei possessing angular momenta of 40 or 50 units are forced into unusual rearrangments of internal structure by the very large centrifugal and Coriolis forces.
- Nuclear matter in new states of very high density, as may occur in high-energy collisions between heavy nuclei.
- Nuclear structures formed from unconventional particles, for example, strange particles or antiprotons.

INSTRUMENTATION FOR NUCLEAR RESEARCH

For several decades the tools of nuclear science, the electrostatic generator, cyclotron, and electron accelerator facilities, have provided a wealth of information on nuclear structures and nuclear dynamics. However, some of the most challenging scientific problems require capabilities that are beyond those of current nuclear facilities but could be attacked by a modest upgrading of them. And, although high-energy accelerators (greater than a few billion electron volts) have been constructed for elementary-particle research, they are becoming increasingly valuable tools for the acceleration of heavy ions for nuclear science research. Machines for producing heavier ions of high energy are now under development for this purpose.

Within recent years, intense beams of pi-mesons (pions) have become available for the study of nuclear structure. The strong interaction of pions with nuclei provides an effective means for measuring the forces acting at short distances from nuclear particles. In addition, the transfer of a large amount of energy from pion to nucleus can result in the creation of unusual nuclei that lie far from the Valley of Stability. This work is still in its early stages, and in the next few years it should produce important insights into nuclear structure.

The very strength of the pion–nucleon interaction becomes a liability when the intention is to study the nucleus in its undisturbed state. For such studies, the much weaker electromagnetic interaction is the preferred probe. In particular, beams of electrons and photons (gamma rays) with energies up to a few billion electron volts serve as a very sensitive means for resolving the fine details of nuclear structure. Some studies of this kind have been carried out under special arrangements at existing large electron accelerators, but there is an important need for an electron accelerator more closely matched to the purpose and dedicated to experimental studies of this kind.

Studies of nuclear structure would also benefit greatly from experiments in which the nucleus is probed by beams of neutrons of greater intensity than now available, by neutrinos, and by high-resolution laser spectroscopy. Each of these areas is now being investigated for its scientific potential and for the kinds of instrumentation that would be required for successful experimentation. In the case of neutrino beams, the required intensity would be very large because of the weak neutrino–nucleus interaction. At present, there are no sources of neutrinos of the required intensity and correct energy range.

PARTICLE PHYSICS

The structural units of matter in its usual forms have been described thus far as electrons, protons, and neutrons. From these, nuclei, atoms, molecules, and ordinary macroscopic matter are constructed. As late as the middle 1940's these units were thought to be indivisible and were therefore called elementary particles.[6] In addition to those known particles, there was at that time evidence from cosmic rays and radioactive decay of the existence of other forms of matter—the pion, the neutrino (one of the products of beta decay), the muon (a heavy electron), the positron (a positively charged electron), and, of course, the photon.

Throughout the subsequent decades, many new kinds of particles were discovered, most of them related in some way to the two nuclear particles, the proton and neutron. These discoveries came slowly at first, then at an ever-increasing rate. By 1960 the number of known particles had grown from 4 to about 30, and most physicists had begun to have serious doubts that so many different kinds of particles could all be elementary in any meaningful sense.

Although the list of particles has now increased to more than 200, evidence has been accumulating during the past decade in support of the idea that *none* of these nuclear particles, not even the proton itself, is truly elementary. Instead, they all appear to be composite structures built up from simple combinations of only a few kinds of basic constituents called quarks. Quarks have not been directly observed in experiments and are thus still hypothetical particles, but the circumstantial evidence for their existence has now become so various and persuasive that there are not many doubters left.

FORCES AND INTERACTIONS

The forces holding the building blocks of matter together have been described as gravitational, electromagnetic, and strong interactions. Just as the photon is the quantum or packet of electromagnetic energy, the pion was originally thought to be the quantum of the interaction energy associated with the strong force of the nucleus. This is another concept that has been modified by recent developments but, nevertheless, the pions are a form of matter participating importantly in strong interaction physics.

The electron, positron, and neutrino are emitted by nuclei in the process of beta decay, or radioactive decay. This decay is ascribed to some kind of interaction with protons and neutrons, just as the emission of light by atoms is ascribed to the electromagnetic interaction. However, this interaction cannot be represented as one or a combination of the three forces mentioned above. It is a fourth form of interaction, called the weak interaction since the magnitude of the interaction energy required to account for the rate of beta decay is found to be very much smaller than that associated with the electromagnetic interaction. This means that beta decay processes occur very slowly.

In terms of the interactions that they are subject to, it is clear that, except for photons, particles fall into two general classes: those that are not subject to the strong interactions and those that are. Those not influenced by the strong interaction are called leptons, and include electrons, neutrinos, and muons, the latter behaving in every way like a heavy electron.

The particles undergoing strong interactions, which include nucleons, protons and neutrons, pions, and many others to be described later, are called hadrons, from the Greek form *hadr-* for "strong."

MATTER AND ANTIMATTER

As far as we now know, every particle has associated with it an antiparticle. For example, the positron is the antiparticle to the electron. The antiparticle has the property that it can annihilate the particle—this means that when the two combine they disappear and can be replaced by photons; and photons in turn can create another particle–antiparticle pair. One can imagine a universe of material made up entirely of antiprotons, antineutrons, and positrons composed to form antiatoms and antimolecules; in other words, antimatter.

The search for simplicity of the laws of physics led naturally to this concept of antimatter and to the simple principle that a universe of antimatter should be indistinguishable from a universe of matter. This symmetry concept is borne out by experience except for a very subtle unexplained property of weak interactions that does distinguish matter from antimatter. However, it should be kept in mind that an unexplained phenomenon appearing to be subtle and unimportant may signal a future revolution in science.

DISCOVERY OF THE HADRON FAMILIES

A series of unexpected discoveries concerning the hadrons began in the mid-1940's with the observation of some particles produced by cosmio rays that behaved in such an unusual way that they could not be identified with any particle known at the time. This behavior was reproduced a few years later in the laboratory by particles produced in collisions between hadrons (protons, pions, and neutrons) at very high energy. Because of their unaccountable behavior they were called strange particles.

Every type of strange particle is unstable and decays in a short time (approximately 10^{-10} seconds or less) into the ordinary, or nonstrange, particles. This lifetime, while short on the ordinary scale of time, is still very long on the nuclear scale and is therefore ascribed to the weak interactions.

There are two types of strange particles, one called the baryon, for which, roughly speaking, one of the decay products is a nucleon and the other called a meson, which, again roughly speaking, never decays into a nucleon. More precisely, a baryon can decay into an odd number of nucleonic particles (usually one) while a meson can on' decay into an even number (usually zero).

Further investigations of the consequences of high-energy collisions between particles have led to an enormous proliferation of new kinds of hadrons, now numbering over 200 different species. This is a virtual zoo requiring a great effort by a large number of experimenters to untangle and identify the particles and the relationships between them. Although they probably occurred in large quantity in the early universe, most such particles are seldom observed now outside the laboratory because they have short lifetimes and the conditions necessary for their creation occur infrequently. All of the particles decay into combinations of lighter particles through a chain of events ending in the stable building blocks of ordinary matter and neutrinos.

QUARKS AS BUILDING BLOCKS

As the zoo of over 200 baryons and mesons of different masses, both strange and nonstrange, revealed itself through high energy experiments, certain regularities in the patterns of masses and other properties of these particles became apparent, indicating that there must be an underlying structure, just as the regularities in the behavior of the chemical elements or the regularities in the patterns of stable nuclei indicated the important underlying structures of those systems.

An early step in arriving at a description of families of hadrons was the recognition that symmetry could account for many features of the observed regularities. This symmetry was a generalization of the simple symmetry between proton and neutron already recognized as a fundamental aspect of nuclear structure. As we have noted, the proton and neutron differ only because one is electrically charged and the other is not. Their strong interaction properties, that is, their characteristics as hadrons, are the same, although there is a slight mass difference presumed to be caused by electromagnetic effects.

In an analogous way, similarities were found among other groupings of particles and these similarities were regarded as an expression of a basic internal symmetry among the particles. In its original form, this symmetry could be accounted for by treating all baryons as composites of just three particles, the quarks. The name quark arose from the quotation "Three Quarks for Muster Mark" from *Finnegan's Wake* because there were just three kinds, now called "up," "down," and "strange" quarks.

At the time this model was proposed, baryons corresponding to all combinations of quarks except that of three strange quarks had been observed. From the model, not only was it predicted that such a baryon should exist, but also it was possible to estimate the mass of this unknown particle. This prediction set off a frantic scramble among the high-energy laboratories to produce and identify the particle, already named the "omega-minus." It was soon found, with a mass within one-tenth of a percent of that predicted. This great triumph lent strong credence to the basic symmetry ideas underlying the quark model but it did not convince physicists of the existence of the quark, especially since no one had seen one.

While baryons were accounted for as combinations of three quarks, the very different characteristics of mesons could be accounted for if each meson was assumed to be made up of a quark–antiquark pair. The existence of an antiquark to go with each quark was a natural and necessary consequence of the assumption of symmetry between matter and antimatter.

Charmed Quarks

The notion that a fourth kind of quark must exist first arose in connection with attempts to relate the weak and electromagnetic interactions. The fourth quark was called a "charmed" quark by the theorists who posed the hypothesis, and charm emerged as a reality as the result of experiments establishing the existence of an entirely new kind of meson, called the "J" or "Ψ," or even the "J/Ψ," in 1975.[7] This meson has a mass greater than that of three nucleons, and a lifetime for decay into other mesons that is much too long to be accounted for as a pair of normal quarks, but that is in accord with the predictions of the charm hypothesis, if it is assumed to be a combination of a charmed quark with a charmed antiquark. Predictions were also made about the masses to be expected for excited states of this system. These predictions were quickly confirmed by experiments.

It is also to be expected that a charmed quark could combine with one of the original three antiquarks to form a charmed meson, and such a particle, the D meson, has also been found (Figure 15). Similarly, charmed baryons are on the menu, and there is recent evidence for their existence.

Six Quarks?

The existence of four quarks is generally accepted by physicists as being well established. However, as a result of another surprise, in 1977, there now appear to be more. At that time a meson having a mass more than nine times that of the nucleon was discovered, requiring the introduction of a fifth quark into the picture.

Finally, the unity of the theory of weak and electromagnetic interactions turns out to call for an even number of quarks. Therefore it is generally assumed that evidence for a sixth quark remains to be discovered.[8]

Where Are the Quarks?

Many experiments have been carried out in an effort to detect a free quark but have not been successful at

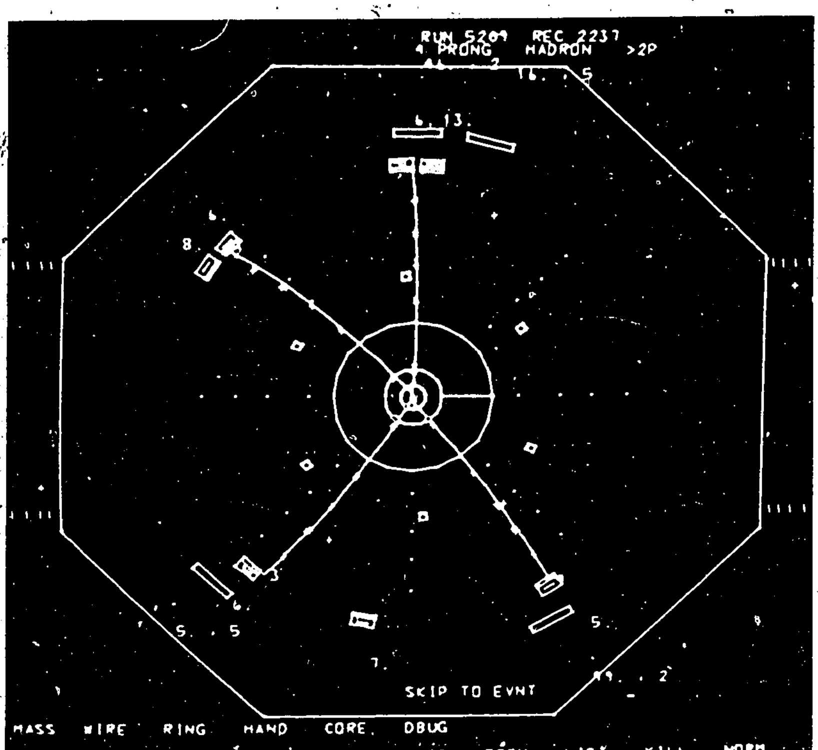


FIGURE 15 Decay of a charmed particle is reconstructed in a computer-generated display. The innermost circle represents the beam pipe in which electrons and their antiparticles, positions, collide to produce other particles. The particle track at 12 o'clock has been identified as that of a negatively charged K meson; and the track at 7 o'clock as that of a positively charged pion. These particles are thought to be the decay products of a D^0 particle, bearing a property called charm. The D^0 particle decays too quickly for it to be detected directly. (Jon Brenneis. From "Fundamental Particles with Charm," *Scientific American*, October 1977.)

the time of this writing.[9] The list of experiments includes not only those carried out at accelerators but also those involving analyses of seawater and of moon rocks, magnetic levitation of small metal balls, cosmic-ray studies, chemical reactions, and so on.

The key to identifying quarks is their remarkable assumed property of having electrical charges of one-third or two-thirds of the "indivisible" unit of electrical charge, that is, the charge of the electron. This eccentricity of charge means that the electrical charge of a quark cannot be neutralized in ordinary matter so that a quark should exist until it meets another quark with which to combine. This also provides a means for identifying the particle.

Since the chemical properties of atoms and molecules are controlled by the electrostatic forces holding them together, it is clear that quark matter would have quite different properties. One can speculate about the possibilities and they are quite interesting, indeed. The only difficulty is that no one has been able to collect even one quark.

Evidently the quarks are bound together to form baryons and mesons in an unshatterable way. Of course, the energy required to smash them into quarks may simply be much higher than is presently available. Or it may be that quarks are a mathematical fiction, yielding the correct properties in a mathematical way, but not real otherwise.

Neither of these explanations is satisfactory and there is a growing general belief among physicists that the explanation lies in an unusual direction. If the quarks are bound together by a force that increases in strength as the distance between them increases, much as would happen if they were bound together by rubber bands, then one could understand that they could never be separated. Theories along these lines are being pursued vigorously and, in the near future, it will probably be possible to propose definitive tests to validate them.

Such forces would represent the basic strong interaction, which binds the quarks together. Just as the electromagnetic fields binding electrically charged particles together can be described in terms of quanta, the photons, so this basic strong interaction field must have its quanta, which are called "gluons" because they glue the quarks together (logic of nomenclature at long last!). Gluons as particles would necessarily have very unusual properties in order to account for the rubber bandlike forces proposed to exist between quarks. Although a free gluon may not be observed until quarks become available to generate them, there are indirect ways of detecting their existence within particle structures and these effects are being investigated.

LEPTONS AND WEAK INTERACTIONS

The leptons provide the only examples of observed, truly elementary particles; they behave in all but some very subtle ways as "point" particles, that is, as structureless objects.

The existence of muons, or heavy electrons, came as a special surprise since there seemed to be nothing in nature requiring it. One famous physicist is quoted as saying, "Who ordered that?" when it was discovered. Another made a habit for years of keeping "μ" written in a corner of his blackboard to remind him of the mysterious role of this heavy electron in physics.

Another question that has often been raised is: If there exists one heavy electron, why not more? The question recently has been answered: A lepton, called the "tau," having a mass 3,500 times that of the electron, or twice that of the nucleon, has recently been discovered.

Some inkling of the role in nature of the muon results from the knowledge that the neutrino associated with the muon is a different particle from the one associated with the electron. Since leptons undergo interaction in pairs, this pairing off of an electron with one kind of neutrino and the muon with another suggests that these pairs of particles may have some fundamental significance in the structure of the physical universe.

The statement that leptons undergo interaction in pairs applies both to their electromagnetic and to their weak interactions. In the case of the former this may be seen by noting that an electron–positron pair is annihilated, or created, in interaction with a photon. For the weak interactions the appropriate neutrino is always associated with the emitted electron or muon in beta decay.

Since hadrons are subject to the weak interaction, so must be the quarks that compose them. The suggestion that there must be a fourth quark, the charmed one, arose from the desire to give the weak interactions of quarks a structure analogous to that of the weak interactions of leptons, thereby requiring that quarks also occur in pairs.

All of this leads to a rather simple picture of nature: Matter is made up of two parallel families of elementary particles: the leptons and the quarks. And there is a one-to-one relationship between them (Figure 16). The quarks, being subject to strong interactions, are buried in hadrons, but the leptons are free to roam. When the tau lepton was discovered, it seemed inevitable that another (fifth) quark would be found, and it was. The assumption that there must be a neutrino to go with the tau lepton then leads again to the need for a sixth quark, as already mentioned.

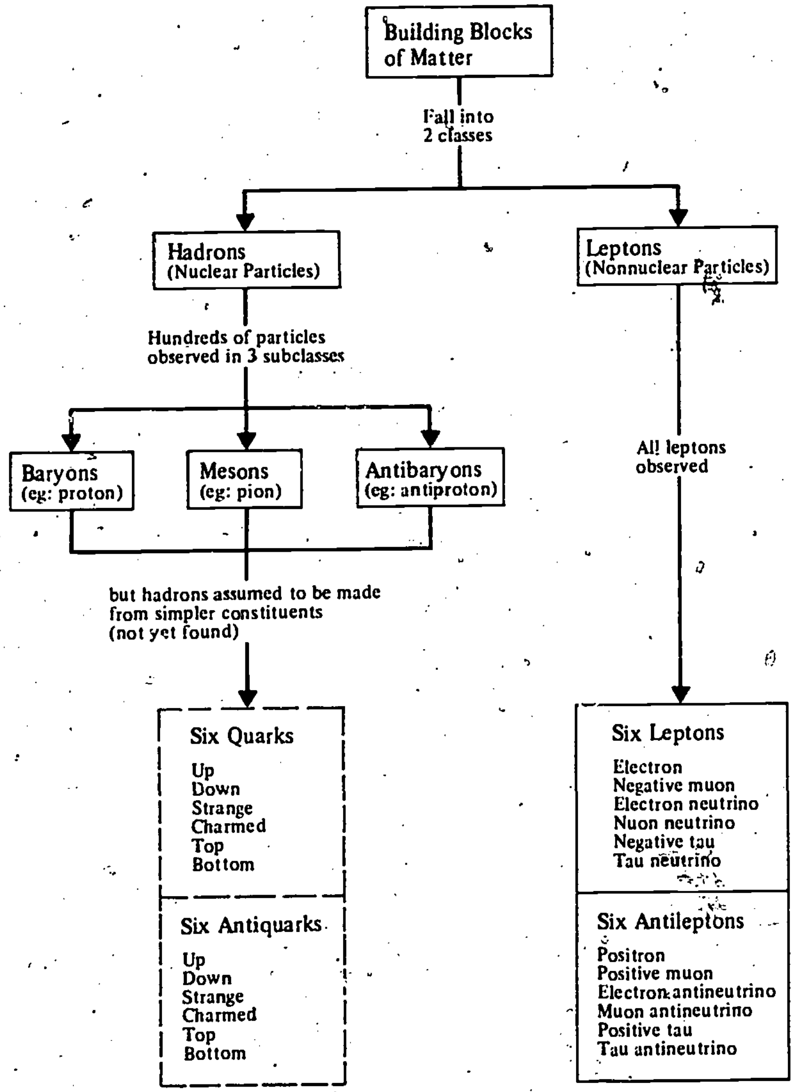


FIGURE 16 Basic building blocks of all matter. The evidence for quarks remains circumstantial.

Unification of Weak and Electromagnetic Interactions

It has been recognized for some years that there are certain similarities between the weak and the electromagnetic processes, but attempts to formulate unified theories and experiments designed to demonstrate their validity ran into serious difficulties until recently. Now it can be said that a very promising unified theory, making use of the concept of pairing of particles as above, has been developed.[10]

One of the earliest questions was why, if these phenomena are so similar, is the weak interaction so very much weaker than the electromagnetic interaction? From the beginning the answer to this question was assumed to be related to the nature of the quantum of the weak field. This would be the analog of the photon except that it would need to be electrically charged if it were to couple to an electron and neutrino the way a photon couples to an electron and positron.

If this charged quantum called the W^+ or W^- has large mass, the effective beta-decay coupling would be correspondingly weak or small. However, there were serious difficulties until a recent reformulation of the theory that included an electrically neutral W. This implied the existence of weak decay processes which did not involve charge, such as the emission of neutrino–antineutrino pairs. These are known as "neutral weak current effects." Such effects can be, and have recently been, observed in high-energy neutrino experiments.

Thus a promising theory has now emerged which, among other things, predicts the mass of the W to be about 75 times that of the proton. There is at present no accelerator capable of producing a particle of so large a mass, but efforts in this direction will be one of the principal objectives of proposed intersecting-beam machines that will be able to attain the required level of energy.

CREATION AND ANNIHILATION OF PARTICLES

There are two ways in which the unstable forms of matter are produced, by decay of more massive unstable forms or by creation processes at high energy. Several decay processes leading to stable particles are familiar to us, for example the decay of an atom in an excited quantum energy state to a lower quantum energy state with the emission of a photon. This photon, having an energy equal to the difference in energy between the atomic states, may be said to be created in the process. By using the equivalence of mass and energy, one may also say that the total mass of matter was unchanged, the mass of the atom in the excited state being greater than the mass of the atom in the lower state by just the right amount to account for the mass of the photon.

In a similar way, electrons and antineutrinos are created in the beta decay of a nucleus, when it is energetically possible. The positron, which is a stable particle, is also produced, along with a neutrino, in beta decay when the unstable nucleus has a larger positive charge than the product nucleus.

Positrons don't persist in matter because they are annihilated by contact with ordinary matter containing electrons, their antiparticles. This process, and its reverse, electron–positron pair creation, is an example of the processes involved in the creation and annihilation of particles.

We know that a positron and electron can swallow one another, leaving behind nothing but the energy represented by the two masses. This energy appears as electromagnetic radiation or in the form of other particles. One can say that "matter" (the electron and positron) has been annihilated. Similarly, a photon of energy greater than twice the electron mass can, under the right conditions, be converted into an electron–positron pair; thus matter is "created."

Of course one can take the view that photons, having mass, are also matter and that there has been no change in the amount, or mass, of matter; there has simply been a change of form. Thus we arrive at the point of view that matter is conserved, but it may appear in a great variety of forms as long as they all have the same total energy.

Exotic particles may be created by such processes or by collisions between energetic particles. The mass of the pion is about 270 electron masses and it can therefore be produced in collision between protons if they have enough energy to produce this mass by a process like

$$p + p \longrightarrow p + n + \pi^+$$

which means that two protons collide to produce a proton, neutron, and positively charged pion. Protons of sufficient energy occur in the cosmic radiation falling on the earth from outer space, and it is for that reason that pions were first observed in cosmic ray experiments. The pion decays into a muon and neutrino in about 10^{-8} seconds. Muons, which have a lifetime ~100 times longer, were discovered much earlier than pions, also in cosmic rays. The muon decays into an electron and two neutrinos.

PRODUCTION OF PARTICLES IN THE LABORATORY

In order to produce pions in the laboratory by the process just described, it is necessary to have machines capable of producing about 400 million volts to yield a proton beam of adequate energy. Since the pion is the lightest of the mesons, machines of much higher energy are required to produce the other mesons and baryons. With sufficient energy (a machine of over 5 billion volts) it is possible to produce antiprotons and antineutrons paired off with protons and neutrons, respectively. At even higher energies the antiparticles of the strange baryons and other members of the zoo are produced.

One may ask how it is possible to identify and work with particles whose existence is so transitory. For those of very short lifetime, observations and measurements are made on the decay products. Particles of lifetimes 10^{-10} seconds or greater are observable although that time is only long enough for a light signal to travel three centimeters. However, at high speed the particles actually travel much farther than this because of the relativistic effect of time dilatation predicted by Einstein: A moving clock appears slower to a stationary observer than does a stationary clock. This effect becomes large at speeds close to the speed of light. Because these particles are produced in very high-energy collisions, they move with such high speeds. The relativistic effect can be so large that, for example, it is possible to produce intense beams of pions over long distances despite their short lifetime (at rest) of 10^{-8} seconds. Such beams are used to bombard targets to produce other kinds of particles.

RESEARCH INSTRUMENTS FOR PARTICLE PHYSICS

The primary beams of high-energy particles required for particle production experiments and other high-energy studies are produced by large particle accelerators that are direct descendants of the atom smashers developed in the early days of nuclear physics. As we have seen, the more deeply one probes the substructure of nuclei, the greater the energy required—to take advantage of the shorter wavelengths of higher-energy particles so that structural details can be "seen" over short distances and to have enough energy to knock apart the strongly bound part of a system or to create particles of high mass (in accordance with $E = mc^2$).

The primary beams may consist of either electrons or protons. The accelerators may be either circular machines (synchrotrons) or linear machines. Since electrons copiously emit energy in the form of synchrotron radiation when moving at high speed on a curved path, there is a point of diminishing return in trying to accelerate electrons in circular machines.

In the United States the highest-energy (22 billion electron volts or GeV) electron accelerator is located at the Stanford Electron Accelerator Center (SLAC), and there is a lower-energy (about 15 GeV) electron synchrotron at Cornell University. The highest-energy proton synchrotron at the Fermi National Laboratory usually operates at about 400 GeV although it has been pushed to 500 GeV. Additional proton machines are the 30 GeV Alternating Gradient Synchrotron (AGS) at Brookhaven National Laboratory and the 12.5 GeV Zero Gradient Synchrotron (ZGS) at Argonne National Laboratory.

Developments for the future include the construction of an energy doubler at the Fermi Laboratory to increase the proton energy available to 1,000 GeV = 1 TeV (1 trillion electron volts). At the other accelerator laboratories, several colliding beam devices are planned that increase the effective energy even more for reasons to be given later. To increase the available resources needed to carry out these plans, it has been decided to shut down the ZGS at the end of fiscal year 1979.

Electron and proton accelerators using a fixed target are very effective because the primary beams have a very high rate of interaction with a target and therefore provide a large number of events for measurement. They also produce intense secondary beams of particles moving at high speed. However, there is a disadvantage to these machines associated with the high speed of the protons, namely, much of the energy of the primary beam goes into the motion of the secondary particles and not into their mass.

To realize maximum use of a machine's energy, the beam can collide with another high-energy beam moving in the opposite direction instead of making it strike a fixed target. When two similar beams of the same energy collide there is no wasted energy; all of the collision energy can go into the mass of the produced particles. Devices making use of this principle offer the ideal way to explore for particles of extremely high mass.

As a means for following up the discovery of new particles with precision measurements, colliding beams do have limitations. The most serious is the rather small interaction rate resulting from the low "target" density, since the target of one beam is the other beam. This can only be overcome by substantially increasing both beam intensities, but under no circumstances can an event rate comparable to that of fixed target machines be anticipated. The other limitation is that the secondary particles produced are moving too slowly to be useful as beams. If they have short lifetimes, they decay quickly because there is no appreciable time dilatation.

Colliding beams of electrons with positrons, each with 4 GeV of energy, are provided by the facility called SPEAR at SLAC. PEP, also an electron–positron colliding beam machine, with each beam having an energy of 15 GeV, is also under construction at SLAC. A facility comparable to PEP is operating at Hamburg, Germany. The only proton–proton colliding beams presently in existence are provided by the intersecting storage rings (ISR) at the international facility, CERN, in Geneva, Switzerland. The energies of the two proton beams are 31 GeV each. To produce the mass equivalent to the 62 GeV available at the ISR would require a fixed target machine of about 2,000-GeV energy.

Construction has just begun at Brookhaven National Laboratory on an intersecting proton beam machine designed to produce colliding beams of 400 GeV protons. At the same time, an effort is under way at CERN to use the 400 GeV Super Proton Synchrotron (the equivalent of the Fermi Laboratory machine) to circulate protons and antiprotons in opposite directions and bring them into collision. Thus the region of very high masses, such as that of the hypothetical W to be described later, will be explored.

MEASUREMENTS AND INSTRUMENTS

Measurement of the properties of the particles produced in collisions of the beams with targets or with each other is the principal objective of the science of particle physics. Such measurements including numbers, directions, momenta, masses, and energies of particles must be made with high precision. It is also necessary to separate events at incredibly close time intervals. The instruments required for the purpose become increasingly complex as the energy increases. There are at least two reasons for this. One is that the number of particles produced in a collision increases, making it more difficult to sort out the particles of interest. The other is that the higher the energy, the more difficult it is to measure the momentum of a particle by bending its trajectory in a magnetic field. In order to capture the necessary information, the instruments must increase in size as well as in complexity.

Required instruments include particle detectors, huge magnets, very fast response electronic circuits and on-line computers to digest the incoming information. The detectors depend on the ionization that charged particles produce in going through matter, while the detection of nonionizing particles, such as photons, neutral pions, or neutrinos, depends on their production of charged particles, which are then detected. A charged particle passing through a gas or liquid leaves a trail of ionizing particles that can be made visible as a vapor trail in a cloud chamber, a trail of bubbles in a bubble chamber, or as sparks in a spark chamber. As the need to observe a multitude of these trajectories more and more accurately has grown, special electronic devices have been developed, the most recent being the drift chamber, which operates inside of strong magnetic fields to give a very precise measure of points along particle trajectory (to better than 0.1 millimeter), and handles many trajectories at the same time. This information is combined with that obtained by other methods, such as total energy calorimetry, which measures the energy absorbed in a mass large enough to absorb all of the energy of a jet of many particles. With these and other techniques it is possible to collect information at an enormous rate. It then takes powerful, fast computers to analyze these data to the extent that the experimenter knows what has happened in his apparatus. The final step, scientific interpretation of the data, requires the most careful attention of the best scientific minds.

A major experiment at a high-energy accelerator requires a large team of physicists backed up by engineers, technicians, and computer personnel to plan and design an experiment, develop and design the apparatus, design and prepare the data-acquisition system, and analyze and interpret the data. One experiment may involve 30 or 40 physicists, plus a comparable number of support people and require two years of planning and preparation, another two years for running the experiment, and at least a year for analysis and interpretation. This is a far different picture from that of the lone scholar in his laboratory, but the scholarly consequences in terms of the depths of insight are different, too, and such formidable efforts are required to reach to these depths.[11]

CONCLUSION

THE ROLE OF MEASUREMENT

Precise measurement is the hallmark of the remarkable advancement in understanding the physical universe in modern time. Quantitative description of matter leads to mathematical formulations of the laws of nature, which, after many tests of their validity, may tentatively be regarded as established truth. Reproducibility and predictability of data go hand in hand with precision of measurement. Understanding grows as the character of observation shifts increasingly from the qualitative to the more precise and quantitative. What is perceived as a color becomes a precisely measured wavelength of light. A shape is related to carefully measured angles of cleavage planes of a crystal. Such measurements attach numbers such as length, time, and angle to physical phenomena, and these numbers may be remeasured many times to test their reproducibility and validity, thus helping lay the foundation of a science.

SIMPLICITY OF THE LAWS OF NATURE: THE ROLE OF SYMMETRY

Precisely measured numbers alone are not enough, however. A complete description of matter requires that it be related, according to a set of rules formulated in terms of simple mathematical expressions. These rules transcend the boundaries of scale, and it is an axiom of science, consistent with a vast amount of experience, that they are based on principles that apply everywhere and for all time. The simplicity of the mathematical expressions may be manifest in either their structure or the statement of symmetry implied, or both, when the two are related.

The elementary concept of symmetry is that of geometrical form. For example, the geocentric and heliocentric models of the universe take the sphere as the natural geometrical form. The eventual realization of the role of spherical symmetry in nature is expressed in terms of the homogeneity of the space around us—no matter which way we turn, the character of empty space is the same. This leads directly to the pervasive fundamental law of conservation of angular momentum, which plays as important a role in the structure of atoms as it does in the structure of the planetary system.

There are many other important geometrical expressions of symmetry in physical science, including the description of the structure of crystals, the behavior of conduction electrons (those carrying electricity), and the structure of molecules. The whole understanding of the chemical behavior of the elements turned out to depend on a very general internal rather than geometric symmetry property of electrons, related to their interchangeability—the fact that two electrons are indistinguishable. This concept of a symmetry resulting from the interchangeability of particles has been generalized, as in the case of the neutron and proton, with remarkable consequences for our understanding of the fundamental building blocks of matter.

UNITY OF THE PHYSICAL SCIENCES

The fact that the complex behavior of the physical universe is governed by a set of universal principles and mathematical laws implies unity among the various fields of physical science, but unity at a deep theoretical level not readily apparent. Nevertheless, because of a commonality of the underlying principles, there is a unity of method both in theory and experiment. The result is that the subfields interact in a supportive and synergistic fashion. Physical science has an organic quality, with the whole being much more than the sum of the parts.

The unity of method and structure is manifest in the understanding of cooperative quantum behavior of many-particle systems, whether they be electrons in solids, atoms in liquids, nucleons in nuclei, or neutrons in neutron stars. Symmetries and "broken" symmetries play very similar roles in the theories of elementary particles and the theory of macroscopic phenomena such as magnetism. The properties of the elementary particles play an essential role in theories of the origin of the universe, and some recent ideas arising from the unified theories mentioned below suggest a previously unanticipated picture of the way in which matter may have been created in the beginning.

This unity of method is a reflection of an even deeper unity of fundamental principles. The history of physics has been marked by a struggle for ever deeper unifying principles. Examples of successes are manifold: Newton's connection between the earth's gravity and the motion of planets around the sun; the unification of heat and energy by Clausius; the unification of electricity, magnetism, and optics by Maxwell and Faraday; the unification of the corpuscular and wave theories of light by quantum theory; the unification of chemistry and physics by quantum mechanics; and the current, apparently successful unification of electromagnetic and weak interaction theories. Recent work has led to the hope for an early extension of this last concept to include the strong interactions of particles.

Einstein sought the unification of gravitational and electromagnetic theories as his ultimate aim. Through the study of quantum effects of gravitational fields, we may see his hope realized on a much grander scale than he ever imagined by the unification of gravitation, electromagnetism, weak and strong interactions—all of the known forces of nature.

OUTLOOK

The following outlook section on structure of matter is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

ASTROPHYSICS AND COSMOLOGY

Rapid developments in astronomy and space science during the past few decades have led to an entirely new perception of the cosmos. Observations of almost the entire electromagnetic spectrum from radio and infrared through visible and ultraviolet to X-ray and gamma rays have been made possible by new ground-based instruments and space probes. Similarly, important progress in the study of cosmic rays has also been made possible by new instrumentation and the mounting of instruments in space vehicles.

The detection of 3°K radiation has provided convincing support for the Big-Bang concept of the origin of the universe. The evidence for the existence of neutron stars confirms theories of stellar evolution and the strong suggestion of evidence for black holes, tends to confirm one of the more remarkable predictions of Einstein's general theory of relativity.

The prospects are for continued rapid advances in this field. Observation and measurement are just beginning at both the infrared and gamma radiation ends of the spectrum. The planned gamma-ray space observatory will permit astronomers to make measurements with about 10 times the sensitivity now possible. The space telescope scheduled for launch in the space shuttle, the new ground-based large multiple-mirror telescope, and the very large array radio interferometer will permit much deeper penetration into the universe.

These and other new instruments may help resolve some of the important cosmological and astrophysical questions concerning the structure of the universe, for example, whether it is open or closed, and the actual existence of black holes.

Also significant questions concerning stellar processes such as nuclear synthesis of the elements will be addressed by cosmic ray experiments. Neutrino detection methods may help resolve questions about energy sources of the sun raised by recent experimental results.

CONDENSED MATTER

Research on matter in the large, on the macroscopic scale, has an impact on all the ways in which materials are used. There have been an enormous number of advances in our understanding and application of the science of condensed matter. Some of the most recent developments showing promise for substantial progress in the next five years are studies of amorphous materials, surfaces, and exotic new materials, as well as atomic and molecular engineering and study of systems far from equilibrium. The trend is toward dealing with systems of increasing complexity at higher levels of precision.

An important theoretical advance has been our improved understanding of phase transitions, such as transitions from gases to liquids or nonmagnetic to magnetic phases in solids. This fundamental work is applicable to a remarkably wide range of phenomena.

Since instrumentation has become more elaborate and expensive, ways of sharing apparatus among institutions are being explored. Already possibilities for qualitatively new experiments on condensed matter with major centralized facilities are being realized. For example, the new research frontiers provided by synchrotron radiation sources of high-intensity ultraviolet and X-radiation, which are powerful tools for studying the structure of condensed matter, are being approached by several new facilities presently under construction.

Intense neutron sources that can be used to study different aspects of the structure of condensed matter represent another example. They also are a very powerful tool for the study of atomic dynamics of liquids and solids. Great progress in this field has been made in Western Europe, which for some time has had the most intense steady source of neutrons, at the Institut Laue-Langevin in France. In addition, a new high-intensity pulsed source is under construction at the Rutherford Laboratory in England. Somewhat more modest steady sources have been available in this country for many years, and a considerably more modest pulsed source is under construction.

Since condensed matter science serves as a technology base both for other fields of science and for industrial development, it is essential to maintain our strong capability in the field.

MOLECULAR AND ATOMIC STRUCTURE

Progress has been rapid in the study of the structure of molecules, especially large organic molecules such as those occurring in biological systems. Developments in instruments and spectroscopic techniques paralleling those used in investigating condensed matter have been largely responsible for this. Exploitation of nuclear magnetic and electron paramagnetic resonance methods has been especially fruitful, but, again, increased sophistication of these techniques has raised the cost of instrumentation so that new institutional arrangements to replace the one-scientist–one-instrument tradition must be explored.

Synchrotron radiation sources and intense neutron sources also hold great promise for research in this subfield. The use of neutrons is now at its most primitive stage, beginning with small-angle scattering to obtain gross features of large molecules. An order-of-magnitude increase in the intensity of neutron beams is needed to realize the potential of this technique.

Because of the coherence, intensity, and spectral purity of the light they produce, lasers have become a powerful tool for high-precision molecular spectroscopy, making it possible to observe molecular fragments in, for example, chemical reactions occurring in very short time intervals of 10^{-12} seconds—a picosecond.

The continued deepening of our understanding at the basic level of processes taking place in and between molecules, especially large organic molecules, may make it possible to mimic some biological processes.

A number of elegant developments in atomic physics make possible very precise tests of some of the fundamental laws of nature. Again, lasers have greatly increased the precision of atomic spectroscopy. They have also made it possible to isolate and study a single atom at a time, and to investigate the structure of giant atoms in highly excited energy states, which occur naturally only in outer space. Laser techniques also may enable the separation of isotopes of an atom for practical applications.

Very unusual atomic forms, exotic atoms, made up of unstable particles that can be produced by high-energy machines are being studied. These, along with other forms of advanced atomic spectroscopy, are making it possible to determine the fundamental constants of nature with ever increasing precision.

NUCLEAR STRUCTURE

Here, too, the physicist has moved into an era of great understanding. The availability of high-energy particles enables the study of already known nuclei and the production of heretofore unfamilar ones. Pion, electron, and photon beams are being used to probe the nucleus. Nuclei, and therefore atoms, of much greater mass than those occurring naturally may be produced. Production of such nuclei and even of grossly deformed nuclei with beams of fast heavy ions is just beginning. It may be expected that the discovery of new states of nuclear matter will shed further light on fundamental questions concerning combinations of large numbers of particles.

The problems of nuclear theory have much in common with those of condensed matter, especially those associated with quantum fluids, and there has been a successful sharing between these fields.

PARTICLE PHYSICS

Of the many particles known, one group continues to be regarded as elementary, i.e., indivisible. They are the leptons: electron, muon, the tau particle, and the neutrinos associated with each. The "zoo" of more than 200 other particles called hadrons differ explicitly from the leptons in that the forces between them are very strong, like the forces holding protons and neutrons together in nuclei. Hadrons are not believed to be elementary; indeed, very recent developments have convinced most physicists that they are made up of quarks, although they have not been seen as separate entities. The recent developments in particle physics have resulted from a beautiful interplay between theory and experiment—theory governed to a considerable extent by the idea that there is a simplicity and symmetry in nature waiting to be revealed. For example, there appears to be a match-up between the quarks and the leptons; thus, the discovery of the tau particle, a form of the electron 3,500 times heavier, is associated with a new quark suggested by the discovery of the upsilon.

The quest for simplicity and symmetry also stimulates the quest for unification of apparently different kinds of fields and forces, such as the gravitational and electromagnetic fields. Although success has not as yet been achieved, there has been a recent important advance in tying together weak interactions with electromagnetism; weak interactions being responsible for beta radioactivity of nuclei and electromagnetic interaction for the decay of excited states of atoms by emission of light.

It is expected that there will be rapid progress during the next few years in unraveling the relationships between particles, and between forces. We seem to be on the threshold of tying together the weak, electromagnetic, and strong interactions. Theoretical progress in avoiding some of the poor approximations used in solving these problems is expected. Whether new particles yet to be discovered will modify current ideas about quarks cannot be predicted. There will be a major effort to search for the *W* particle, the quantum of weak interaction, at the international accelerator facility in Geneva. The *W* is believed to have about 75 times the mass of a proton, and no existing machine is capable of providing the energy to confirm its existence. However, a method capable of doing this, a beam of protons colliding with a beam of antiprotons, is being developed. The discovery of this particle would be a great step forward in the unification of the fundamental forces of matter.

INFLUENCES AND TRENDS IN PHYSICAL SCIENCE

The present health of U.S. physical science depends on three major factors that have played a decisive role in establishing the present remarkable strengths of the field. The first is the quality of the personnel. We have benefited from a rich diversity of scientific manpower, selected on a worldwide basis, and from very special educational opportunities.

The second factor is the nature of available instruments and facilities. These depend on the industrial and technological environment and the technical ingenuity of the scientists, as well as on financial resources.

Finally, there is the difficult to define but important element of style. This includes balancing risk and certainty in the selection of a task, choosing between simple and complex methods and techniques of accomplishing it, and the judicious selection and use of needed resources. Style also relates to the way in which the interaction between experimental and theoretical aspects of the science takes place. It is influenced by the objectives of the work, whether it is oriented toward basic knowledge or applications. It reflects the degree of commitment, motivation, and faith in a successful outcome on the part of investigators.

The traditional setting for fundamental research is the university, and the burden for maintaining the thrust of the research as well as the quality of the personnel falls on universities. A current trend toward reducing this traditional role has significant implications for the future. For one thing, inflation, funding that has leveled off, and fewer academic jobs have reduced the opportunities for academic work and careers. This situation is affecting what has been for three or four decades the very high quality of our scientific manpower, and emphasizes the need for new institutional arrangements.

One such arrangement that is having an important influence on some academic institutions is the addition of a number of "permanent," nonacademic, postdoctoral research positions, thereby changing the balance between the faculty, research associates, and students. This trend has been given strong impetus, not only by the changing job market, but also by the growth in research activities centered at large and complex off-campus facilities, since the team effort associated with these activities usually must make use of some experimental full-time scientists.

The growing need for more complex, larger, and more expensive instruments has been manifest for many years in fields such as astronomy and high-energy physics. However, in many other fields as well, the advanced status of research requires more sophisticated instruments in the laboratory and also an increasing need for large centralized facilities.

Except for astronomy, where the need for centralized facilities has been generally recognized, scientists have usually resisted the transition from the small one-man laboratory to the large centralized facility. The idea of the independent scientist and a few of his students working together with apparatus that each can comprehend completely and control is rooted in tradition and continues to be attractive. However, as physical science has become more complex, it has become increasingly evident that the complexity of the apparatus and extent of the facilities must increase if certain research opportunities are to be fully exploited. This leads to a major shift in style—to team research, to the commuting professor, to the student resident at a facility hundreds or even thousands of miles from his university. Such adjustments were made in the field of particle physics and in space sciences some years ago; they are currently taking place in nuclear physics and are about to occur in some special areas of condensed matter science.

Of course, team effort is not new to engineering development. It has been the necessary style in applications of technology on a large scale. However, the need to speed up the translation of the results of recent scientific research into high technology, as in the development of nuclear fusion energy, is introducing new relationships, leading to increasing importance of interdisciplinary work between the basic sciences and engineering.

The ubiquitous computer has also profoundly affected the style of physical research as has almost every other information-related human activity. High-powered computers, which make possible the solution of scientific problems that cannot be treated by other methods, have become powerful scientific tools.

Recording and analysis of experimental data have taken on new dimensions through the development of small computers. Computers linked with scientific apparatus are now essential whenever large amounts of data are involved. They also make possible continuous monitoring of experiments so that adjustments can be made during the course of the experiments. Furthermore, much of the analysis of a complex experiment can be completed within a short time after the data are taken.

In summary, it can be said about the prospects for the science of structure of matter that while the field is in good health in the United States now, there are trends suggesting a weakening of the fabric in the immediate future. Their cause and effects must be given close attention if we are to overcome them.

REFERENCES

1. Gott, J.R. III. Will the Universe Expand Forever? *Scientific American* 234(3):62–79, 1976.

2. Weinberg, S. *The First Three Minutes: A Modern View of the Origin of the Universe*. New York: Basic Books, 1977.

3. Kaufman, W. *Cosmic Frontiers of General Relatively*. Boston: Little Brown, 1977.

4. Hartline, B.K. In Search of Solar Neutrinos. *Science* 204(4388):42–44, 1979.

5. Schawlow, A.L. (ed.). Lasers and Light. San Francisco: W.H. Freeman, 1969.

6. Drell, S.D. When is a Particle? *Physics Today* 31(6):23–32, 1978.

7. Schwitters, R.F. Fundamental Particles with Charm. *Scientific American* 237(4):56–70, 1977.

8. Glashow, S.L. Quarks with Color and Flavor. *Scientific American* 233(4):38–50, 1975.

9. Nambu, Y. The Confinement of Quarks. *Scientific American* 235(5):48–60, 1976.

10. Weinberg, S. Unified Theories of Elementary-Particle Interaction. *Scientific American* 231(1):50–59, 1974.

11. Kirk, W. Parity Violation in Polarized Electron Scattering. *SLAC Beam Line* (Stanford Linear Accelerator Center). Report No. 8, October 1978. (An example of the nature and difficulty of experiments in high-energy physics.)

BIBLIOGRAPHY

Eisenberger, P., and B.M. Kincaid. EXAFS: New Horizons in Structure Determinations. *Science* 200(4349):1441–1447, 1978.

Estrup, P.J. The Geometry of Surface Layers. *Physics Today* 28(4):33–41, 1975.

Ferris, T. *The Red Limit: The Search for the Edge of the Universe*. New York: Wm. Morrow, 1977.

Field, G., D. Verschinier, and C. Ponnamperuma. *Cosmic Evolution*. New York: Houghton Mifflin, 1978.

Friedman, H. *The Amazing Universe*. Washington, D.C.: National Geographic Society, 1975.

Mermin, N.D., and D.M. Lee. Superfluid Helium 3. *Scientific American* 235(6):56–71, 1976.

Physics in Perspective, Volume I and Volume II, Part A. (NRC Physics Survey Committee). Washington, D.C.: National Academy of Sciences, 1972.

Rhodin, T.N., and D.S.Y. Tong. Structure Analysis of Solid Surfaces. *Physics Today* 29(10):23–32, 1975.

Tauc, J. Amorphous Semiconductors. *Physics Today* 29(10):23–31, 1976.

II TECHNOLOGY

4 Computers and Communications

INTRODUCTION

Over the next five years, we can expect that technical factors will continue to drive computer development. The rapidly falling cost of computational power already has led to amazingly inexpensive and versatile computer circuitry and to significant though less dramatic reductions in the cost of data storage, input, and display equipment. All this is making for a swift development in breadth, and allowing computers and computer-controlled equipment to penetrate even more broadly into the daily reality of factories, offices, and homes. In the next few years, the computer uses we already see about us—reservation systems for airlines and hotels, microcomputers that control cooking times and temperatures in microwave ovens, chips that adjust the flow of fuel to automobile engines—will expand rapidly.

These near-term uses, in which millions of microcomputers may come to be used during the next decade, will also have an impact on many large systems. Telephone automatic dialing systems that allow call-backs and waiting for busy signals are already in use. The technology that makes possible today's cash-dispensing terminals will grow, as will office text-editing and document preparation and distribution systems. Continually improved computer capability and lower cost are the keys to this expansion.

CURRENT TECHNOLOGY AND SYSTEMS

ORIGINS FROM SOLID-STATE PHYSICS

Mechanical aids to computation such as the abacus have been in use since ancient times. And a great many machines for adding and multiplying had already been proposed by the end of the eighteenth century. Many of the most fundamental ideas used in today's computers can be traced back to the nineteenth-century British inventor Babbage, who with Admiralty support attempted to build a computer organized in a manner very similar to today's.

After World War II, a whole generation of vacuum-tube based computers was developed that routinely accomplished large computational tasks. But practical computers of today's high speeds are very much the product of sophisticated electronic technology.

HISTORY OF INTEGRATED CIRCUITRY

Early in the history of electronics and radio, experimenters recognized that certain materials had special electrical properties. The crystal detectors—the "cat's whiskers"—of the first radios, and the selenium rectifiers of the 1930's, prized because they allowed electrical current to pass in only one direction, were soon in wide use.

As the quantum theory of crystalline materials subsequently developed, physicists came to understand that semiconducting elements like selenium and germanium constituted a potentially useful borderline between the conducting materials, such as aluminum and copper, which pass electrons freely, and the insulators, which do not. Research on these semiconducting materials in the mid-1940's revealed that it might be possible to build electrical amplifier circuits by forming junctions between semiconducting zones of different properties and that solid-state amplifiers of this sort might replace the inefficient and bulky vacuum tube. Starting from their observations on the use of semiconductors, Shockley, Bardeen, and Brattain of Bell Laboratories mounted a systematic theoretical and experimental thrust toward this goal. By December 1947 their work, and in particular their ability to manage the science of materials, resulted in their creating the first transistor.

This Nobel Prize-winning breakthrough was at once recognized as a revolution in electronics. Not only were transistors much smaller and more reliable than the vacuum tubes they replaced; but, more important, they did not require a hot cathode as a source of electrons, and hence consumed very little power. This revolutionized the form and function of everything electronic, from the portable radio up. Moreover, the transistor's advantages in size, reliability, and low power-consumption made it practical to put together much more complex electronic assemblages than had previously been attempted. This fact is a technological cornerstone of the computer age.

STEPS TO LARGE-SCALE INTEGRATION

Between 1947 and 1957 germanium was the dominant semiconductor material in transistors. But it was soon realized that silicon, another attractive semiconducting material, had one great advantage: When it is heated in the presence of oxygen, a layer of quartz forms upon its surface. Quartz is one of the best-known insulating materials and also is relatively impervious to most of the environmental impurities that can degrade transistors. The possibility of encapsulating silicon in very thin quartz films simply by heating, and then of etching controlled patterns through the quartz, led in 1960 to silicon planar technology, the basis of all current microelectronic technology. In this technology, small regions with special electrical properties are formed in silicon by diffusing impurities, such as phosphorus, through holes etched in the quartz layer on top of a silicon wafer. Metallic conductors connecting these regions are then deposited in patterned strips atop this same quartz layer. This technique, which can be used to create very complex circuits with individual components of microscopic size (e.g., 0.004 millimeters), is the basis for today's sophisticated large-scale integrated (LSI) technology.

CHIP SIZE AND COSTS

Solid-state technology has surged forward during the past decade, revolutionizing the computer industry. A key factor has been an increasing ability to install complex logical functions on a single chip. (In the manufacture of semiconductor devices, large sausage-shaped single crystals of silicon are sliced into thin circular wafers with the circuitry then developed on their surfaces. After processing these wafers are cut into small chips, about one-fourth inch on a side.) Starting from a one-transistor-per-chip technology in 1960, the number of transistors on a single chip has approximately doubled each year. Chips manufactured today routinely contain tens of thousands of transistors and other components, complete with electrical connections. Today's LSI chips can, for example, store 64,000 binary bits of information (Figure 1-7) and make them available to a computer within a small fraction of a microsecond. A complete microcomputer, with processor, memory, and input–output circuitry, can now be built on a single chip. Moreover, LSI chips are manufactured by a kind of photolithography, in which cost is relatively independent of the pattern being developed. As the capability of chips has risen dramatically, chip cost has remained nearly constant.

More complex functions are attained merely by imprinting a more complex, finer-grained pattern on each chip, producing a larger number of individual electronic elements. Smaller elements tend to increase operating speeds and decrease power requirements, which makes the dominant trend to finer structures especially important. To maintain the momentum of this trend, it is necessary to maintain the accuracy with which parts are shaped and the accuracy with which successive layers of microscopic parts are aligned. It also means increasing the number of layers of metal and glass that can be accurately built up on the surface of a chip. Where processes involving five and six layers dominated in the late 1960's, new chip manufacturing processes typically involve 8–12 layers. This trend may be expected to slow since it is difficult to manufacture many layers without raising chip costs

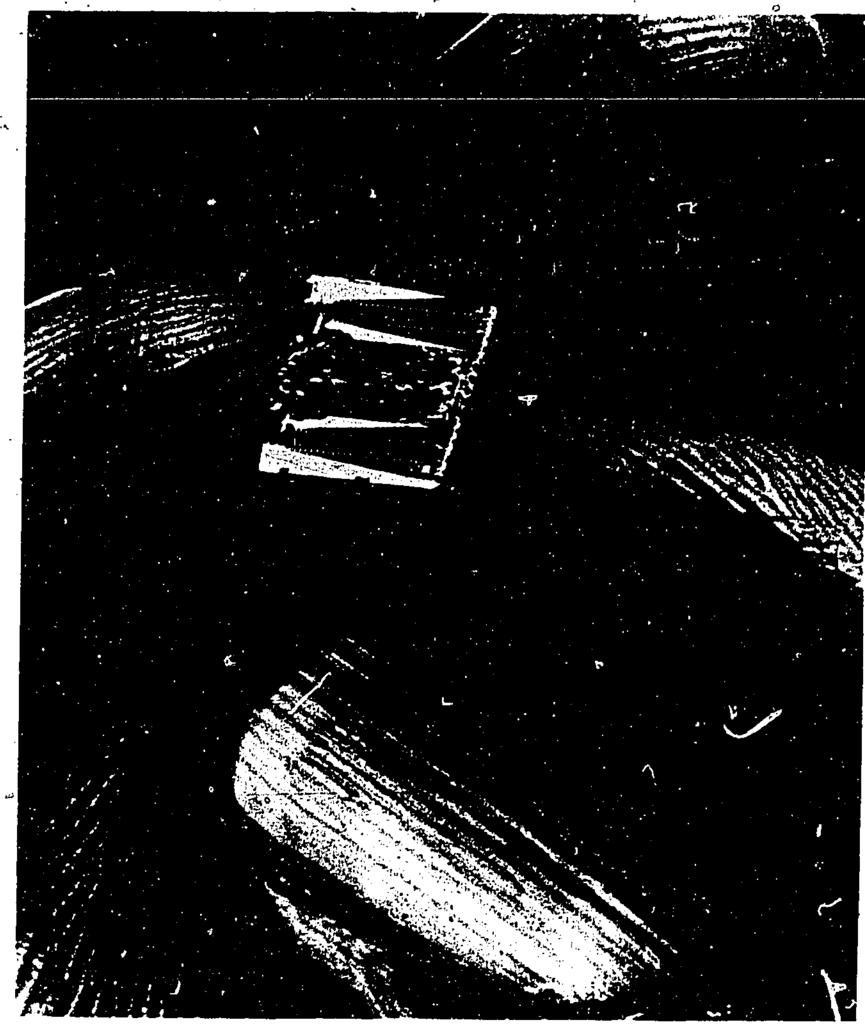


FIGURE 17 64K-bit chip stores information roughly equivalent to 1,000 eight-letter words. (International Business Machines Corp.)

and lowering chip yield (i.e., the number of chips that pass electronic inspection at the end of the manufacturing process).

Components used in today's high-volume complex integrated circuits are as narrow as four micrometers, and new optical image projection systems are expected to reduce this to one–two micrometers over the next several years. Historically, the size of integrated circuit elements, i.e., single transistors, has been cut in half every six years. Progress at about this rate can be expected to continue through the next decade, putting circuits with one-micrometer dimensions into mass production by about 1990. Nonoptical techniques (electron-beam writing and X-ray lithography) may provide even smaller structures, those in the submicrometer range.

Chip costs can be controlled and even reduced by decreasing the number of defects on a processed wafer of silicon through improved process control and manufacturing cleanliness. A significant factor here is the broader application of new processing techniques, such as projection masking, plasma etching, and ion implantation.

An additional favorable cost factor is the use of larger silicon wafers, which allows more chips to be produced per processing step. Wafer sizes have increased from the two-inch standard diameter of 1968 to three–four inches today. Equipment for processing four-inch wafers is now generally available; further expansion to five- or six-inch wafers in the early 1980's is expected. However, upgrading production facilities will require much present equipment to be replaced by larger, more expensive equipment.

The rapid growth within the solid-state industry of smaller devices, greater complexity, and more functions per chip is well illustrated by Figure 18, showing the rising number of bits per solid-state memory chip. This development has taken us from 1,024 bits per chip in 1971 to 64,000 bits per chip in 1978, has kept chip cost approximately constant, and has thus

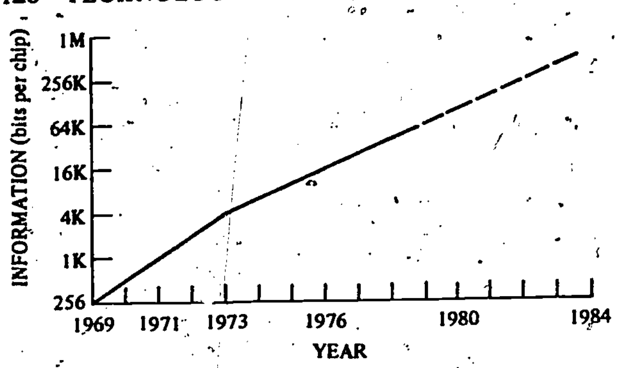


FIGURE 18 Growth of semiconductor technology.

decreased cost per bit by a factor of roughly 16. It seems reasonable to extrapolate this trend, and to anticipate 256,000-bit memory chips in the early 1980's and million-bit memory chips by the mid-1980's.

The development of microcomputers shows a similar trend. Microcomputers have evolved from minimal units involving 2,000 transistors in 1971 to devices that today integrate 30,000 transistors on a chip. Here also progress will continue for the foreseeable future.

Speed has also increased substantially, from the microsecond speeds that typified early integrated circuits to today's fastest commercial chips, which are capable of adding a pair of decimal digits in less than five-billionths of a second. This development should also continue, but at a decelerating pace unless the industry progresses to a new technology, such as the Josephson devices discussed below.

ALTERNATIVE TECHNOLOGIES

Throughout its history the silicon integrated circuit has faced potential competition, but thus far none of these competitors has become important. Nevertheless, it is worth having a look at two of these alternative technologies, one oriented toward storage applications, the other toward extremely high-speed computation.

The first of these is the so-called magnetic bubble technology. This technology exploits the fact that discrete regions of deformed magnetic fields can be formed within thin films of crystalline garnet and then moved under the influence of external magnetic fields along precisely controlled paths. These special magnetic regions or bubbles can be used to store information in much the same way as the magnetized spots in a conventional magnetic tape; except that in a bubble memory it is the stored spots that are moved magnetically, thus the medium that contains them need not move. This allows much faster data motion and makes for much more reliable devices. Unlike silicon circuit memories, bubble memories retain information even when power is turned off; in this respect, they are similar to magnetic discs or tapes.

Another advantage of bubble memories is that their extremely simple physical structure may enable very high storage densities, with several million bits stored per garnet chip. Their disadvantage is that they do not offer the same type of ultrafast information access, within 0.1 microsecond, that silicon memory devices do. Rather, they must cycle through a potentially large number of stored words of information to reach a particular item. This requirement increases typical information retrieval times for magnetic bubble devices to hundreds of microseconds. They thus represent potential all-electronic substitutes for electromechanical bulk storage devices and for the cheapest forms of electronic storage, rather than for fast memories.

The first commercial bubble memories appeared in 1976, some eight years after research had first demonstrated their feasibility. Bubble devices storing 256,000 bits are expected to be available in 1979, and 1-million bit structures are expected a year or two later.

Josephson junction technology is a second alternative to today's silicon chip. It represents a superspeed circuit technology that may begin to be used, at least for special applications, within the next decade. Josephson devices have been shown in the laboratory to switch at speeds of fractions of one-billionth of a second. Moreover, the power dissipated by devices of this kind is very low. The most likely initial use for this technology is in special ultraspeed applications beyond the capacity of silicon devices; for example, in extremely high bandwidth multiplexers for satellite data reception. The rate at which Josephson devices are perfected will be determined by the extent to which resources are committed to this development during the next few years; but it does seem unlikely that any commercial computer will use this technology during the next five years.

It is worth noting that the finite speed of light limits computer performance in a manner that is steadily becoming more significant as circuit speeds increase. Present-day circuits switch in about one-billionth of a second, which is the time required for light to travel about one foot. Josephson circuits could switch in the time it takes light to travel an inch. Since light is the fastest of all physical signals, ultraspeed computers will also have to be physically small.

We have discussed some of the more radical alternatives, but many variations are also possible within the area of semiconductors themselves. These include logic designs based on gallium arsenide and on silicon on sapphire and a great variety of process innovations and new logic designs.

MICROCOMPUTERS AND MINICOMPUTERS

By 1971 semiconductor technology reached a point at which it became practical to build a complete small computer central processing unit—or microcomputer—on just one or two silicon chips. When combined with appropriate memories, themselves consisting of only a few LSI chips, the resulting microprocessors were capable of performing a very wide variety of high-speed control or computational functions. The advent of inexpensive microcomputers has made it practical to employ computer control in hundreds of applications, from sewing machines to television games, which had previously been out of reach for size, cost, or other reasons. Total design costs are kept low because a microprocessor consisting of just a few standardized chips could be used in a great variety of applications, ensuring that a single design could be produced in hundreds of thousands of identical copies.

The first microcomputers had simple overall structures involving relatively few transistors. Evolution since these first designs has been rapid and is expected to continue for many years. Microcomputers have already developed to a level of function overlapping that of yesterday's minicomputers. The computer size and cost spectrum now runs from the largest machines costing several million dollars to the one-chip machines costing less than five dollars.

Directions of Microcomputer Evolution

The evolution of microcomputers is dominated by two major design considerations. The first is that the manufacturer must choose between offering a fixed processor versus a compatible family of chips that can be combined in many ways. The second is the size of the arithmetic quantities that can be manipulated directly on a single chip.

Fixed-design microcomputers are now being delivered by several manufacturers. Generally these are intended for specific control applications, where very large volume use is anticipated, for example, the control of automobile engine spark timing, fuel injection, or exhaust gases. As performance requirements for these high-volume applications expand, single-chip microcomputers will grow in capability and in memory size.

Microcomputers oriented toward complex control applications and toward data processing are evolving into families of compatible chips, allowing very flexible ranges of small systems to be built. One popular approach is to provide chips that can be connected end to end to form arithmetic processing units capable of manipulating quantities of any desired size. In addition to general-purpose microcomputer chips, such families will include memory chips, chips to control peripheral devices such as printers or data-storage discs attached to computers, communication controllers, and special processors, such as high-speed arithmetic units.

Another major trend is toward the development of more sophisticated basic microcomputer designs featuring increased word size. The first microcomputers operated on only four binary bits (essentially one decimal digit) of information at a time. With such short words, many cycles were required to handle multidigit arithmetic data. Consequently, these first microcomputers were relatively slow and also hard to program.

The second generation of microcomputers, still in wide use today, provided units capable of processing 8-bit words (essentially two decimal digits) in parallel. Recently microcomputers have been extended to 16-bit words. These new microcomputers, which are also more advanced architecturally, exploit the new semiconductor technology that allows the manufacture of semiconductor devices 10 to 50 times more complex than the first such devices. Some of these 16-bit microcomputers are more than 10 times as fast as the earlier 8-bit ones.

Micros and Minis

Today microcomputer performance overlaps a significant portion of the range that has been covered by minicomputers; distinctions between mini- and microcomputers based on hardware performance will tend to blur over the next five years. Increasingly, the distinction between micro-, mini-, and larger computers will depend on the peripheral devices to which they are attached and the software systems with which they are offered. Minicomputers will be those small computers that are supported by substantial peripheral devices, including high-speed discs and printers, and extensive software systems; while microcomputers will be those small computers, consisting of a very few chips, which are buried within some larger device and not separately visible. High-performance computers, some perhaps consisting of only a few dozen chips, may manage very large storage systems and numerous attached peripheral devices representing most of a system's total bulk and cost. Moreover, such large systems will typically be provided with extensive applications-oriented software, translation programs for languages, and complex multi-user operating systems. Microcomputers typically are supplied with only minimal software support packages; minicomputers are in between.

We can also expect microcomputer technology to reach out in a number of important specialized directions. Special signal-processing applications can exploit many of the design and logical control techniques developed in connection with microcomputers, and this will allow an increasing part of currently analog-signal processing to be accomplished digitally. This makes possible great improvements in the fidelity of signal recording and transmission, a technical development that may eventually lead to home sound- and video-recording systems of the highest quality. Another specialized direction growing in importance is represented by fast arithmetic units to supplement microcomputers.

Over the next five years, microcomputers of truly impressive performance, incorporating advanced designs, will appear. It seems only a matter of time until microcomputers possess the 32-bit word length now employed in many large data-processing systems. Indeed, in the next decade, micros will possess the computing capabilities of all but the largest existing machines. The microcomputer, by immensely extending the range of economical computer applications, will become a major computational workhorse for the future.

VERY LARGE COMMERCIAL AND SCIENTIFIC COMPUTERS

An important trend in commercial data processing is toward increased use of large, constantly active data bases. Therefore, effective use of memory hierarchies—i.e., data-storage systems that combine very fast electronic memories for storage of transient data with very large but much slower rotating disc memories—will be a central design concern. Inexpensive, magnetic-bubble memories or systems based upon arrays of charge-coupled devices may win a place for themselves in the price and performance range somewhere between fast electronic storage and bulk disc storage. If three-level memory systems involving fast electronic, magnetic bubble, and bulk disc memories come to be widely used, hardware and software systems that help manage the resulting problem of data motion between memories of different speeds will be required.

The Reliability Problem

Organizations are becoming increasingly dependent upon the uninterrupted functioning of their computers. The trend to larger volumes of on-line data in normal commercial operations increases this dependency. Hardware and software reliability will therefore become a central issue in computer system design and can be expected to absorb a good deal of design attention. To achieve hardware reliability in the present environment of fixed design costs and falling circuit costs, designers will probably tend to use redundant circuitry more and more. The problem of software reliability is not amenable to any equally easy solution, but useful techniques of data duplication, which allow file recovery after hardware or software failure, are starting to develop. Special hardware and software features that ease the need for such data-backup operations will be considered intensively.

Distribution of computational tasks over computer networks is another approach to system reliability. Consider the case of a distributed system that interconnects several computers. If one machine fails, its tasks can be transferred to the others in the network. This achieves many of the benefits of circuit redundancy.

Competing Designs

Scientific computation exhibits an insatiable demand for raw computational power. In the search for ways to meet this demand, the computer designer may find it appropriate to use computing circuitry much more lavishly than would otherwise be reasonable—for example, duplicating elements so that many computations or computation fragments can proceed simultaneously.

Two main types of general-purpose scientific computers have been developed: single data-stream computers and vector processors. The fastest single data-stream computers, which execute a series of instructions to produce a single data result, have now attained instruction rates in the neighborhood of 20–40 million instructions per second. Vector processors, which execute similar sequences of instructions but apply them in parallel to produce multiple outputs, can generate results 10 times as fast. In a single-stream processor, computation speed can be increased, up to a certain point, by adding circuitry.

This added circuitry can serve a variety of purposes. Arithmetic operations such as multiplication can be accelerated by combining the action of large numbers of individual circuits; multiple arithmetic processing units can be provided and their use scheduled by special hardware. Overlapped or look-ahead computers containing multiple arithmetic units can be programmed in the same straightforward way as strictly serial computers, but gain additional speed by internally overlapping the various subphases of instruction execution. If pursued vigorously, the use of overlapped instruction scheduling should allow serial computers capable of executing 100–300 million instructions per second, as compared with the present 20–40 million instructions per second, to be developed by the mid-1980's.

However, this overlapped serial approach to computer design does have its limits. As the number of circuits used in a computer grows beyond 100,000, the marginal effectiveness with which computation can be accelerated by integrating additional computing elements diminishes. In this range the alternative vector-processor design begins to appear attractive, especially since the inherently repetitive design of vector processors allows relatively low-cost, mass-produced chips to be used. Moreover, since vector processors apply each instruction to many data items, their requirement for high-speed instruction handling is diminished.

The main problem of such vector processor systems is not their design or production, but the software problems connected with their use. Present problem-solving approaches cannot easily be adapted to run efficiently on vector processors. Straightforward program adaptation often leads to programs that use only a small percentage of the computing power actually available from a vector processor, but no very systematic program translation procedure for doing better than this is available. Moreover, our theoretical understanding of programs and computational procedures appropriate for vector processors is still slight. For this reason, high-speed overlapped computer architectures, which can be programmed effectively using well-established current techniques, are bound to remain attractive for scientific and commercial processors, in spite of the advantages of simplicity and ease of manufacture that favor vector processors.

Microcomputers, which are produced in very large volume and attain very high levels of integration, provide exceptionally cheap arithmetic computing capability. For this reason, it has been suggested that large-scale scientific computations be performed by large networks of microprocessors working in parallel. However, the problem of programming such systems effectively is even more severe than that of vector processors. Nevertheless, the strikingly low cost of large-volume microcomputers emphasizes the considerable significance of design cost in the total cost of large computers. Given that the quantities of computers produced and sold will range from about a thousand for intermediate sizes of machines down to a dozen or so for the very largest supercomputers, a concerted effort toward automation in chip and computer design may significantly lower costs and increase performance.

Current design procedures for integrated circuits still tend to incorporate many manual steps, driving design costs to a level that is sometimes estimated as high as $100 per transistor. This cost level is clearly unacceptable in a technology that is making it possible to put hundreds of thousands of individual transistors on a chip. We need much more thoroughly automated design practices; ones that will in time resemble the computerized translation techniques we use to generate machine-level computer codes from programs written in less detailed programming languages.

SOFTWARE: PROBLEMS AND TECHNIQUES

The rapid rise of computer science has now made the words algorithm and computer program familiar to most educated people. But there are several other words that almost, but not quite, capture these concepts: procedure, recipe, process, routine, method. Like them, an algorithm is a set of rules or directions for getting a desired output from a specified input. A program is the statement of that algorithm in some well-defined language that a computer is able to handle either directly or after mechanized translation. Although each computer program represents an algorithm, the algorithm itself is a mental concept that exists independently of any specific representation.

The distinguishing feature of both algorithms and programs is that they cannot tolerate the slightest degree of vagueness. In other words, they must be so well defined that a mere machine can follow them as written.

The difficulty of software, that is, program, development springs directly from this and from the fact that large programs are among the most complex objects that mankind has ever attempted to build. In this regard, programming contrasts sharply with other kinds of engineering development. No matter how complex, most other engineering projects will always be governed by a few dozen physical principles, whose mastery will open up a relatively smooth road to a valid design. By contrast, completion of a large programming project can require the mastery of thousands or tens of thousands of interwoven details.

Software engineering aims to control this avalanche of detail by discovering principles that allow programming to be approached systematically, organized in standardized ways, and reduced in mass by elimination of redundant details. Related goals are to find effective ways of using a computer to check the internal consistency of large programs, to pinpoint discordant details, and to define formal techniques whereby programs can be proved correct or the generation of incorrect programs rendered impossible. But progress toward these goals has been limited, so that programming costs have been a steadily growing fraction of the costs experienced by private and governmental computer users.

PROGRAMMING LANGUAGES

Early in the history of computing, the buyer of a computer expected to receive only the basic hardware. He counted himself fortunate if that hardware was fast, not too expensive, reasonably reliable, and service was available. In the mid-1950's, however, a dramatic change began. Customers began to realize that the effective use of computer hardware also required software packages, including both operating system programs to regulate the flow of work through a computer and specialized language translators that could convert formal programming language into detailed sequences of instructions for a particular machine. Except in the case of very inexpensive (mini- and micro-) computers and some very large one-of-a-kind systems, it became conventional for the computer manufacturer to supply this software. The quality and diversity of this software has become an important factor in deciding between competing computers. Recently computer purchasers have even come to expect vendors to supply major items of application-oriented software: data-base management systems, inventory control systems. or typesetting systems, to name but a few. This trend will continue.

To understand the techniques and difficulties of developing programs, one can think of the processes as an operation whose stages resemble the three main phases of architectural development: stage 1 corresponds to an architect's rough rendering; stage 2, to detailed architectural plans showing all dimensions, walls, shafts, and plumbing lines; stage 3, to full fabrication drawings showing every bolt-hole and electrical fitting.

- Stage 1. Program development usually begins with a statement of a problem in rough terms: Read the information file on company employees; then read the pay-period records on hours worked and overtime; correlate these records with the employee records, and print out a payroll check for each employee.
- Stage 2. A programmer must then convert that rough specification into a running computer program. To do so many problems of representation and of method must be faced and solved. A typical representation question would be: In what format should the employee information file be laid out?

Method questions would resemble the following: By what sequence of actions will one convert hours worked and overtime into a paycheck for an individual employee, complete with the necessary tax, insurance, and other deductions? How will error situations, e.g., overtime records that do not correspond to any employee in the employee file, be handled? What backup files will be maintained to guard against record loss, and how will they be maintained?

- Stage 3. The final step is conversion of the sequence of instructions produced by the programmer into a minutely detailed sequence of machine instructions, which answers such questions as: In what precise order will the subparts of expressions like NET PAY = GROSS PAY (1.00 — TAX RATE) — INSURANCE + BONUS be evaluated within the computer? Where, within the memory of the computer, are the various quantities that appear in such expressions to be stored? In what pattern, and to what memory locations, is the computer to jump between its elementary instruction subsequences?

Stage 2 in the above sequence is normally accomplished manually, and stage 3 automatically. A crucial question is where to draw the line between these two levels of detail. For example, how are data layouts and detailed instruction sequences to be chosen: automatically, by the computer, or manually, by a programmer?

Were efficiency no problem, the designer of a programming language could reduce the distance between stage 1 and stage 2 by systematic use of "very high level" programming languages. These languages make significant abstract structures—such as mappings, patterns of characters, curves and surfaces in space, all of which can be represented in many ways within a computer—directly available to the programmer. They also allow free use of new techniques for the combination of processes, such as parallel exploration in many directions at once and the direct programmer manipulation of programs as objects. However, free use of such powerful but abstract operations creates programs that do not always use computers in ways that attain maximum efficiency.

The opposite approach is to use a so-called assembly language, much more similar to the internal instruction language of the computer itself. Such a language makes it possible for a programmer to control all of the computer's features and to attain efficiency in a particular application. Use of a high-level language will eliminate much expensive detail, but by the same token will give the programmer less control over the machine operations and reduce the extent to which coding skill can be used to generate highly efficient sequences of computer instructions.

Efficiency-Related Considerations

Until recently, efficient use of computational power has been crucial. For this reason most of the programming languages in broad use today are sufficiently detailed to be translated easily into efficient internal instructions for a computer. However, the rapidly falling cost of computation can be expected to motivate more aggressive attempts to replace highly detailed programming with more abstract approaches. A crucial technical question is the extent to which the process of translation from condensed, somewhat abstract programs to internal machine language can be made more sophisticated. Today's program translation routines look at the texts that they translate rather myopically, concentrating on one detail at a time. To do better, one needs to develop translation routines able to uncover those significant overall facts about particular programs that good human programmers exploit. For example, to choose the most appropriate layout for a collection of data, one needs to determine all the operations a program will apply to these data, the order of these operations, and the places in the program at which all or part of the data is no longer required.

These questions belong to the discipline of program analysis and optimization. Much attention has been devoted to this area, and it is developing, but its problems are inherently difficult. Although progress to date has been slow, the current research in program optimization techniques should lead to increased understanding of this area in the next few years. Such research seeks to develop systems that will efficiently and reliably transform programs written in powerful, abstract programming languages into efficient machine codes.

COMPUTER CONTROL OF CONCURRENT PROCESSES

Despite the difficulties outlined above, programming language designers have several substantial accomplishments to their credit. There can be little doubt, for example, that by reducing the mass of detail required to put business and scientific application programs into a computer, the developers of the COBOL and FORTRAN programming languages greatly accelerated the growth of the computer industry.

The much more recent development of languages oriented toward the control of concurrent processes is another success story, and one that is still unfolding. These languages are important to those classes of programs that must function in real time and manage many peripheral devices in parallel. This requirement typifies computer operating systems. These regulate the flow of work steps through a computer and see to the simultaneous reading of input, printing or display of output, and movement of data between storage devices. The requirement also typifies military software, which may have to regulate computer systems consisting of dozens of small and large computers distributed over many command posts, aircraft, ships, and weapons. Software of this sort has been notoriously difficult to produce, delaying major computer applications or making them prohibitively expensive. Recently, however, programming languages that make concurrent-control software considerably easier have begun to appear.

These languages embody several significant ideas. The first is that of coordinated parallel processes. These arise naturally out of the need to regulate external devices having their own timing constraints. Consider an operating system that manages rotating electromechanical data-storage discs. When an input/output operation addressed to a disc is executed, a long wait for particular data to appear will be followed by a period in which these data must be quickly transferred. A natural way to organize this is by parallel processes. In this case, it is natural to think of an input/output process as an activity that monitors a turning disc until a critical position is reached, at which point the process is activated and transfers data. However, this process must alternate its activity with that of other computational, communication, and device management steps, which are in turn activated, suspended, or resumed. A second important notion is that of shared data structures, through which processes acting in parallel can communicate with one another.

These two ideas made conveniently and directly available bring a higher degree of organization to the writing of concurrent software packages; they markedly ease the writing of parallel control software, and allow flexible software systems to be created for a great variety of applications. This exemplifies the ability of programming language design to supply programmers with clean conceptual approaches to their practical problems, thereby accelerating the application of computers to new areas.

THEORETICAL COMPUTER SCIENCE AND THE CONTRIBUTION FROM MATHEMATICS

Computer science has been able to draw many of its most fundamental concepts ready made from mathematics. This fact has greatly accelerated the practical development of computing. The computer industry, for example, could not have developed as swiftly as it did without the principles of Boolean algebra from nineteenth-century mathematics. Similarly, programming found its universal possibilities and its limits laid out at its very beginning, in the work of the great mathematical logicians Godel and Turing. Generally speaking, those theoretical computer science areas that have been able to draw upon the mathematical tradition have advanced surely and rapidly.

ALGORITHM DESIGN AND EFFICIENCY

The contribution from mathematics has been particularly helpful in the design and analysis of algorithms, which, again, are patterns of steps or directions that govern the actions of a computer.

It turns out that for any desired result many possible patterns of steps or algorithms—some relatively obvious, others highly ingenious—can be used. However, the total number of elementary steps required will vary enormously with the algorithm used. For important problems, the right approach may be billions of times more efficient than a more obvious approach. Thus the choice of an algorithm to achieve a desired end is in many cases the single decision having the most impact on computer efficiency, one that can vastly exceed the impact of machine speed or code quality.

To see why this is so, consider an important but elementary problem: that of searching for a given name—for example, the name of a particular client of some large public agency—within a computer's memory. Computer memories are normally organized in sequentially numbered small subareas called cells or words. Suppose that we need to deal with 1 million such names, occupying 1 million cells. If a name is not found, the computer is to respond: "No such customer." This is, for example, a very important operation in computerized cash-dispensing systems.

If the names are held in random order within the cells, then, in searching for a given name, the computer must examine every cell, comparing the name in this cell with the name sought. No cell can be ignored, since any omitted cell may contain the sought-for name. A million comparisons will therefore be required to ensure that a particular name is not present, and half this number of operations, on the average, to locate a name that is present. Even on one of today's fast computers, this will take about a second.

Now suppose that the names are kept in alphabetical order. The search can begin by looking at the entry in the middle of the table. If the desired name is alphabetically before this middle entry, the entire second half of the table can be eliminated and need not be searched; similarly, if the name is alphabetically greater than the middle entry, one can eliminate the entire first half of the table. Thus a single comparison yields a search problem that is only half as large as the original one. The same technique can now be applied to the remaining half of the table, and so on until the desired name is either located or proved to be absent.

Since one, doubled 10 times, is 1,024, another 10 doublings gives a number more than a million. Therefore, to search an alphabetized table of a million entries by this so-called binary search method, only 20 comparisons are required. By proceeding in this more efficient way, the same computer could handle at least 50,000 different name searches in a second, rather than just one search.

As with all algorithms, two questions arise in connection with this binary search procedure. In how wide a range of circumstances will it retain full efficiency? Is binary search actually the best possible search procedure, or can we devise some other still more ingenious search method? It is clear that if an unchanging collection—e.g., names or words in a dictionary—is to be searched repeatedly, the initial labor of setting up an alphabetical arrangement is easily worthwhile, since the time saved in later searches will more than recover the cost. Suppose, however, that the collection is changing rapidly because of insertions and deletions. Is it then worth maintaining the alphabetical order. If so, how can this order be maintained most efficiently? Maintenance of order is not a trivial matter, since a naive attempt to insert a single new element into the middle of a million alphabetized elements could make it necessary to move as many as 500,000 elements up or down one place.

It turns out, however, that there are special arrangements of alphabetical data that allow arbitrary insertions and deletions to proceed just as rapidly as binary searches, so that alphabetical order can be maintained. The idea is to lay out the data items as the twigs of a branching, treelike arrangement of data. The details of this arrangement, and of the pattern of actions needed to maintain it as insertions and deletions are made, are by no means obvious. This whole approach must be regarded as a significant algorithmic invention, intellectually comparable to the discovery of a mathematical theorem.

Algorithm Analysis

Algorithm design has established itself as a major branch of computer science. Something like a thousand ingenious algorithms and groups of related algorithms have been discovered and described. This includes algorithms for many kinds of numerical computation, for searching and sorting, manipulating algebraic and logical formulae, developing optimal plans and schedules, compressing and expanding data, checking for and correcting errors, translating computer languages, working out effective geometric layouts for computer and other integrated circuits, calculating properties of important abstract mathematical structures such as graphs, and dozens of other purposes. The overall aim of algorithm design is to analyze every computational task and to find the best possible way to accomplish it. We can expect this effort, which makes effective use of mathematical tools and ideas, to continue to grow as one of the most stable and productive branches of computer science.

Since many algorithms can be devised for any task, we need to be able to calculate the efficiency of competing algorithms. In doing this, computer science has used all the most highly developed enumerative tools of mathematics, including techniques originally used in combinatorics, number theory, probability theory, and mathematical analysis. These applications have in turn begun to have measurable impact on mathematics, one that is bound to grow; for example, computer science has been important in the remarkable revival of combinatorial mathematics during the past two decades.

To calculate the efficiency of an algorithm that performs a given task is one thing; to prove that the algorithm is the most efficient among all possible ways to do this task is considerably harder. To see this, again consider the problem of searching. The binary search technique is quite efficient; in particular, the work it expends in searching a table of given size grows only slowly with the size of the table. A table of million entries can be searched with only 20 probes, and only 30 probes are needed to search a billion entries. But is this the best method? In fact, there is a technique, even more ingenious than binary searching, that is able to search a properly laid-out table of a billion or a trillion names just as fast as it can search among a thousand or a million names. It deliberately scrambles the names in a predetermined manner that allows a name to be looked up, on the average, in a fixed number of probes independent of table size. The point is that the designer of algorithms must never be content with an algorithm that is merely good. Until it can be proven mathematically, there is always the possibility of a much better algorithm.

LIMITS OF CALCULABILITY

A rigorous mathematical proof that a given algorithm cannot be bettered will always tend to be difficult, since it must examine all possible ways of accomplishing a given task. Nevertheless, by using the theoretical tools originally developed by logicians for proofs of undecidability and unsolvability, it has been possible to give such proofs in a number of cases. These same tools have been used to prove certain problems inherently difficult or impossible, and thus to define the theoretical limits of computability (Figure 19).

Proofs of this kind are particularly interesting when they show that significant problems arising in practice are inherently difficult. The bin-packing problem is an example. Here we are given a collection of sticks of various lengths and then required to arrange them end to end into groups of no more than a yard in length (although groups of less than a yard in length are allowed). The objective is to end up with as few groups as possible. (This problem arises, for example, when strips of paper of various widths have to be cut from yard-wide rolls, with as few rolls cut as possible. Another example is that of loading trucks without exceeding a stated weight limit.) Small problems of this kind can, of course, be solved by enumerating all possible groupings into two, three, or more subgroups and by testing each such group to see if it meets the criterion that no group should have a total length exceeding one yard.

However, for a case that is even moderately large (e.g., 100 sticks) this approach is totally unworkable, since the number of possible groupings is enormous. For example, the total number of ways in which 100 sticks can be separated into two groups is 2^{100}, or approximately

$$1{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000.$$

This raises the question of whether there exists any algorithm that will find a best possible solution in all cases without having to examine some substantial part of the collection of all possible groupings. Very substantial theoretical evidence indicates that the answer to this question is "no," i.e., that the problem of best possible stick packing is inherently hard. However, quite good approximate solutions can be found in most cases, as can exact solutions in many, but not all, cases.

The stick-packing problem has been related to a very wide variety of other packing, graph-analysis, and arrangement problems, all of which have been shown to be equally difficult in a precise theoretical sense. It is likely that none of these problems can be solved much more efficiently than by a comprehensive search over an exponentially enormous space of possibilities. Other slightly deeper problems—for example, various problems arising in computer manipulation of mathematical formulae—are known to be unimaginably difficult. Such results remind us that the outer boundaries of the calculable are more restricted than initial enthusiasms for the raw speed of computers might suggest. Other algorithmic approaches, such as heuristics and probabilistic algorithms, may provide approximate solutions to some of these problems. Nevertheless, since these results apply to calculating devices and robots of every kind, they also warn us

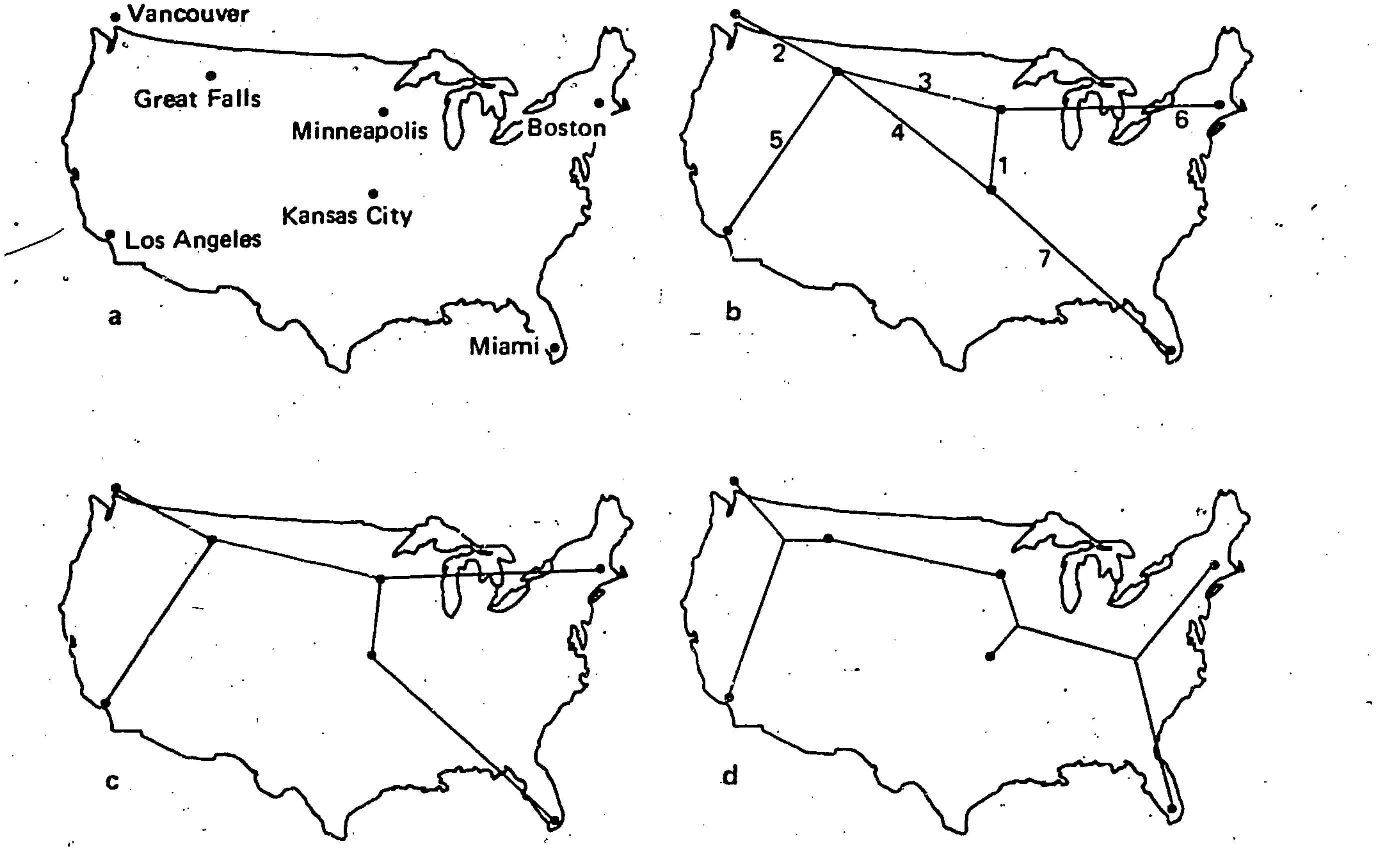


FIGURE 19 Algorithm efficiency illustrated by the spanning-tree problem. The problem is to find the shortest network of lines connecting a set of points. For example, the shortest railroad system linking a set of cities (a). If the lines meet only at cities, the problem is solvable with an efficient algorithm. First, the closest pair of cities are joined, then the next-closest, and so on (b). Lines joining cities already connected indirectly are then omitted, producing the optimum spanning tree (c). An even shorter network is possible if lines can meet outside of cities. However, there is no efficient algorithm for finding the points where the lines should meet. The optimum network (d) was found only by exhaustive search. (From "The Efficiency of Algorithms," by Harry R. Lewis and Christos H. Papadimitriou, © January 1978, Scientific American, Inc.)

that the path to development of artificial intelligence cannot be broad or trivial.

Faced with theoretical obstacles of this fundamental sort, computer scientists have begun to search for ways in which the very notion of calculability could be relaxed so as to allow a wider range of possibilities. An interesting recently discovered possibility is that of working with algorithms that need not yield correct results in all cases, provided that the probability of error can be estimated rigorously in advance and can be shown to be extremely small. A number of interesting examples, such as testing an integer for primality or a polynomial formula for validity, have been found in which by allowing even very minute probabilities of error (e.g., one error in every 10^{1000} cases processed) much faster algorithms can be devised than seem otherwise possible. Algorithms of this type illustrate some of the imaginative new possibilities that algorithm designers may be able to exploit.

COMPUTERS AND COMMUNICATIONS

The enormous advances in integrated chip technology have combined in recent years with major advances in communication technology to create a computer-communications environment that promises to have major commercial and social impact. This technological development has made large computers more usable by decentralizing access to them and moving the facilities to the everyday locations of their actual users. It has also helped the trend toward interconnecting computers from different locations into computer networks. This trend—fueled by rapidly falling computer equipment costs and also by a somewhat less steep decline in communications costs—will continue over the next five years.

These developments make it possible to connect banks, retail stores, travel reservation systems, credit verification services, and inventory control systems within a unified data-communication environment. This, in turn, has given rise to a number of experimental national and international computer networks, some of which have been in operation for some time: ARPANET, TELENET, and TYMNET in the United States and DATAPAC in Canada. Other such networks, only just beginning, are TRANSPAC in France, EPSS in England, EURONET elsewhere in Europe, and DDN in Japan.

Improvements in communications technology that have strengthened the trend toward the network use of computers include digital encoding and transmission of signals, grouping of signals into intense bursts that can share a common channel with other such bursts, new switching techniques, and satellite transmission. The basic technique of digital transmission, i.e., transmission of precisely controlled streams of bits rather than waveforms as with voice or TV, allows digital computers to be interconnected over compatible digital communications channels. It also allows digital error-correction to preserve the accuracy of information transmitted over long distances. Burst multiplexing and switching increase the efficiency of communication channel use by allowing single channels to be shared among multiple data sources and receivers. A variety of useful multiplexing techniques such as electronic fast-connect circuit switching and highly flexible time-division networks have been developed, but in recent years the packet-switching technique has become particularly important. In packet switching, data to be transmitted are partitioned into small bundles, called packets, each of which carries its own destination address and can therefore make its independent way through a communication system. Pulse-code modulation converts analog signals, such as voice or TV signals, into streams of bits, thus allowing analog traffic to flow along the same transmission channels used by computers.

SATELLITE COMMUNICATIONS

Communications satellites have developed enormously since their initial appearance in the late sixties (see p. 35). In 1970, 14 satellites had been placed in geostationary orbit, and 96 additional satellites will probably be in orbit by 1980. The decade since the launching of the first stationary communication satellite has also seen a 10-fold increase in the communication capacity of satellites, an almost 100-fold increase in their effective radiated power, and a roughly 100-fold decrease in costs. Increasing satellite power has been particularly significant since this has decreased the size and cost of satellite earth stations, from the 30-meter diameter, $3,000,000 station required for the original INTELSAT system, to the 5-meter, $30,000 station available today. Earth–station costs will continue to drop as we go to higher frequency (12–14 and 20–30 billion cycles per second) systems. Among other advantages, low earth–station costs will allow direct satellite communication between business establishments.

In addition to these important advances, the next five years will see significant advances in the following areas: larger satellites with greater on-board power generation and with on-board transmission-processing capabilities, multiple-spot satellite communications beams, and more sophisticated access and channel sharing schemes.

LOCAL COMMUNICATIONS

Satellite communication and long-haul cable and microwave systems address the long-distance communication problem successfully. But a great need remains for technical advances in short-distance communications. The local communication problem is no longer that of connecting the telephone central office to another central office, but rather economically connecting these central nodes to large numbers of individual end-users. Currently, the most prevalent local communication technology is the same one in use since the invention of the telephone and telegraph over a hundred years ago: signal transmission over pairs of copper wires. Still, wire-based communication techniques will continue to improve and will supplement the medium-range wireless and long-range satellite digital communication techniques.

One local communication problem receiving a great deal of attention today is in-plant communication, i.e., the flexible transmission and switching of large volumes of voice, image, and data traffic among a single contiguous group of user-owned premises, such as adjacent factories or offices. Current in-plant schemes commonly involve a time-shared wire or group of wires, ranging in length from a few tens of feet to a mile or so, connected to a group of terminals or other signalling devices that share access to a single long-distance, possibly broadcast, channel. A common idea in such systems is to provide a means for burst communication at full channel capacity, but to share this capacity among many data sources and receivers under the control of some mechanism that allocates the channel on a demand basis. A properly designed system of this kind can replace the large numbers of wires that are ordinarily used for local data and telephonic transmission. This technology has great potential for in-house communication in office buildings, factories, or ships.

We can also expect a continuing development of ground radio systems for computer communication;

some will divide a geographic area into local cells, while others will make use of a packet-switching technology like that suggested for long-distance communications. In the near future these will be attractive to data-communication users who must remain mobile, such as doctors, taxis, and ships.

OPTICAL FIBERS

Fiber optic transmission systems are a relatively new technology that may enhance the information-carrying capacity both of local wired communication networks and of somewhat longer-range networks (see p. 176). These systems use modulated light sources to transmit digital information along fine fibers of optical material. Compared with copper-wire systems, they offer large bandwidth (large information-carrying capacity), low weight, small diameter, freedom from crosstalk, and low cost. Fiber optics provide an information-carrying capacity as high as that of expensive shielded video cable at a cost comparable to that of the conventional copper-wire pairs used for telephones. Their ability to carry enormous amounts of data in a physically small cable suits them ideally for use in metropolitan areas, where underground conduit space is limited and difficult to expand. However, optical fibers still require special handling, protection, and splicing equipment, and are not yet fully ready to compete with copper wires.

Experimental communication links using optical fibers have already been implemented at all of the standard wideband data communication speeds, ranging from 1.5 to 274 million bits of data transmitted per second. In principle, fibers can accommodate still more enormous rates of data transmission, although at present the optical dispersiveness of fibers, as well as the limitations of the transmitting and receiving technology, do not permit more than a hundred million pulses per second on a single fiber. Current research is aimed at development of more powerful and longer lasting light-emitting elements, more complex fibers that will reduce the attenuation of optical signals below their present level of approximately 10 percent signal loss per kilometer, better fiber-splicing techniques, and more adequate optical amplifiers for use as receivers.

ELECTRONIC MAIL

The sending of messages directly between computers or terminals has a number of advantages over today's letter mail. A message can be delivered reliably and almost immediately, can be read by its recipient at leisure on a terminal, and can be acknowledged. Messages can be filed for later retrieval, answered and automatically addressed back to the original sender, and copied for retention by a sender or for transmission to third parties. Multiple addresses can be easily specified, and address lists for special purposes easily generated and maintained by each sender. Messages can be sent, not to a specific terminal, but to a person wherever he happens to be located; the receiver can retrieve his mail using a terminal wherever he happens to be. All of these facilities exist today on a number of systems.

The use of such systems is expanding rapidly, particularly within corporations. Indeed, most of the currently important computer-based electronic mail systems are used to connect dispersed locations of single firms. Rapid growth in the interconnection of these systems will soon permit electronic mail between firms. These electronic mail systems already handle computer-generated information and information directly keyed into the system. The development of facsimile scanning equipment, inexpensive high-resolution display terminals and printers, and new methods for efficiently encoding graphic information is also strengthening the capacity to transmit printed and handwritten documents. As such devices develop, they may lead to integrated communication services in which not only data and facsimile, but also voice and video, are carried on a unified network. Unified transmission of different types of traffic is already here, but the development of well-defined unified applications is taking longer than predicted.

SECURITY AND ENCRYPTION

The sharing of an electronic communication system by many people, and also the transmission of great volumes of data over easily intercepted satellite channels, create serious issues of privacy and require special techniques both to guarantee security and to authenticate message authorship. This is largely a software, rather than a hardware, problem and has been addressed through improved encryption techniques. Such techniques protect computer communications both from unauthorized reading and from attempts to insert, delete, or alter their contents. Such protection will become increasingly important as electronic transmission of mail, stock market transactions, and banking transfers begin to be more common.

Promising research is now under way in the development of electronic signature systems that can be used to sign or notarize transactions—for example, in messages between stockbrokers. Certain types of arithmetic computations, which are inherently difficult to reverse, probably can be made the basis for unforgeable electronic signatures. Such signatures can be added to any messages and used by the sender to authenticate his presence while a message's receiver can use a corresponding decryption technique to verify the source of a message.

OFFICE SYSTEMS

The past few years have demonstrated the usefulness of office typewriter systems based on computer-driven machines with televisionlike display screens. Savings in labor and speed of operation have justified the installation of such systems at many large- to medium-sized offices. These systems, which allow a great deal of flexibility in text editing, make it possible to change a particular word automatically wherever it appears in a manuscript or completely rearrange paragraphs and sentences. Other new systems can verify and correct the spelling of every word in a manuscript. Office directories are being made available on computer displays; here the ease of making new entries, deletions, and fast look-ups, to say nothing of the savings in printing costs for directory reprints, are all very attractive. These new systems are already beginning to change the offices in which they have been installed and are starting to have noticeable impact on day-to-day office procedures, on the way in which information circulates within offices, and on the daily work of secretaries.

ELECTRONIC FUNDS TRANSFER SYSTEMS

Numerous forecasts proclaim that electronic funds transfer systems will shortly usher in an era of checkless banking, if not a completely cashless society. Almost as numerous are articles dismissing this perspective as illusory, arguing that the public does not trust computers, wants to have pieces of paper in hand as evidence of funds available, and prefers to delay payment of a bill until some time after it has been received. Neither of these polar viewpoints captures the fact that electronic funds transfer is not a single entity that will either take over or fail completely. Rudimentary electronic funds transfer systems, e.g., funds transfer by telegraph, have been with us for over a century. The question now is in what form and how rapidly such transfers will grow, in what areas they will grow most rapidly, and what their impact on commerce will be.

Certain environments in which the growth of electronic funds transfer has been particularly rapid are easy to note. One such is the telephonically—and soon to be electronically—verified credit card. Whenever a customer submits such a credit card for verification and it passes, the credit card company is in effect instantaneously creating money, giving credit to the customer and guaranteeing payment to the vendor, whether or not the customer has cash on hand at the moment. At the end of the month the customer receives a consolidated statement (which testifies to his continuing desire to see and validate a written statement before authorizing payment) and discharges his obligation, generally by writing a check on a bank. Further developments are possible here, including the presentation of monthly statements at a terminal within an electronic mail system and authorization of payment from a terminal without any paper check being necessary.

Large corporations are increasingly making credit arrangements directly with business customers rather than paying a bank to arrange them. Computers, which make it easy for an accounting office to do complex, bankerlike calculations, are one factor working towards this do-it-yourself financing. Of course businesses, like individuals, are eager to maintain their access to cash float; but this can be handled in a computerized business environment by appropriate systems of discounts or fees for early or late transfer of funds.

All this makes it plain that the growing use of electronic funds transfer will affect ordinary consumers, businesses, and banks in ways that should be better understood. For example, consumer privacy may be compromised by systems that allow all the details of a consumer's pattern of purchases to be marshalled centrally. The pattern of cash float—the interval between billing and payment—upon which firms now depend may change radically. These issues, which have already received considerable attention, need to be explored further, since they may have significant impact on the whole structure of major financial institutions.[1]

HOME COMPUTING

We can expect that within a decade most homes will contain at least one computer and that within two decades virtually every home will. The homeowner may not know that it is there if, for example, a computer chip controls the tuning of a color TV or the defrosting of a refrigerator. In this section, however, we concern ourselves with more explicit home use of computers. Many such applications are conceivable although speculative:

- The combination of computer and telephone may be useful for remote control of household appliances and heating systems, delivery of personal reminders, medical monitoring, security alarms, and baby sitting.
- A wide variety of useful information services may be delivered via home computer terminals. These can

provide anything from encyclopedias or libraries of relatively fixed information to highly dynamic information on current news, market conditions, supermarket specials, and want ads. Hundreds of thousands of pages of useful, highly organized information may be accessible to the subscribers to such systems through their telephones, home computers, and attached displays.

• Computer games and clubs linked by computer may provide an attractive social milieu for many enthusiasts.

Home computers, which are steadily becoming more powerful, can be terminals communicating with larger and more powerful central computers. This raises the question of the degree to which home computers will offer extensive computational capability versus the extent to which they will remain limited devices depending strongly on a central facility. Related to this is the question of how extensively home computers will be tied to the nation's telecommunications network. Currently advertised home computers—over 100,000 have already been sold in the United States—are sold as stand-alone devices, with little emphasis on their communications capabilities or even on the fact that they are often linked on hobbyist networks. Nevertheless, it is possible that the use of home computers for communication will grow very large. They may, for example, come to be used as electronic mail stations.

It is also true that the home computer of the future is likely to provide a substantial computing capability in its own right. Available for only a few hundred dollars, future home computers may have speeds now associated only with large computers. Nevertheless the home computer will need to become easier to use if its use is to spread beyond hobbyists.

In time, the need for a link to a central computer will become greater since a central computer can provide major information services, such as daily news bulletins and market quotations. Moreover, natural language processing, graphic capabilities, and sophisticated error handling are all important for most people, and these require substantial quantities of memory likely to be available during the next five years only through large central computers.

Sophisticated software will be required for home computing, and this raises significant questions. Vendors of home computers now cater to a market of hobbyists who enjoy programming. Typically, such vendors put BASIC or some even more primitive programming language on their machines and let the buyer write useful programs. Very little by way of prefabricated programs is offered. However, the broader public will not want to program home computers but to use them directly and flexibly. This points to a very large market for software packages, including programs capable of dealing with various types of natural language queries. Development of such packages, and of the advanced programming tools needed to build them, should be spurred by the increasing use of home computers, but not much has yet been done. Likewise the legal protections required for the adequate functioning of a broad software market, e.g., protection for intellectual property maintained in electronic form, have not yet been adequately defined. Because of these unresolved problems and its inherent technical complexity, software is likely to be a thorny obstacle to the spread of home computing.

Traffic patterns on other communications networks may be substantially altered as home information systems become practical, particularly to the degree that they are linked to central or other remote home computers. Over time, this can have a major effect on communication facilities that have been designed to accommodate traffic patterns characteristic of voice communications. This could substantially change the economics of proposed systems. Although there is little understanding of what these effects will be, they cannot be ignored.

A related question is that of the balance between specialized, one-program devices like computer games, and multifunctional, general-purpose terminals. Depending on the direction in which software for home computers evolves, the trend may be either to single computers, or groups of identical computers, providing many services, or to a variety of separate devices that compute, but which individually provide very different services and are called by different names. Regulations established by government, as well as pricing decisions taken by the administrators of centralized computing services, will affect this outcome.

REGULATORY ISSUES

For computer communication systems to attain their full potential they will have to be ubiquitous, cheap, and very easy for most persons to use. A number of key regulatory uncertainties will have to be resolved for this to be possible. Are systems of computers and terminals data-processing systems (and so unlicensed) or communication systems (and therefore licensed)? Should electronic mail be viewed as a service, having the character of a national monopoly and therefore to be offered on a universal basis by a monopoly carrier, be it the Post Office or the telephone industry, or simply as a set of products to be developed and offered by a wide range of competitive suppliers? What will be the impact of electronic mail on the postal system?

What charge structure is appropriate? How can the needs of small users be served best?

Distinguishing Computers and Communications

In the all-digital systems of the future, it will be hard to draw a clear technical distinction between communication services and computer services. Both will use computers to manipulate bits and to transmit them between points. Identical processing and transmission equipment will be used to pull segments together into a text, edit it, address it, and distribute it. Yet the Communications Act of 1934, by authorizing the FCC to regulate electronic communication, requires the commission to seek some legal formula distinguishing between digital communication and digital data processing if it is not also to be drawn into regulation of the computer industry. That the FCC should seek some such formula is understandable and proper, but it would be an illusion to believe that such a formula, whatever it is, can easily be applied to technical reality or that this formula could be effective for very long, considering the rapid and unpredictable course of technology.

While some portions of the communication/computer industry naturally require regulation, others are best served by fostering vigorous technical and economic cooperation. Regulation of limited and eventually congested satellite "parking space" and communication bandwidth is of course required. As with ground radio, a limited spectrum must be organized and allocated. Continued regulation is also likely to be required in the use of public right-of-way for information distribution systems, whether copper-wire pairs, coaxial cable, or optical fibers. In its role as guardian of the quality of public communication services, the FCC must also bear in mind the potential impact of sudden and heavy computer-generated demands for communication bandwidth on the services received by more conventional systems users. Moreover, alongside the electronic compatibility and standardization requirements of the past, a new and far more complex set of compatibility issues—this time related to software—is growing. However, in responding to these problems by new regulations, the FCC must also avoid hampering the use of new digital communication equipment or the provision of new computerized communication services.

Still another range of problems relate to the First Amendment. The Constitution guarantees the freedom of the press, but not particularly the freedom of business machines. Freedom of the press has been construed to include radio and television. Are computer communications also protected? In this connection the rather different legal history of the telegraph and of radio is interesting. If one looks back at the history of telegraph law, few references to First Amendment issues are seen. Yet such references are frequent, and have often been central, in the legal history of radio. The reason is that the telegraph has been thought of as a business machine, not a device to which a citizen would rush to express his concern. By contrast, radio was a means of expression from the beginning. Although computer communications were first regarded as adjuncts of business operations, there is good reason to believe that such communication will eventually encompass a great part of our society's means of interpersonal contact and expression. If home computer terminals become a major channel for mail, for consumer information, for education, and perhaps for political campaigning, they may end up being as important to the twenty-first century as the printing press has been for the past 500 years. If so, the First Amendment will become important to practices and arrangements for the regulation of computer communications.

Besides regulation and competition, another important force affecting the rate of development of new computer/communications services is the U.S. government's own large equipment and service purchases. The government has procured the most sophisticated technologies available for its national security and space programs, but not for its ordinary civil activities. Here government has a significant chance to promote technological innovation; to develop a technical base upon which more sophisticated applications must rest.

ARTIFICIAL INTELLIGENCE

Since this report is concerned with the next five years rather than the next few decades, we have focused most of our attention on practical technological developments that are most vigorous today. Even so, a short survey of the revolutionary long-term efforts to create humanlike capabilities in computers is in order.

The computer is not just an adding machine; it is a universal information-processing engine, one which may be capable of duplicating many of the most characteristic capabilities of the human brain. This possibility, intuitively grasped by such pioneers of the computer age as Turing and von Neumann, has continued to motivate and fascinate computer scientists.

But even if we admit construction of artificial intelligences and superintelligences to be the ultimate perspective of the computer age, it still does not follow that one or another branch of computer science should be emphasized, nor can we tell what discoveries, still hidden from us, will emerge, nor whether the necessary development will take decades or centuries.

Work on programming computers to display humanlike problem-solving characteristics is widespread and ranges from work aimed at modeling human performance in specific problem-solving areas, such as medical diagnosis, to programs whose search techniques are based on heuristics and "rules of thumb" and even to the use of computerlike models of thinking as a tool in cognitive psychology. The use of human intelligence in real world situations requires an intimate coordination of thinking with seeing or hearing and motor action. The branch of artificial intelligence research that seeks to design systems capable of sensory-motor coordination is generally called robotry.

While only a small fraction of artificial intelligence research has been directed toward sensory-motor tasks and robotry, this domain does provide a rather broad view of the nature of the research problems across the whole field. For this reason, and because of the growing practical importance of industrial robotry, we will look a little more closely at the current status of artificial intelligence research on robots.

ROBOTS

The general goal of research on robotry is the development of versatile robots that can perform varied mechanical tasks in homes, factories, offices, and out-of-doors. To be successful such robots require a broad range of abilities. They must be able to plan a sequence of actions and to simulate and test the execution of these actions in a model of the environment. Continuing correspondence of the model to a dynamically changing environment will depend on a robot's ability to acquire and analyze sensory information. When surprises occur, the robot must be able to plan and carry out corrective actions.

Even though these abilities do not seem to require high "intelligence," their realization in machines is a complex matter. Techniques powerful enough to permit development of truly flexible robots still elude us. However, specialized robots that can automatically perform many assembly, materials handling, and inspection tasks are now coming within reach.[2] These systems are much more versatile than the numerically controlled machine tools now in use; they are controlled by small, general purpose computers that respond to touch, force, and television inputs and are guided by general descriptions of the task. The evolution of robot assembly systems should henceforth be continuous. Ultimately these can be expected to develop into "automatic factories"—versatile computer controlled systems for processing raw materials into finished goods.

The more elusive goal of a general service robot will require programs capable of analyzing major sensory inputs and responding to them with something like human sophistication. Researchers are already attempting to duplicate all important sensory functions, including speech recognition, language analysis, and analysis of visual patterns. Machines are now capable of recognizing small numbers of words and short phrases spoken in isolation, but correct response to continuous speech is much more difficult. A recent five-year research effort demonstrated the feasibility of experimental systems capable of responding to spoken vocabularies of several hundred words, but much more basic work is needed before practical speech recognition systems can be built. Challenges will continue to inspire work over many decades.

LANGUAGE AND ROBOTS

It is of course much easier for computers to deal with information in printed than in spoken form; but even for written material the subtlety of language puts truly flexible natural language processing somewhat beyond a computer's grasp. Translation between natural languages and internal computer languages cannot be accomplished by rudimentary dictionary look-up techniques. To understand natural discourse, one must have a reasonably good model of the content of the discourse. (To see this, think of what is needed to understand the following sentence properly: "The rancher, returning his pen to his pocket, stepped into the pen.")

The development of such models and their integration into language-analysis programs is a formidable task, on which only experimental starts have been made. However, there are information retrieval systems that allow one to communicate successfully with a computer using subsets of natural English dealing with restricted subjects. Meanwhile, researchers are currently experimenting with more sophisticated natural language systems possessing larger knowledge bases. Here it is reasonable to hope that within the next ten years computer teaching systems, where natural language capability is highly desirable, will be able to carry on adequate English dialogues with students in certain specialized subjects.

Large bases of knowledge concerning particular subject domains are also of central importance to evolving computer systems of the consultant type, in which fragments of knowledge collected from expert practitioners in a specialized field, such as medical diagnosis or ore prospecting, are collected and organized to attempt duplication of human judgments.

SIGHT AND ROBOTS

General purpose robots would certainly have to process visual information. The problem of visual perception by computer is a minefield of subproblems. At the simple end of the scale, systems already exist for comparing what the robot's eye would see with stored pictures; these are used today to guide a missile toward a target. The problem of interpreting objects in photographs collected routinely by weather and surveillance satellites is somewhat more difficult, but still feasible. Reliable techniques have been developed to automate parts of these photo-interpretation tasks. Similar systems can also be used to spot roads, bridges, or railroads in aerial photographs. Related techniques allow machine reading of printed text.

Somewhat more sophisticated vision systems are now being introduced into factories to control industrial robot arms during materials handling, inspection, and assembly. Research in machine vision is concerned with such issues as reconstruction of the three-dimensional shape of a solid body from the pattern of lines and corners that constitute its image on an electronic retina. Work has also begun on the detection of the presence and form of bodies, parts of which are obscured by other bodies. Visual abilities of this level of sophistication will be required by general purpose robots intended for household or outdoor use.

The widespread research effort to endow machines with abilities of this sort is contributing to the development of a computational theory of intelligence that is beginning to have an impact on psychological theory. Significantly, one of the most important scientific insights resulting from the work is the realization of how complex the processes underlying intelligent behavior must be and what a vast, but somehow patterned, storm of perception and selection must underlie every conscious act or thought. However, beyond this sheer complexity, there seems to be no conceptual barrier to the eventual synthesis of intelligence in machines, which may very well be the final outcome of the march of technology that we have described.

For now we can epitomize the attainments, hopes, and disappointments surrounding the field of artificial intelligence by glancing at one of its amusing, yet typical and challenging efforts—the design of chess-playing computer programs.

The first serious attempts to develop game-playing programs go back to the 1950's. Progress with simpler games (such as checkers) was swift and dramatic. Chess proved more difficult. In fact, computers still cannot match the ability of the human chess master to discern critical sequences of positions without explicitly examining a very large number of cases. Nevertheless, chess-playing programs have improved steadily. In 1977, a small but significant landmark was passed: CHESS 4.6, a program developed at Northwestern University, drew one game and won another (although losing its match) against British chess master David Levy. And these programs continue to improve.

OUTLOOK

The following outlook section on computers and communications is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

Over the past two decades, electronic computers, initially the esoteric tools of a small scientific community, have become familiar presences in daily life. As a result, the shopper who keys in for a quick infusion from a computer-controlled cash dispenser and the child whose exposure to computers begins with a video game are both beneficiaries of this new technology. More and more of the information upon which business and administration depend is available in machine-readable form and managed by computers. The technology that has made this development possible has drawn ideas from many areas ranging from mathematics to linguistics. But its driving force during the last decade has been the amazing decline in the cost of semiconductor chips. As chip costs continue to fall during the next five years, this trend will put more and more computing capability in the hands of the public. In turn, this will lead to further improvements in all the areas now touched by computers and an expansion into new fields.

OUTLOOK FOR TECHNOLOGY

The development of semiconductor technology has taken us from 1,024 bits of information per chip in 1971 to 64,000 bits per chip in 1978 and has kept chip cost approximately constant, achieving a 16-fold cost reduction. Looking forward, it seems reasonable to anticipate 256,000-bit memory chips in the early 1980's and a million-bit memory chips by the mid-1980's.

During the same period, microcomputers evolved from minimal units involving 2,000 transistors in 1971 to devices that today integrate 30,000 transistors on a chip. This progress has made it possible to enjoy the size savings and low cost of microcomputer control in hundreds of applications, from sewing machines to television games. Similar progress in size, costs, and applications will continue over the next five years. The microcomputer of the mid-1980's, often consisting of only a few chips, will be capable of truly impressive performance and will employ advanced architectures and likely the 32-bit word length used by the larger systems of today. This microcomputer will not only extend the range of economic computer applications, but will also become a major computational workhorse of the future.

Device speed has improved from the microsecond speeds of early integrated circuits to today's fastest commercial chips, capable of adding a pair of decimal digits in less than five-billionths of a second. This development should also continue, but at a decelerating pace unless the industry introduces new technology.

Two possible future technologies are bubble memories and Josephson junctions. Magnetic bubbles, which allow data to be stored very densely and moved magnetically, can provide fast, reliable bulk-data storage and retrieval. Josephson junction technology is a superspeed logic technology that may begin to be used, at least for special applications, within the next decade.

OUTLOOK FOR COMPUTER SCIENCE

Programming, the writing of instructions to tell computers what to do, has become a major element in the cost of computer operations. Software engineering is intended to control the avalanche of detail that makes programming so expensive by discovering principles that allow programming to be approached systematically, organized in standardized ways, and reduced in mass by elimination of all redundant details. Related goals are to find effective ways of using a computer to check the internal consistency of large programs, to pinpoint discordant details as fully as possible, and to define formal techniques whereby programs can be proved correct or the generation of incorrect programs rendered impossible. Progress toward these important goals has, however, been limited.

An important field of research is the development of new, more powerful programming languages that can make program development easier without losing efficiency. Improved techniques for writing programs to control many simultaneously active devices are also being developed. But since all this is basic research, some time will elapse before its impact is felt by either the industry or the public.

Computer science in the past has made use of a great many silent borrowings from mathematics. Theoretically work is continuing in many areas related to the mathematics of computers and is gradually transforming the *ad hoc* design approaches of a decade ago into a more mature, theoretically based undertaking.

The possibility of using computers to create artificial intelligences displaying humanlike capabilities has continued to fascinate and inspire computer researchers. However, except in a very few areas, we are still far from being able to do this.

SOCIAL IMPACTS

The use of computers in communications has already begun to have an important effect on the way we live. The airline reservation systems, long-distance bibliographic searches, and credit-card verification schemes of today will expand during the next five years, as will office text-editing systems. The coming period will also see the spread of computing power to more people than ever before.

Centralized computers—connecting users by various telephone and direct-wire links—characterize both the early days of computer use and many systems still in use today. There is, however, a trend away from centralized computers toward decentralized systems in which significant computing power is located as close as possible to the actual users.

Computer technology has made many new ways of communicating possible. Examples include data networks, satellite and digitally encoded telephone communications, optical fibers, electronic mail, text editing and distribution systems, and electronic funds-transfer systems.

The role of these computer communications systems raises basic regulatory questions; for example, whether the First Amendment applies to computer communications, and whether security and privacy are adequately guarded by current encryption techniques. These questions, however, are overridden by the fact that progress has come so rapidly that computer technology and communications technology are converging. This raises important questions for an agency like the Federal Communications Commission as it seeks to protect the public interest.

REFERENCES

1. *EFT in the United States: Policy Recommendations and the Public Interest.* Washington, D.C.: National Commission on Electronic Fund Transfer, 1977.

2. Will, P.M., and D.D. Grossman. Experimental System for Computer Controlled Mechanical Assembly. *I.E.E.E. Transactions on Computers* C-24(9):879–888, 1975.

5 Energy*

INTRODUCTION

The U.S. energy sector is changing rapidly. Oil and natural gas, which provide three-fourths of our energy and upon which depend most of our technology for transportation, space heating, and industrial heating, are being depleted and must be conserved. At the moment, and for the next few decades, the principal alternatives are coal and nuclear fission, mainly in the form of electricity.[1]

Both of these alternatives have limited futures given the current technologies. By the turn of the century or somewhat later, depending on energy consumption rates in the meantime, they in turn must begin to be replaced. We must then be well on the way to deploying sustainable, environmentally acceptable, long-term energy sources, compatible—in the interest of international stability—with the changing roles and aspirations of the developing countries.

Meanwhile we must use what is available—coal, nuclear fission, vigorous energy conservation, some solar energy, and small amounts of geothermal heat—to reduce our dependence on oil and natural gas. However, the rising concern with environmental quality, and the political and economic friction of making large industrial shifts, will make this difficult and will tend to focus attention on short-term expedients. It will be necessary to take broader, longer views of this problem than those to which we are accustomed.

Many of our energy consumption systems cannot be substantially changed in one or two decades. Large urban areas, where most energy is used, have made their greatest growth in the past 40 years. During this period the real price of energy fell as large oil and gas discoveries were made worldwide, electric utilities took advantage of technical refinements and economies of scale, and the environmental costs of energy use were largely unknown or ignored. People thus paid no penalty for energy use that under today's conditions—and tomorrow's even more—would be considered very inefficient. Long-distance commuting by energy-inefficient automobiles became commonplace. Too much insulation was a waste of money. The size of the average house rose, and more and more were detached, with all four walls exposed to the elements. Most industries optimized their capital equipment with energy consumption as a minor consideration.

*This chapter benefitted greatly from the work of the National Research Council Committee on Nuclear and Alternative Energy Systems (CONAES). However, the contents of this chapter have not been reviewed by the Committee and do not necessarily reflect its views.

That is now changing. The price of oil has quintupled since 1973, with coal and gas prices not far behind, and U.S. dependence on oil imports has increasingly apparent political and strategic implications. Nuclear fission, once promoted as a source of clean and nearly limitless power, faces major and increasing difficulties, although it could have a productive future. While coal is abundant, it is ironic in this time of general concern about the environment to be forced back to a fuel earlier abandoned for most uses in part because of its potential for pollution.

Still, only coal and nuclear fission can be widely available over the next two or three decades, and we must find some acceptable combination of the two, with much less reliance on oil and natural gas. We must also do the work that will be necessary to put sustainable long-term energy sources in place.

SOME IMPORTANT TIME PERSPECTIVES

Difficult choices arise. Short-term expedients may foreclose serious long-term options, and the nation must take great care that its accommodation to today's realities does not limit its scope in dealing with the more distant future. For example, strong emphasis on expanding domestic oil production, without comparable attention to how it is to be used, may tend to deepen our dependence on oil and to intensify the harshness of the eventual and inevitable withdrawal.

Financial and political institutions are more responsive to the short-term view dictated by market rates of return and by electoral intervals than to the longer and broader view determined by the depletion of domestic resources and the international implications of growing reliance on imported energy. It will require an educated public and enlightened officials to temper this disparity through responsible efforts to avoid the future economic and political difficulties to which short-range expediency can commit us.

The shortest meaningful time perspective for serious social decisions in the private sector is determined by normal market expectations of rate of return on investments. This leads typically to time horizons usually of five years, and rarely as great as 10 years, in present industrial practice. Investments maturing later are relatively unattractive compared to more immediate payoffs. Exploratory research aimed at very distant returns can be justified, but the cost must be low by normal market rules. The political sector has a similar time perspective, determined by the intervals between elections.

For energy technology the shortest time perspective is given by the time it takes to develop and commercialize new major technologies: solar power, new and cleaner ways of using coal, more resource-efficient nuclear fission reactors and fuel cycles, nuclear fusion, and mature energy conservation technologies. This time varies somewhat depending on the technology, but many take 20 years or more, except under the most exceptional conditions and strong government incentives. Civilian nuclear power, for example, which was accorded remarkable federal priorities, was first seriously explored in the United States in the early 1950's and only by the early 1970's was well enough developed to contribute 2 percent of the nation's electric power. (But the industrial momentum built up during all that time led to a 12 percent contribution in 1978.) The magnitude of a development program similar to that of nuclear, and the length of time before it could repay the initial investment, place it clearly beyond the view imposed by the market.

This disparity between the economic and technological time perspectives leads to market exploitation of existing options and neglect—possibly foreclosure—of potential new ones. Governments can correct this imbalance by either underwriting the long-term development itself, as a nonmarket social good, or constructing appropriate market signals—tax and other incentives, regulations, and so on—to stimulate option development by the private sector. Governments in fact do both.

Several yet-longer time perspectives are also important. The time it takes to deplete a particular energy resource, so that more supplies become too expensive for their accustomed uses, should exceed the time it takes to develop and deploy alternatives or to adjust to a social and technological position of doing without. Thus, the perception that the resource depletion time for oil and natural gas in the United States may be comparable to or shorter than the time it will take to produce a replacement is a principal cause of concern. It leads, for example, to demands for new technologies on short notice.

Another important time perspective is even longer. The urban and industrial infrastructures of civilizations, which heavily influence our energy use, can persist for centuries. Cities, for example, last for centuries and change fundamentally only over generations.

MID-TERM ENERGY OPTIONS

In the next two or three decades, the range of energy options available to us is rather strictly limited to existing technologies. Oil, natural gas, coal in various forms, and nuclear fission must among them supply most of the nation's energy to 2000 and somewhat beyond. The declining supplies of oil and gas further limit the scope of our actions, forcing us to rely increasingly heavily on coal and nuclear fission for any real increase in energy supplies. Given the many environmental, political, and resource uncertainties plaguing coal and nuclear fission power, it is obvious that energy conservation must play a large and increasing role in the national energy programs. But no one source of energy or conservation alone will suffice.

DISCOVERING PRIORITIES: CONSERVATION FIRST

The broadest questions that can be asked about energy policy deal with the balance of emphasis between supply and conservation. In general, supply is relatively overemphasized, because the supply sector is simpler and better organized to provide its products and to receive and recognize rewards. Energy conservation tends to offer rewards that are received later in time by a diffuse and ill-organized population. Thus effective energy conservation requires either generating a strong public awareness or providing government financial incentives, or both.

Energy conservation consists of two different categories of activity. The first is relatively simple cessation of present waste: Turn down the thermostat instead of opening windows, turn out the lights when leaving the room, don't leave the car's engine running when you go back into the house on a last-minute errand. These steps reflect simple social concern and thoughtfulness, and should become more general as people are made aware of the costs of waste. The second category concerns deeper changes in the economic structure, the designs of what we use, and public incentives. It is here that we focus our principal attention.

Energy supply costs money, and the marginal cost of energy increases as more is demanded; however, in the relatively pristine field of energy conservation large savings can be captured at low cost. An analysis by Gibbons and his colleagues[2] suggests that even if real energy prices remain roughly constant from now to the year 2010, all sectors of the economy will find it profitable to raise their energy efficiencies significantly—by as much as 25 percent in some industries.

There are many good technological opportunities for using energy more efficiently and thus conserving; for example, leveling electricity demand over the day, storing energy, using electric automobiles in cities, developing new thermal cycles, and using new less-energy-intensive processes for industry.

Related to energy conservation is the question of how energy use and GNP (a convenient but imperfect measure of public well-being) are related. In the short run, reducing the number of Btu consumed reduces GNP, but the long-term possibilities are much better. As capital equipment is retired it can be replaced by equipment designed to use less energy, generally at higher cost.

The time scales for such transitions are different for different things. Well-built houses last a hundred years, but recent experience in residential energy conservation shows that old houses can be much improved; new ones require only half as much energy or less, for heating. That comes out to halving the energy requirement per unit of service in 100 years, or about 0.7 percent per year. Cars last on the average 10 years, and we are requiring that the fleet fuel-efficiency of new cars almost double between 1976 and 1986. The life of industrial plants is intermediate, averaging about 20 years; new equipment uses typically perhaps three-quarters as much energy as the old. Applying those simple numbers, we find that industrial energy per unit output could be approximately halved in about 45 years.

Further detailed analyses,[3] taking into account these and many other replacement possibilities, suggest that, overall, energy used per unit of GNP could be as much as halved in 40–50 years, with negligible effects on GNP. This means that conservation incentives, including rising energy costs, could be structured so that by the year 2020 half as much energy as is now used would suffice for today's national total of goods and services; that is, the same amount of energy could accomplish twice as much. The detailed strategy for doing this would involve frequent reviews and appropriate corrections from time to time.

The key to effective energy conservation is planning for the long term in the sense just described. Attempts to induce efficiency by replacing houses, cars, and factories before they pass their useful lifetimes requires that useful things go unused, which in turn generally implies drops in both energy consumption and GNP growth. As stated earlier and repeated here for emphasis, short-term energy consumption and GNP tend to be closely coupled (as experience in the 1973–74 oil crisis demonstrated); but over periods of decades large beneficial changes can be made.

An example of how one group perceives the range of possibilities will help to focus this discussion. Gibbons and his colleagues,[2] in a study for the National Research Council, projected U.S. energy consumption to the year 2010, using four scenarios representing a range of plausible movements in energy prices to that date. Table 1 gives assumed prices in 2010, with the 1975 prices listed for comparison. The scenarios were developed on the assumption of 2 percent average GNP growth between 1975 and 2010; a variant of scenario III, which allows for 3 percent average GNP growth, is included. Table 2 shows the general results of this scenario analysis. The highest

TABLE 1 Assumed Energy Prices for 2010 Demand Scenarios

	Energy Conservation Policy	Consumer Price ($/Million Btu)			
		Distillate Oil	Natural Gas	Utility Coal	Electricity
Scenario					
I	Very aggressive; deliberately arrived at reduced demand requiring some life-style changes	13.49	14.84	3.24	26.37
II	Aggressive; aimed at maximum efficiency plus minor life-style changes	13.49	14.84	3.24	26.37
III	Slowly incorporates more measures to increase efficiency	6.74	7.42	1.62	15.82
III_3	Same as III, but 3% average GNP growth	6.74	7.42	1.62	15.82
IV	Unchanged from present policies	2.81	3.09	0.81	7.91
Actual					
1975		2.81	1.29	0.81	7.91

TABLE 2 Energy Demand for 2010 (quads)

	Buildings	Industry	Transport	Total	Losses[a]	Primary Consumption[b]
Scenario						
I	6	26	10	42	16	58
II	10	28	14	52	22	74
III	13	33	20	66	28	94
III_3	17	46	27	90	44	134
IV	20	39	26	85	51	136
Actual						
1975	16	21	17	54	17	71

[a] Losses include those due to extraction, refining, conversion, transmission, and distribution. Electricity is converted at 10,500 Btu/kWh; coal is converted to synthetic liquids and gases at 68 percent efficiency.

[b] These totals include only marketed energy. Active solar systems provide additional energy to the buildings and industrial sectors in each scenario. Under the most aggressive assumptions (Scenario I) this amounts to 5 quads.

and lowest scenario totals differ by more than a factor of 2. The panel found the whole range consistent with the assumed rate of economic growth; the variations in energy consumption reflect the variation, from scenario to scenario, in energy prices and other incentives to substitute inputs such as labor and capital for energy in the economy.

THE EXAMPLE OF CITIES

It is in and near cities that most energy is used, and urban life shapes and is shaped by how energy is used, both positively and negatively. Air pollution peaks in cities, but so does the ease and efficiency with which heat, light, and transportation can, at least in principle, be provided. The ease with which cities can adapt to changing energy realities is therefore an important question.

There is a good deal of inertia in the nation's urban and industrial infrastructures. Buildings and industrial plants have lifetimes measured in decades. A realistic conservation strategy must therefore deal not only with the opportunities afforded by the new but also with the resistance to change of the old.

The kinds and amounts of energy demanded depend in complicated ways on ingrained features of urban structure and urban life. Most of the large urban areas in the United States had their greatest growth in the past 30 or 40 years, when energy prices were falling and energy efficiency thus mattered little. Settlement and transportation patterns, building standards, and industrial equipment all developed in this period in essentially their present forms, and to varying degrees they resist fundamental change. Basic transportation and settlement patterns, for example, may persist for centuries.

In the long run, one may speculate that with escalating energy costs, cities will slowly evolve into more compact forms, with less suburbanization or more focused suburban centers. Given the probable energy supply options, cities may come to depend more on electricity. Basic services may be supplied from larger energy units to take advantage of district heating using "waste" heat from nearby power plants and industrial facilities, as is now done in some parts of northern Europe.

But there are many improvements that can now be made. Table 3 lists both short- and long-term prob-

TABLE 3 Energy Conservation

Problem	Relation to Energy	Consequences If Problem Is Unattended	What to Do?	Some Impediments to Implementation
Housing	Poorly insulated buildings cause heat loss, expense (this is but one example)	Housing less desirable because of increased living cost. Poor people suffer most	Insulate, retrofit, new housing codes, public education	Building codes, initial cost, lack of information, lack of low interest loans or tax rebates, landlord-tenant dilemma (who pays?)
Pollution	Mainly caused by high energy use	Bad health, loss of economic values, lowered quality of life	Stricter standards and better energy technology for cities, especially with respect to automobiles. Also cogeneration, district heating where appropriate.	Reliance on cars for commuting, ability of business to escape antipollution laws. Lack of information on pollution effects.
Transportation	Reliance on energy-intensive modes	Pollution, congestion, run out of energy resources	Higher taxes on parking, vehicles. Bikeways. Zone for high population density near public transport. Buses (probably electric) and rapid transit extensions where feasible.	Special interest groups, commuters, desire for personal security and private transportation. Substandard mass transit.
Waste heat from power plants	Obvious	Poorer urban microclimate, higher energy use, pollution	District heating, cogeneration	Utility regulations, piecemeal planning and zoning, short time perspectives, lack of overall policy
Urban waste	Energy content equals several percent of city's needs	Environmental degradation, waste of resource	Recycle, pyrolize, methanate, etc.	Disorganization of recycle industry, organization of electric utilities and of urban refuse sector. Regional political fragmentation.
Cost of municipal government	High energy cost	High tax rate, services cut	Initiate program of energy awareness, audits, capital improvements	Lack of expertise, short time perspectives, lack of funds
Welfare	Higher energy costs hurt those least able to afford them	Worse health problems because of inadequate heating	Winterization programs for the poor, emergency assistance	Insufficient funds, poorly used

lems and potential solutions, all of which might benefit from prompt attention.

MID-TERM SUPPLY STRATEGIES

The next important issue is how to develop energy sources. The difficulties of supply appear to lie primarily in oil and gas, but many of the alternatives being prepared are designed to provide electric power. Nuclear fission, nuclear fusion, geothermal energy, much coal combustion, most large-scale solar concepts, and others—all are either exclusively or mainly for electricity. Synthetic liquid fuels from coal or biomass could amount to only a small fraction of present oil and gas use (about 55 quads per year)*, in any event for several decades and possibly also in the indefinite future. The same can probably be said of synthetic fuels derived from domestic oil shale, even though the emerging *in situ* extraction processes may reduce the water consumption and general environmental damage that have hitherto been serious deterrents to exploiting this huge resource.[4] Biomass from municipal, agricultural, and forest wastes could provide at most about 10 percent of present use of oil and gas, and even then only after the year 2000.[5] So-called energy farms could supply a good deal more in principle, but at the cost of undue competition with food production and ecological risks typical of any monocultured crop.

The prospects for producing coal-based synthetic fuels are limited also. For example, water supply may perhaps be a major limitation in the production of synthetic fuels from coal, and possibly oil shale. A very pessimistic view is presented by Harte and El-Gasseir, who assert that freshwater supplies will limit production of synthetic fuels from coal in the United States to about eight quads per year.[6] Critics of this work feel that insufficient account was taken of the potential for more efficient water use, including recycling, for interbasin transport, and for tapping new sources such as brackish underground water. All would agree, however, that water availability and competition from other water uses will impose severe restrictions on the siting of plants, and will require major technological developments in efficient water use and large water-related investments.

There are further problems in using fossil fuels, especially coal, some complicated by considerations truly global in scope. (We will see that nuclear power too has its problems.) The most urgent of concerns in burning fossil fuels is the continuing buildup of carbon dioxide in the atmosphere, apparently due in large part to fossil-fuel combustion, as well as widespread deforestation, especially in the tropics. A likely result of such a buildup, if it continues, is an overall warming of the earth's atmosphere, which holds the potential for shifts in global wind currents, rainfall patterns, and temperature distributions. The outcome of such climatic shifts are poorly understood; world agriculture, whose capacity is even now strained in many ways, could be severely disrupted. This possibility is being seriously studied, and a continued, coordinated effort at understanding this effect is a necessity.

The increasing demands of developing nations for their share of dwindling oil and natural gas supplies further emphasize the need for this nation to move away from reliance on fossil fuels. The current world oil supply and price situation already forces some nations to choose between slower economic progress and rapidly growing indebtedness.

The United States uses 55 quads of petroleum and natural gas annually. This seems considerably in excess of what is producible from politically secure reserves or from coal-derived synthetics over the long term. Faced with this apparent imbalance between the need for and the supply of liquid and gaseous fuels, we can either try to shift technological development more toward synthetic liquids and gas (mainly from coal with much better technology), or to restructure the technological basis of the civilization to reduce permanently the current heavy dependence on such chemical fuels. In practice, both these approaches will be necessary, but the balance of emphasis between them is complicated, important, and ill-understood.

Summing up, the central issues in U.S. energy policy are conserving, preparing for long-term shifts in energy sources, anticipating a possibly more electric future, and planning to phase out fossil fuels. These are all difficult, and opinions differ on the extent to which each is needed or feasible. However, they remain the key issues, and policy as yet displays no very coherent or comprehensive view of their relative importance.

Oil and Gas

Recent estimates of total remaining recoverable resources* of petroleum available in the United States at a cost not higher than about twice the present world

*The term "quad" is short for "quadrillion Btu's." It is used throughout this chapter as a means of putting different energy sources on an equivalent basis. One quad is equivalent to about 186 million barrels of oil, a trillion cubic feet of gas, or 40–50 million tons of coal (depending on type).

*The quantities of resources are defined in terms of reserves and resources. In general, reserves are the amounts that can be extracted economically (that is, with current technology and at prevailing prices). Resources represent the total amounts known or estimated to be in place, without regard to the feasibility of extracting them.

price—that is, about $30 per barrel[7]—are estimated at 100–125 billion barrels. Present total domestic use is about 7 billion barrels per year, of which nearly half is imported. Although oil will not abruptly stop flowing, one may think in a general way of the recoverable resources as representing a little more than 15 years' supply, or 32 years' at the present import rate. But these figures are only approximate, because of uncertainty about both resources and projected consumption.

Oil obtained by secondary recovery—such as water or gas injection—have been already figured into these estimates; but in general tertiary recovery schemes—such as fire-flooding or use of detergents to release the oil from its rock matrix—have not, because both the cost of these methods and the outlook for substantial additions to reserves from them remain very uncertain.

Domestic oil-shale resources are estimated to contain 3,660 quads of recoverable oil (the equivalent of about 680 billion barrels); this is several times the size of the domestic oil and gas reserves combined. A technology for exploiting the resource does exist, and in principle production could begin almost immediately. But there are serious difficulties. Virtually the entire resource is concentrated in a small, arid, and ecologically fragile area on the western slope of the Rocky Mountains, and the problem of water supply, spent shale disposal, and air and water pollution are formidable. Also, the raw shale oil would need substantial refining even for use in large boilers and utility stations; among other things, it contains large concentrations of arsenic, for which the removal and disposal technology is not yet available. One should not count on large contributions from this source for many years, though a small pilot program might not be out of place.

Tar sands represent a possible major resource to Canada and Venezuela, but domestic deposits are of much poorer quality and are unlikely to figure strongly in the U.S. energy economy.

A principal difficulty with all these unconventional sources of oil (and also those of gas) is that they are very capital-intensive and expensive to produce; it is therefore unlikely that they could more than partially offset the decline in production from conventional sources.

The total world resource of oil is figured by Moody and Geiger and others as about 2,000 billion barrels, about 50–60 years' supply, depending on both patterns of use by industrial countries and growth rates of developing countries.[8] The present high and prospectively higher oil prices have damaged many plans of developing nations, and have led to undesirable increases in the use of local noncommercial energy sources—more deforestation of India and Nepal, for example. The 2,000-billion-barrel estimate presumes that a number of large oil fields will be discovered, such as in Mexico.

The United States' ultimately recoverable resources of natural gas have been variously estimated at 500 trillion cubic feet, and maybe much more, depending on whether large amounts of methane can be extracted from the geopressured brines of the Gulf Coast at reasonable prices, perhaps $5 per thousand cubic feet or less. Prudence advises against counting on much from those brines. Present annual domestic use of natural gas being about 20 trillion cubic feet, we see a resource time of some 25 years. The prospects for importing large amounts of natural gas are problematical, because liquefying and shipping the gas is expensive and hazardous. A possible exception is the prospect of importing large quantities from Mexico by pipeline. Thus finding more gas, making substitute gas, and reducing consumption must all receive urgent attention.

These findings, agreed to by most analyses, coupled with the material in earlier sections of this chapter, pose a series of policy difficulties, most of which will be compounded when we consider coal a little later. As an example of such difficulties, note that with a prospective U.S. resource base for oil and gas of 15–30 years but shorter economic time perspectives, short-term optimization urges expansion of domestic production capacity as rapidly as prudently possible. But that decision perversely tends both to expand, or at least maintain, petroleum- and gas-dependent technologies and to exacerbate the global environmental problems of fossil-fuel use. Also, petroleum- and gas-dependent technologies will be all the harder to readjust later on even shorter time scales.

What seems missing, or at least insufficiently considered, is a strategy or series of possible strategies not only to reduce dependence on imported petroleum via short- and medium-term domestic expansion, but also to greatly reduce total dependence on gas and oil over approximately the next 25 years. Recent delays in gaining public acceptance of even much more modest adjustments to national energy policy suggests such strategies will be hard to implement and will require a better public understanding.

Regarding domestic policies to stimulate petroleum conservation in the short term and domestic exploration for the medium-term transition, fuel taxes have been suggested as effective instruments and require serious study. The long-term objectives of U.S. policy for oil and gas should be to have prices more nearly reflect the replacement cost of supplies. The actual U.S. technology for petroleum and gas exploration and development is in private hands and unexcelled.

Coal

Although coal has many unattractive features, it is the only large, assured fossil energy source available to the United States that can in principle be exploited fairly quickly.

The coal reserve base in the United States, perhaps 500 billion tons (4,000–5,000 quads), exceeds the petroleum and gas reserves. The total coal resource is several times larger than this. Nevertheless, coal does not appear as attractive for the long term as these numbers might indicate. In the first place, an energy growth rate of 2 percent per year (compared to the actual 2.5 percent growth between 1920 and 1972) would exhaust the reserves in one century, assuming that coal must account for almost all the increase. In the second place, the potential climatic impacts of global carbon dioxide emissions from fossil fuels and other sources could dictate a much earlier transition from so much use of fossil fuels; this will be discussed later. In the third place, coal has problems of its own.

The supply situation is not straightforward. The price of coal may tend to rise, for example, to keep just under the oil price by whatever are the incremental costs of transportation, pollution control, and so on. Labor costs in a rapidly expanding coal industry would be an especially important source of upward pressure on prices. Indeed, the unavailability of labor at almost any price could place a ceiling on the rate at which coal can be mined. The Tennessee Valley Authority recently contracted for eastern Appalachian coal that conforms to sulfur emission and other standards at $50 per ton, far above the typical steam-coal price of $20–$30.

Besides large problems of environmental damage on local, regional, and global scales, coal presents a host of lesser ones. Here are some:

- Legislative decisions requiring that sulfur oxides be scrubbed from the stack gases from all new large coal-burning plants (irrespective of the coal's initial sulfur content), which effectively keeps low-sulfur western coal out of eastern markets.
- The enormous need for upgrading railway tracks and roadbeds, if most coal is to continue being carried by rail.
- The problem of polluted-water disposal at the output ends of proposed coal slurry pipelines, if they become a dominant mode of coal transport.

Research, development, and demonstration related to coal suffered until the early 1970's from near-starvation; the Office of Coal Research, for example, was funded at only about $10 million annually through the 1960's. Nuclear options, at the same time, were funded generously. The results now appear in urgent demands for new coal technology on a too-short time scale.

Fluidized Beds Taking the pragmatic view that coal use must be increased for several decades despite a poor long-term outlook, we should attempt to use the best technology likely to be available. One promising approach is fluidized bed combustion, in which a bed of ash plus continuously fed pulverized coal is kept levitated by streams of air blown from below. The coal burns at a relatively low temperature in the bed, thus minimizing the formation of nitrogen oxides. Furthermore, sulfur dioxide absorbers such as crushed limestone can be added to the bed, and most of the sulfur is removed with the dry ash. Some remaining problems are:

- The carry-over of particulates and hydrocarbon into the hot gas stream is still not well enough known.
- The erosion and corrosion of steam-raising pipes that pass through the bed, giving very good heat transfer from the bed to the steam, may require the use of special steels.

Compared with typical problems faced by the nuclear power industry, these seem simple. The main impediments to resolving them are a legacy of very conservative attitudes in the coal-equipment industry and coal-burning utilities.

The other main desulfurization scheme, and the one most used now, is scrubbing the exhaust gases; this, unlike fluidized bed combustion, does not remove nitrogen oxides or some other pollutants. The use of regenerable scrubbing chemicals, as in the so-called double-alkali system, may be the best approach; it tends to minimize the large amount of polluted limestone slurry that comes from the once-through systems now used.[9]

Coal Conversion

The art of making clean synthetic fuels from coal progresses, but slowly. Fuels of this kind are envisioned both as substitutes for both petroleum and natural gas in applications that require liquids or gases and as cleaner burning substitutes for coal. They are not expected to contribute very substantially before the 1990's, and apparently will never be able to replace much more than half of even our current imports of oil and gas, largely because of the perceived limitations of water supply. Furthermore, there is some question about the net benefits to pollution abatement in using such fuels to generate electricity; this might only shift certain environmental impacts from the power plant to the synthetic fuel plant. As supplements to the supply of oil and gas, however, they may be important.

Synthetic Gas The main step in converting coal to gas efficiently consists of adding hydrogen to the coal. This is usually done by partial combustion in the presence of a hydrogen-rich material (usually water). The result is dissociation of the complex molecules in coal into simpler and lighter substances consisting mainly or solely of carbon and hydrogen, plus a variety of residues. In simple gasification processes coal reacts in air with steam to produce carbon monoxide and hydrogen, with heat for the process provided by burning of some of the coal. This produces so-called synthesis gas, which has a heat content about 10–15 percent of that of natural gas. This low-Btu gas is suitable for onsite use, but because of its low heat content is not economical to transport very far. If oxygen and hydrogen are used in the gasification process, a somewhat higher-quality gas can be produced, having about 30–35 percent of natural gas' heat content; such intermediate-Btu gas must also be used onsite, but may have advantages over low-Btu gas in some applications. Finally, synthetic gas, fully equivalent to natural gas and therefore useful in existing pipelines, can be manufactured by adjusting the hydrogen-to-carbon monoxide ratio of the raw gas to about 3 : 1 and passing the mixture over a catalyst at high temperature to yield almost pure methane.

Promising early technological developments in using gas from coal, ranking almost equally with fluidized bed combustors in timing and importance, are some combined-cycle systems using low-Btu gas from coal prepared and burned onsite to generate electricity. Technologies for producing low-Btu gas have been in hand for many years; present environmental and economic conditions require some technological improvements deemed quite feasible. Cost estimates are high, running to \$3–\$4 per million Btu, corresponding to \$17–\$23 per barrel of oil; but other gaseous fuels may be in such critical supply as to be disallowed for electricity generation in the late 1980's. The overall thermal efficiency of a combined-cycle system can exceed 50 percent from input gas to electricity, and 40 percent or more on the basis of raw coal. These efficiencies are satisfactory for a clean system.

Synthetic Liquid Fuels Many coal liquefaction approaches have been evaluated. The three that are favored at present—Gulf's Solvent Refined Coal II (SRC-II), Hydrocarbon Research's H-Coal, and the Exxon Donor Solvent (EDS) process—are hydrogenation processes that yield about 2.5-3.0 barrels of liquid per ton of coal. They all use approximately the same process conditions and enjoy roughly equivalent technical status. The SRC-II process is most advanced in terms of scale, with a Department of Energy-sponsored facility in Tacoma, Washington, operating at about 30 tons of coal per day and a 6,000-ton-per-day plant planned for operation in 1985. Plants of similar or larger size have been proposed for the other two processes.

If it were deemed necessary to begin deploying a liquefaction industry today, the best available technology would be the Fischer-Tropsch process, used in Germany during the Second World War and still in use in the Union of South Africa. This is a two-stage process; low-Btu synthesis gas is produced from coal, and then converted to a variety of liquid fuels. Compared to the more advanced systems under development, the process is low in efficiency and high in cost.

Present projections are for synthetic oil prices to be \$20–\$25 per barrel of fluid that contains many more biologically active (and therefore possibly toxic or environmentally hazardous) species, such as anthracenes and phenols, than does natural crude oil. This material can be chemically refined to open up the molecules into long chains similar to those in natural petroleum. But the cost will be higher, the yield lower because hydrogen must be added, and the plant necessarily more leak-tight than is the custom in present oil refineries. In addition, the chemical and physical stability of these fuels against forming tars and gums during storage is a continuing problem, since they can polymerize more readily.

In summary, while coal is almost sure to be used, it remains in a difficult technological condition, with very substantial environmental, technological, and social questions outstanding.

Nuclear Fission

Nuclear fission power can be viewed as both a mid-term and a long-term energy alternative. It is mid-term because it is available now in appreciable amounts, but this energy option will last only until sometime in the next century, given present technology using naturally fissionable uranium. It can also be long-term, because with breeder reactors the uranium (or thorium) resource would suffice for millions of years. Here we discuss fission as a mid-term option, chiefly to compare it with coal. The long-term aspects are discussed later in this essay.

Nuclear power in the United States faces an uncertain future. Fostered and promoted by the government for three decades, the industry and the technology have lately suffered a number of setbacks. These are reflected in the last five years' precipitous decline in utility orders for reactors, from 34 in 1973 to only 2 in 1978, and there are no signs of improvement in this situation. Thus, although most projections suggest that life in the next few decades would be very difficult without nuclear fission, even if the correspondingly increased fossil-fuel use proves acceptable, nuclear power may be barred from contributing significantly.

The reasons for the industry's pessimistic outlook are complicated. In general, utilities are inhibited from ordering reactors by several considerations. First by a wide margin is uncertainty about the future regulatory and political environment of nuclear power. The technology has lately become subject to public criticism on a number of grounds; organized antinuclear activist groups have demanded and received easier access to participation in the licensing and siting processes, and the consequent delays for litigation have raised the risks of financial losses and capacity shortfalls. In addition, some states and localities have placed legislative limits—some amounting to virtual bans—on various nuclear power activities. At the moment, this kind of opposition centers around the question of reactor safety—especially, as this report is written, in the aftermath of the Three-Mile Island reactor accident near Harrisburg, Pennsylvania, March 28, 1979—and waste disposal; opposition often also centers around nuclear fission's high technology per se and its supposed social attributes. The overall result is a lack of a reasonably sure and uniform siting and licensing policy for nuclear plants and related facilities, and periodic requirements for retrofitted safety devices on existing reactors. The utility industry is as averse to this institutional uncertainty as it is to technical unreliability; thus the present *de facto* moratorium on new nuclear plant orders in the United States.

Also, the breeder reactor development program, on which utilities have counted as a follow-on technology and as a source of plutonium for fuel when the nation's limited uranium resources are depleted, has been delayed. Recycling of spent reactor fuel has also been deferred. Both actions were taken in part because of the Administration's concern about weapons proliferation.

Another factor sometimes mentioned as a vital consideration is the fact that projections of electricity demand are significantly lower now than they were five or ten years ago. However, the rate of new orders is at present inadequate to satisfy even these lower projections.

Underlying the nuclear industry's troubles are serious social concerns. As a technology for producing electricity, nuclear fission is at least as reliable and economical as coal, its closest counterpart in the production of baseload power.[16] But concern with its safety is persistent, and the issues of waste disposal, reactor safety, and weapons proliferation must be convincingly and publicly addressed if the technology is to have a productive future.[17]

Waste Disposal No feature of nuclear power contributes more uncertainty to its future than the problem of permanent disposal of nuclear wastes. While many promising technical options exist for geological storage of solidified wastes or spent fuel, the federal government until very recently has neglected their demonstration in a systems engineering sense. As in the case of catastrophic reactor accidents with many casualties, the possible scenarios of what could go wrong are hypothetical and speculative. The hazards cannot be tested by experience, and we are forced to rely on rather theoretical calculations, with only some of the individual steps in a long sequence testable by experiment. There is always room for the criticism that the theory may have overlooked something. In the case of reactor safety the accumulation of operating experience free of accidents with severe public consequences could eventually justify theoretical optimism. In the case of waste disposal, however, there is no wholly experiential way of validating the calculations, because the safe storage period extends over many generations. The main mitigating feature on the other side is that the consequences of a storage malfunction are much less severe, even under extreme assumptions, than the consequences of reactor accidents, hydroelectric dam failures, or breakdowns in many other present and proposed industrial waste systems.

Reactor Safety The hazards of accidents with nuclear power reactors are also important sources of public concern. The most severe accidents conceivable—for example, a loss of coolant resulting in a core meltdown followed by a breach of the reactor's containment—could be catastrophic indeed, with possibly thousands of deaths and property damage measured in billions of dollars, assuming the most unfavorable possible conditions of weather and population density. The calculated chances of such a catastrophic accident are exceedingly small—of the order of one in a billion per reactor-year. This assessment of the risk, however, is based not on actual experience, but on calculated probabilities of hypothetical events involving human fallibility. The uncertainties in such calculations are very large, and although the overall actuarial risk is almost certainly small compared to that of almost any other energy technology, the uncertainties surrounding the probabilities of the largest accidents is a serious deterrent to public confidence. The reality of these uncertainties and of the depth of public concern was illustrated in the Three-Mile incident, mentioned earlier. To be sure, a sequence of events rather similar to what actually happened has been studied, but not for that specific make of reactor, and not accounting for some of the human errors that compounded it—nor for the later human ingenuity that limited it. This uncertainty should be reduced rapidly in the next five years, as analysis of operating experience continues and the theoretical study of accidents is refined.

The question of civilian nuclear power's connection with the spread of nuclear weapons is equally difficult. This is not really an issue affecting light water reactors since they now use a once-through fuel cycle. Reprocessing of spent fuel for its fissile content, however, would, with the processes now available, separate plutonium for a short time in a form suitable for weapons manufacture. Some fear that a nation might be tempted to use such material in a clandestine effort to acquire nuclear weapons, or that a determined and well-organized terrorist group might be able to steal such material for its own purposes. Since the breeder, and thus the long-term prospects of nuclear fission, depend on fuel reprocessing, the resolution of this question will strongly determine the technology's future. Reprocessing facilities naturally must be extremely secure to minimize this risk. But, overstating the risk is unproductive; nuclear power systems are clumsy and not very private ways to acquire nuclear weapons, and it is not clear how this particular problem contributes to the large and complex issue of international security. It is important to recognize, for example, that civilian nuclear power, as an economical source of energy, may in contributing to economic development actually lessen international tensions.

The public debate about nuclear power, and the concomitant irresolution of goals and policies, may continue until we settle the question of whether enough coal can or should be mined to produce most baseload electric power. Much more definite answers should be available by 1985, but by that time the nuclear supply industry is likely to have attenuated drastically. Recovering it could then be very costly.

Nuclear power must be considered in the context of the other national and international energy concerns. It is a source of relatively cheap and environmentally clean energy, qualities hard to find in other available energy sources.

SOME OUTSTANDING ENVIRONMENTAL AND HEALTH EFFECTS

As suggested earlier, the industrial and resource problems of energy supply, while a cause for concern, do not fully express the difficulties this nation will face in providing energy over the next few decades. Energy production and use in the past have had unforeseen impacts on the human environment. Many of these remain poorly understood in detail, but some grosser cases of energy-related ecological damage, public health problems, and of a range of impacts on agriculture and industry are recognized; and regulatory remedies have been imposed in the forms of emission standards, mine reclamation requirements, and the like.

It is not yet possible to rank the various available energy sources precisely according to the relative severity of their impacts, and it may not be especially desirable to do so. Different energy supplies are used in different ways to meet different needs, so that their relative benefits—which must be set against their impacts—are not strictly comparable. However, there are grave reasons for special concern about the impacts of certain energy sources, given current energy projections. Especially important are the hazards of fossil-fuel combustion. Local and regional hazards, for example, arise from emissions of nitrogen and sulfur oxides and particulates into the air. By far the most worrisome global hazard is the prospective buildup of CO_2 in the atmosphere, in part as a result of fossil-fuel combustion.

Coal is by far the most destructive fuel in ecological and public health terms, and the likelihood of large increases in its use presents some unpleasant prospects; we know that even now we pay a rather substantial environmental price for its use, including some air-pollution related deaths and crop damage. We pay a large price also in deaths and injuries of miners, in water pollution from mine drainage, and in disruption of mined land and underground hydrological structures. All of these problems have been dealt with, with various degrees of success, by regulation; the cost and the difficulty of this regulation will grow along with the role of coal in the national economy.

The CO_2 problem is so important as to require special mention. This is an unavoidable consequence of fossil fuel use, and there is no realistic means of control other than limiting the use of these fuels.[10] Again, coal is the worst offender per unit of energy, though oil and natural gas, with about 80 and 50 percent, respectively, of coal's CO_2 emissions, also make important contributions.

Atmospheric CO_2 helps to regulate the temperature profile of the earth's atmosphere and surface. This happens because, although the atmosphere is transparent to the wavelengths of incoming solar radiation, the CO_2 and water vapor in the atmosphere absorb the longwave infrared radiation (heat) reradiated from the earth's surface. They thus trap heat and raise surface temperatures; this is usually referred to as the greenhouse effect. Woodwell calculates that today burning fossil fuels and oxidizing biomass, mainly by destroying tropical forests and the associated humus, each contributes approximately comparable amounts to the global atmospheric CO_2 increase. The amount will double within two generations, if present trends continue; the result could be a significant climatic readjustment, perhaps a general warming of a few degrees Celsius, with several times as much near the poles.[11]

Government energy planning throughout the world virtually ignores this problem. But it presses now, because the fundamental changes that would constitute a remedy would take a long time. For example, global agriculture, by complicated geographic, social, and institutional arrangements, matches crops to particular areas. Experience coupled with simple analyses shows that total production decreases in times of changing climate, because neither the pattern of land use nor the fertility can change rapidly enough to accommodate. The system has inertia. If rainfall shifted from Iowa to Arizona, for example, the corn would not; it would stop growing in Iowa, but it would take many decades at least for suitable soil to develop in Arizona.

If these projections are valid, hard choices lie ahead, such as substantially reducing the combustion of fossil fuels worldwide, in the face of growing demand for them, especially by developing countries. The options of solar power (probably without using biomass, so as to maintain photosynthetic carbon uptake at the highest practicable level), energy conservation (especially in industrialized countries), and nuclear power thus take on special significance. But the first and most difficult task may be to make people all over the world aware of the problem. No very effective institutional mechanisms appear available to deal with such global matters.

Most air-quality degradation arises from conversion or combustion of fossil fuels, but we know little about the consequences. The effects are most apparent in and near cities, where much energy is used in relatively little space. Thus, patterns of energy use profoundly affect urban living; one uncertain consequence is that of urban air quality degradation on health.[12] Evidence abounds that persons with pulmonary and cardiac dysfunctions suffer notably, and die more readily, when the air is bad. But it is difficult to quantify the role of air pollution in deaths among this population.

One should also consider why people got sick in the first place. That is a very difficult question, involving economic status, social and dietary habits, and many other things. The combination of epidemiological, clinical, and basic etiological studies has so far been insufficient to provide very good answers. Compared with studies of the health hazards connected with nuclear power, the health hazards related to fossil-fuel use are very poorly known.

The emissions from fossil-fuel burning include sulfur and nitrogen oxides, among other things, and these interact with moisture and other atmospheric constituents and contaminants to yield smog and acid rain. Thus, these emissions are increasingly strictly controlled as the years pass. But the bases for setting air-quality standards are themselves inadequate, and even the list of controlled effluents has notable omissions. It is worth noting that, because of fossil-fuel emissions of sulfur and nitrogen oxides, in about one-sixth of the United States—defined approximately by a western boundary just west of the Mississippi River and a southern boundary at the latitude of northern Alabama—the rainfall is distinctly acid, with a pH 5.0 or less, and that, in a region about one-tenth that size in the northeastern United States, the acidity is severely high, a pH about 4.0 or less. These regions are considerably larger than they were 20 years ago. The consequent damage to forests and sports fishing in northeastern lakes has been estimated, probably conservatively, at \$100 million annually, and corrosion of the surfaces of buildings and other structures at between \$100 million and \$500 million per year.[13] These are high costs to pay and expensive phenomena not to understand.

Yet another general environmental difficulty, related especially to energy, is the question of water supply. The topic has been often mentioned, but its gravity is not well appreciated. This matter was discussed earlier in this essay.

LONG-TERM ENERGY OPTIONS

Long-term prospects for using large amounts of oil and gas are not bright. We have seen that the supplies of petroleum and natural gas are limited, both nationally and globally, especially at costs anything like present ones. We have also seen that the prospects for making large amounts of such fuels from coal, oil shale, or biomass are not good. Finally, the CO_2-driven climatic problem appears to depend on the long-term rate of using all fossil fuels (or destroying biomass without replenishment), irrespective of other considerations.

Thus we turn to inquire what energy options *are* available in the long term—options upon which it is possible at least in principle to construct civilizations sustainable for the long term. Only two major classes of options exist: solar power in its various forms, and nuclear power, both fission and fusion. These can be, and are now, augmented in some places by relatively modest amounts of geothermal power, but the normally small heat flow from inside the earth makes solar power much more promising, except for some special locations. Tidal power is very limited; a dam built one kilometer offshore around the entire United States would, with 100-percent-efficient low-head turbines, hardly generate enough electric power to meet present demand in Massachusetts, and would do so at environmental and capital costs likely to cause much complaining in other states. Much more power is available from waves, though, and the possibility of extracting it is being seriously studied in the United Kingdom, where suitable conditions exist—a small land area with concentrated population and industry, surrounded by stormy, shallow seas.

The long time it will take to develop and install solar or advanced nuclear power technologies, comparable to the time in which the classic fossil fuels must probably be phased out, provide a sense of urgency. A lively debate exists about whether the relative research and development priorities for the prospective long-term energy sources reflect proper social purposes. In this area, as in many other parts of the energy scene, we find people advocating options on social as much as on technical grounds. As we have seen, however, the larger danger may lie in having too few rather than too many good energy options.

In what follows, it is important to note that many of the most promising long-term options produce electricity as their most natural product. This fact bears strongly on the small-versus-large, diffuse-versus-concentrated debates now fashionable. All the nuclear systems are large; and while many of the solar technologies can be modular and small, the questions of interconnections and of backup energy sources bring us back to the necessity of accepting substantial centralization. In all but the most primitivistic views of future society we must accept substantial centralization of energy supply and delivery. Electric utility systems, for example, will remain with us.

Nuclear Fusion

To understand the development of controlled fusion, the hugeness of the task and the attitudes of those associated with the effort have to be appreciated. Perhaps it is merciful that the scientific difficulties now actually being solved were not recognized from the start. But now it seems quite clear that an acceptable fusion plasma—a gas of deuterium and tritium ions 10 times hotter than the core of the sun—can be produced and confined in a vacuum by strong magnetic fields, can be kept from touching any physical surface, and can be confined long enough for an interestingly large fraction of it to react—producing mainly helium, neutrons, and energy. Thus, in a scientific sense, controlled fusion appears possible. It has taken about 25 years and several billion dollars of effort in several countries, virtually without international hindrance of any kind.

Fusion Reactors But the most expensive and difficult parts—the technology and the practical engineering of a working reactor producing power—are still to come: huge superconducting magnets, vacuum wall materials exceptionally resistant to radiation damage, and so forth.

The basic technological feasibility of controlled fusion will probably not become really clear for at least another decade, and the engineering and economic feasibility even later. Fusion will come, if at all, later than breeder reactors and advanced solar power systems, both of which may be called for by about the end of this century. Still, fusion's most pressing uncertainties may be resolved by then, so it remains a contender for the long term.

Fusion reactors will probably come in large sizes, perhaps 1,000 megawatts electric, or about the size of large modern coal and nuclear power plants. Their high technology suggests that cost will also be high by today's standards. However, the scarcity of acceptable, well-developed long-term energy options gives the word "high" considerable flexibility.

Fusion Hazards Controlled fusion produces radioactive materials and other hazards. But the wastes consist not of highly radioactive fission products, but of tritium and of the activated reactor itself. All these hazards, taken together, seem much less than those of fission. Thus in an environmental and social sense, fusion would be intermediate between solar and nuclear energy.

The connection between controlled fusion power and nuclear weapons is much less relevant than in the case of fission. To be sure, a flood of neutrons emerges from any fusion reactor, and these could be used to produce plutonium for weapons, by surrounding the reactor with a uranium blanket. However, any nation that has mastered the sophisticated and expensive practical technology of fusion will have available many easier routes to fission (and fusion) weapons.

Alternatives At present, the most favored configuration to confine the plasma is the tokamak, a doughnutlike structure with very complicated external electric windings, conceived and first developed in the Soviet Union. More is known about this confinement system than about any other type, but it is still unclear whether a practical fusion reactor will be a tokamak or some other device yet to be developed. It would be best to develop other and possibly simpler toroidal confinement schemes, at least to the stage of determining their real feasibility vis-à-vis tokamaks.

Several nontoroidal experimental plasma confinement schemes have been proposed as prime candidates. The so-called magnetic mirror schemes suffer from particle losses at the ends, giving poor energy efficiency. The inertial confinement schemes, using either high-powered pulsed lasers or ion beams, must be severely pulsed through many millions of cycles, which introduces the difficulty of early fatigue of structural materials, mainly because of thermal stress.

Fusion has been developed to the point at which it requires large and expensive experiments; while unnecessary duplication of effort should be avoided, such experiments will mean expenditures at current and higher levels. As with the other major options, there is no cheap way out.

The effectiveness of R&D in this field has been fostered by almost completely open international research, and at present all parties are eager to cooperate more closely. An internationally planned and coordinated program would avoid costly duplication, speed the dissemination of results, and provide a basis for true collaboration on very large and very expensive engineering trials. Such an international activity is now under way (as of March 1979), chiefly under the aegis of the International Atomic Energy Agency.

Solar Energy

Solar radiation is a diffuse source of energy and large amounts of materials are needed for collectors, storage devices, and so on. To build equipment that can capture and convert solar energy to useful forms requires capital investments embodying nonrenewable resources that are far from free, even though sunlight itself is free. In a similar sense, the uranium to fuel breeder reactors is practically free, because the fuel costs are insignificant compared to the capital costs and the resources that the reactors represent. The real attractiveness of solar power, besides its ubiquity, is the relative ease with which it can be transformed for a number of uses. However, this attraction has often been oversold by various high-technology schemes. A more realistic view is imperative.

Solar power, by convention, includes not only direct conversion of sunlight into useful forms but also hydroelectric power, winds, and biomass (organic matter). Discussion of this alternative thus tends to be extensive, since applications are diverse, and options therefore are hard to compare.

Low Technology The most immediately promising solar application is the production of low- and intermediate-temperature heat, from about 70°C for domestic hot water to about 200°C for a variety of commercial, agricultural, and industrial purposes. The simplest systems use flat-plate collectors like those on the increasingly familiar rooftop water heaters. More advanced ones use mirrors or simple lenses to concentrate the solar heat and provide higher temperatures. Most require little or no further science or advanced engineering; they will succeed if their design and construction are ingenious and simple enough to make them economically attractive. Their principal competition now is fossil fuels, mainly oil and gas. Another decade of rising oil and gas prices plus improvements in commercial solar systems should be sufficient for this technology to develop a strong commercial position, assuming that the systems' reliability turns out to be adequate. The federal government has acted to provide some economic stimuli to these low- and intermediate-technology items.

High Technology In the past the solar energy program, partly through a history of influence by NASA and other high-technology agencies, concentrated on tasks with formidable (but interesting) science and engineering problems. At this extreme, we find advanced photovoltaic conversion schemes, the power tower concept (in which a vast array of steerable mirrors focuses sunlight onto a boiler atop a several-hundred-meter tower), and ocean thermal electric conversion systems, which exploit the temperature difference between the surface and the deeps of tropical ocean water.

One of the most promising technologies is photovoltaic electricity generation, which is technically feasible in a variety of installation sizes ranging from the individual household to large central station generators (Figure 20). Photovoltaic generators are technically feasible today and have in fact been marketed for specialized applications, such as power sources for satellites or for electrical equipment in remote locations, where the high cost is justified. Wider applications must await cost reductions of roughly an order of magnitude, depending on what happens to the cost of other alternatives in the meanwhile. In one scheme, single crystals of silicon or gallium arsenide would be produced at much lower cost than today. It may pay to concentrate the sunlight with simple optical systems, since this would permit a higher cost per unit area for the cells. In many cases the cost of the optical system and supporting equipment may dominate the cost of the cells themselves, so that both need equal engineering attention.

The other photovoltaic approach aims at the scientific understanding, development, and eventual use of amorphous, or noncrystalline, photovoltaic materials, especially as thin films that can be evaporated cheaply

FIGURE 20 Plastic fresnel lenses (right) focus the sun's rays on silicon solar cells (bright spots on left) to convert sunlight directly into electricity. (Sandia Laboratory photo, Department of Energy)

onto inexpensive backing material such as mylar film. This is still in the research stage, because the science of noncrystalline materials is generally more difficult than that of pure crystals. Some success has been achieved, but experimental conversion efficiencies remain smaller than those of crystalline photovoltaic material. Technical progress in both single-crystal and polycrystalline or amorphous cell materials has been very rapid in the last 5 years, based on the large background of knowledge derived from materials and solid-state research over the preceding 15 years. There is now great optimism about the long-term future of photovoltaics, but according to a recent study[14] their market penetration is unlikely to exceed 1 percent by 2000. Even if the future economics were much more firmly established, the problems of creating a large industry and supporting institutional structure and integrating it into the existing electric grid would probably preclude much more rapid market penetration in the present century. Early deployment will be primarily for peaking capacity, and will require conventional backup. Larger market penetration in the next century will increasingly depend on the availability of economical energy storage systems.

The power tower does not look promising. A planned 10-megawatt pilot plant in Barstow, California, will cost $120 million. It will be hard to replicate much more cheaply, because its costs depend so much on those of concrete, steelwork, mirrors, and so on—all of which represent well-developed technologies for which significant cost reductions cannot be expected. A particularly difficult engineering problem is the boiler atop the tower; it must withstand large and rapid thermal fluctuations as clouds pass over the field of mirrors.

Both the ocean thermal energy conversion scheme and, even more, the solar-power satellite scheme are unlikely energy options; they are very capital-intensive and full of serious and poorly understood scientific and technical problems.

Windpower devices occupy an intermediate status, neither high technology nor low, and like most other solar technologies should be regarded as augmenting conventional power supply. The official line of development, which tends toward devices with vanes as large as the largest airplane wings, may not be the best approach. Much cheaper and more reliable devices can serve local areas; solid-state electric circuitry can match the electric output of any reasonable-size windmill to the frequency and voltage of power lines.

Hydropower is usually imagined as the environmentally and ecologically ideal way to produce electric power. But quite apart from the oft-stated risks of dam failures, the dams themselves chop up river systems into ecological bits and unnatural parts, with generally unfavorable consequences too large for the amount of power actually produced. This may be true also of low-head or small-scale hydroelectric installations. Exacerbating the problem of making good judgments on this matter, a recent Corps of Engineers study of power available from low-head hydro dams overestimates by a factor of several (perhaps 10) the average amount available, at least for the New England area.[15]

Regarding biomass, probably 5–10 quads of energy might be producible from farm, forest, and domestic wastes, much of it at a considerable gain in environmental quality; for example, removing the energy from urban wastes and animal feedlot wastes. The concept of obtaining much more energy from intensive silviculture, such as growing sycamore trees for five- to eight-year harvest ("energy farms") looks unattractive. First, we find here the typical ecological problems of monoculture crops. Second, there is the potential for direct competition with production of food, lumber, and pulpwood. Third, restructuring the nation's forestlands for this purpose would take decades. Cutting down scrub forests has been suggested; but the energy content of all the trees in New England would provide that region's energy needs for only about three years.

Concluding with the notable exception of heating and cooling of buildings, note once more how many of the options naturally produce electricity; this reinforces the opinions stated earlier that we seem to be heading for a much more electric economy.

Many solar energy schemes would benefit from having associated energy-storage systems for when the sun does not shine, and virtually all require backup sources of energy for when the storage capacity is exhausted. There are many ways of providing storage, and the need for it varies from application to application and technology to technology. The need for storage has been remarked upon, especially in connection with solar power generation. Without wishing to de-emphasize the importance of energy storage, we point out that until the solar-derived fraction of electric power exceeds about 10 percent of the peak demand, an extended electric grid could absorb the fluctuations in solar output by adjusting the output of its conventional generators. A solar-electric system can be begun with its eventual storage system not yet in place. In the long term the best form of storage in connection with solar energy is the production of fluid or gaseous fuels that can serve as substitutes for hydrocarbons. A wide range of possibilities exists, of which the most important relates to producing hydrogen, either by electrolysis of water or by various photochemical processes. No such process has advanced to the point at which it can be seriously considered for engineering development. This is an important and hitherto neglected area for basic research and exploratory development.

Long-Term Nuclear Fission Technology

The light water reactors (LWR) that now provide almost all of the nation's nuclear generating capacity, operated as they are without reprocessing and recycling of fuel, are not very fuel-efficient. In fact, one recent estimate of the nation's uranium resources suggests that in terms of today's typical 1,000-megawatt power plants fuel supplies will limit LWR capacity to only about 300 plants with relatively assured lifetime supplies of fuel.[18] Other estimates are higher but have less supporting evidence. Fuel supplies for LWR technology can be considered comparable to oil supplies. If nuclear fission is to contribute substantial amounts of electricity at reasonable prices much beyond the end of this century, more efficient ways of using nuclear fuel will be needed.

One way of using uranium ore more efficiently would be to extract the directly fissionable isotope uranium-235 from the nonfissionable part more completely. New technologies—in particular laser isotope separation—have the potential for doing this, but the outcome is not yet certain. In addition, redesigning the LWR reactor cores could improve fuel utilization by perhaps 25 percent.

In fact, development of the LWR depended almost from the beginning on the assumption that spent fuel would eventually be reprocessed and recycled in reactors. This alone would raise the usable amount of the energy potentially available from a given quantity of fuel from about 0.6 percent to nearly 1 percent, thus extending somewhat the life of the available uranium. However, the wisdom of reprocessing has become a matter of public controversy.

Breeders Breeder reactors, which can convert uranium-238 (the common, nonfissile isotope of uranium) to fissile plutonium-239 in quantities larger than are consumed in the reactors, can eventually take advantage of more than 70 percent of the energy potential of uranium ore. There are also thermal breeders, which can convert nonfissile thorium-232 (an element more plentiful than uranium) to fissile uranium-233, and in this way use nearly 70 percent of the energy potentially available from the thorium. This ability to free the energy potential in the so-called fertile isotopes uranium-238 and thorium-232 has a tremendous multiplying effect on available nuclear resources. This is much greater than the approximate factor of 100 implied by the relative fuel efficiencies just cited, because the use of breeder reactors greatly reduces the impact of resource prices on the cost of electricity. This, by making available ores too low in grade to be usable in LWR's, multiplies economically useful reserves of fuel resources.

Until recently, the U.S. breeder development program has concentrated on the liquid metal fast breeder reactor (LMFBR) and its uranium–plutonium fuel cycle. Other nations with breeder programs—notably France and the Soviet Union—have generally done likewise. Thus, at present, this breeder reactor concept is much more fully developed worldwide than any other. LMFBR's could operate for millions of years on the known reserves of uranium in the United States. Even the concentrations of uranium in the stored tailings from today's military and civilian uranium enrichment plants could support a substantial growth in LMFBR capacity for upwards of 200 years, virtually eliminating the environmental and occupational-health impacts of uranium mining and milling.

Unfortunately, breeder reactors require reprocessing and recycling of fuel to achieve their resource-conserving potential and this raises again the hazard of proliferation or theft of nuclear material usable in weapons. Largely for this reason the current Administration has decided to defer reprocessing and delay the planned demonstration of breeder reactor technology while it evaluates, on nonproliferation and other grounds, a range of alternative breeder fuel cycles and reactor concepts.

Advanced converters, which produce fewer fissile atoms than they consume but are more efficient in this respect than LWR's, could significantly extend nuclear fuel resources without requiring reprocessing. The Canadian CANDU heavy water reactor, for example, with alterations in its fuel, could exploit about 1 percent of the energy potentially available in its uranium fuel, compared to 0.6 percent in LWR's, operating on a once-through fuel cycle. This could be further extended by improvements in isotope separation. This improvement, however, would not be very significant in keeping the nuclear option alive unless the growth in electricity demand levels off rapidly after the end of this century or uranium resources turn out to be much larger than now anticipated. Under any other conditions, it will be necessary to have advanced reactors and fuel cycles that exploit the advantages of fuel reprocessing and breeding; time is already growing short for the development of such reactors, even if we consider them as insurance against continued growth in demand for electricity or limited fuel supplies. Advanced converters could also operate on the thorium fuel cycle; although this would require reprocessing in some form, yielding uranium-233 in separated form, which like plutonium-239 is a potential nuclear weapons material.

Despite the likelihood that the LMFBR would be more useful in the widest range of electricity growth and uranium availability and should therefore be developed in any event, we can envisage two possible roles for advanced converters. First, they might serve as a hedge against unforeseen difficulties in breeder development, even if they do offer only partial protection against electricity demand continuing to grow at an appreciable rate or uranium resources of suitable grade turning out to be as limited as they now appear. Second, and more probable, would be use of advanced converters as complements to LMFBR's; for example, if for proliferation reasons LMFBR's, along with fuel recycling and refabrication, were restricted to a small number of secure international sites. LMFBR's could then be used to breed uranium-233 fuel for converter reactors by irradiating thorium blankets. A given amount of LMFBR capacity could support many more advanced converters than light water reactors in this kind of arrangement.

If nuclear power's contribution is to continue much beyond the year 2000, one or more of the following possibilities must be realized by the century's end: major new uranium finds, deployment of advanced converters and later of associated thorium fuel cycles, or development and public acceptance of breeder reactors and associated reprocessing and fabrication facilities with adequate controls on proliferation. The breeder option, in the form of the LMFBR, must be maintained so that if needed it can be deployed early in the twenty-first century.

Even though the present nuclear sector is in trouble, any preparations for the future must recognize that a substantial industry based on light water reactors is in place, and it is unlikely to switch to different reactor systems without credible government assurances and assistance. Several activities sponsored by the federal government make good sense:

- Prompt demonstration of an acceptable waste-treatment method to establish public credibility. Technological options appear to be available; an example is the Swedish proposal to turn the wastes into ceramic, to encapsulate them in titanate–copper jackets (to minimize leaching from water intrusion), and then to entomb them in stable granite rock, with bentonite packing filler providing an ion-exchange medium to trap any leached ionic waste. Other technological options such as some salt deposits are possible.
- Cooperation with Canada in developing an advanced version of their CANDU reactors, which would have the ability to deliver about twice as much electric energy per ton of uranium ore, even without recycling spent fuel—a decided advantage at this time of uncertain uranium resources.
- Further development of the high-temperature gas-cooled reactor preferably in cooperation with the Federal Republic of Germany; a modest prototype exists in Colorado. The principal advantage again is better use of uranium resources. Also, this reactor may ultimately be used with a closed-cycle gas turbine and dry cooling, so as not to require large quantities of cooling water.
- Support promising and more proliferation-resistant technologies for reprocessing spent fuel, such as the Civex process.
- Maintain a breeder R&D program. Success with the second, third, and fourth activities will put off the time when decisions need be made about deploying breeder reactors and give more opportunity for the technology of large-scale solar power to prove or not prove itself. But as seen at present, there exists and will continue to exist a need for breeder development. While the liquid metal fast breeder reactor is closest to final development, enough good ideas for alternative systems exist, some using mainly thorium instead of uranium, that several lines should be followed, and no exclusive choice should now be made.

SOME RESEARCH NEEDS

Research and exploratory development (R&D) are vital to future energy technology. This account has already included some specific areas—for example, the disposal of radioactive wastes—where attention is needed. Now we further illustrate some of the R&D that might be done, with examples in large part derived from material on earlier pages. It is impossible in the restricted compass of this essay to say anything about relative priorities, and no conclusions on such matters should be drawn from what follows.

Our first major topic was time perspectives. We found them generally to be too short for the timely emplacement of new energy technologies. This may also hold for energy R&D;[19] demonstrations and engineering of mid-term technology may be overemphasized to the neglect of more fundamental research that might optimize technology to fit reasonably both the near and long terms. Both the private and public sectors have large stakes in this process, and the balance of activity requires continual study and readjustment.

THE RELATION BETWEEN ENERGY AND ECONOMIC GROWTH

Both short- and long-term research issues arise in conservation, energy demand, and economic growth. In the short term, energy and the economy seem to be closely coupled, but over the long term we found reason to expect more flexibility. However, the relations between economic consumption and energy growth are not well understood. The key assumptions of current models should be examined so that the economic feedback of reducing energy demand by various plausible methods can be more accurately predicted. Also vital is knowledge of the distributional effects of energy pricing policies and regulations intended to reduce demands; changes in such policies will affect different regions and different economic classes in different ways, and compensatory actions must be considered. These and many other social and economic aspects of energy conservation offer a broad scope for basic work in economics and other social sciences.

ENERGY CONSERVATION

Conservation is a fertile field for innovative research. Of most immediate interest are measures that could be taken now and in the near future, at costs quickly paid back in energy savings. There are undoubtedly many applications for new and sophisticated control technologies for optimizing energy use in all sectors. Industries in the United States, for instance, have made important gains in energy efficiency in response to the recent substantial increases in oil prices. In the building sector, more knowledge and better technology would allow the setting of realistic and effective thermal performance standards; no solid scientific basis exists at present for predicting the thermal performance of building subsystems, let alone entire buildings. Because of the expected continuing upward trend in air travel, it would be important to understand the financial and institutional barriers to introducing the next generation of more energy-efficient jet aircraft engines. Comprehensive energy conservation experiments in representative industrial establishments, perhaps with the costs shared by the government and the companies involved, should be seriously considered.

In principle, there is considerable energy saving potential in the introduction of industrial cogeneration—the simultaneous production of electricity and process heat. This results in the more efficient use of the primary energy resources and reduces the need for centralized generating capacity. The barriers to more widespread cogeneration are mainly institutional and economic. Experiments with various institutional arrangments for combined production and use of electricity and heat might well be encouraged and supported by government.

COMBUSTION

Combustion is a basic process of our civilization. If more were known about precisely what happened when fuel is burned, we could build more efficient engines, power plants, and space heating systems; reduce pollutant emissions; and perhaps learn to use lower-grade, less refined fuels in some applications and thus save refining costs. For example, recent advances in combustion control promise to ameliorate emissions of nitrogen oxide pollutants from power plants.

Among the important problems in combustion are the interactions of chemical and flow phenomena in combustion; the high-temperature oxidation kinetics of hydrocarbons, alcohols, ketones, and other compounds; surface catalytic reactions involving oxidation-reduction, pyrolysis, and decomposition; nucleation and condensation of combustion products to form soot; the mechanisms in the formation of pollutants, such as sulfur and nitrogen oxides, hydrocarbons, and other substances; and materials properties for refractories and coatings resistant to corrosion and other kinds of damage in hot gas environments.[20]

CHEMISTRY OF COAL

Understanding of coal has lagged far behind that of petroleum, not only because coal is chemically more complex, but also because the growing use of petroleum and natural gas in the past reduced the economic and scientific incentives to study coal. The available data on the chemical properties of different coals are inadequate, as are procedures for analyzing coals. The chemical reaction paths in important coal conversion processes are incompletely known, as are the thermal and kinetic interactions in combustion equipment and synthetic fuel plants.[20]

Research on these topics can yield a number of important practical benefits. Coal conversion involves reacting some of the hydrogen-deficient coal with water (or some other material whose original source of hydrogen was water) to turn the solids in coal into liquids and gases. This process is expensive in both materials and energy. Improved, more selective catalysts would lower both the requirements for and expense of hydrogen and also facilitate removal of oxygen, nitrogen, and sulfur from the products. Improved surface characterization techniques would be useful in understanding observed kinetic behaviors in catalytic gasification and would thus allow improvement of catalysts for those processes. The inorganic pollutants emitted by coal combustion could be more precisely controlled if the high-temperature interactions of the the mineral matter, and the nitrogen and sulfur, were better understood. The problems of slagging and fouling could also yield to this improved understanding. In virtually all the ways coal is and may be used—direct combustion, gasification, and liquefaction—a better understanding of coal's chemical structure and reactions would contribute to lower costs and better environmental qualities.

R&D Related to Some Other Energy Sources

New clean fuels and clean energy conversion schemes deserve particular attention. These include energy storage, generation and use of hydrogen, and fuel cells. R&D on high-quality work on amorphous semiconductors will aid the development of photovoltaic materials for solar-electric systems.

ENVIRONMENTAL EFFECTS

We have already mentioned the global CO_2 problem, about which we are somewhat uncertain but very concerned.

More generally, we still only poorly understand the environmental impacts of energy use. For example, a vital question, with extremely important economic implications, is the relationship between certain fossil-fuel emissions and public health problems. Combinations of various pollutants—sulfur and nitrogen oxides, particulates, hydrocarbons, and metals—are probably more hazardous than each alone, but the effects of either individual or combined pollutants are very poorly characterized. Better knowledge here could allow more cost-effective control of air quality and probably improved public health. This is especially important given that increased energy use may result in lower-grade fuels.

Another important topic is water consumption by energy facilities, especially synthetic fuel and power plants. Obviously, withdrawing too much water from a river can have extreme environmental consequences; some rivers in the West already demonstrate the effects. Two main areas of research may help us to produce needed energy, while avoiding undue ecological damage and shortages of water for nonenergy use. First, basic work in the reactions involved in synthetic fuel production could decrease greatly the need for water in these processes. Second, water actually available over the long term in river basins and underground aquifers across the country must be known with greater certainty, and in much greater detail, to allow optimal siting of water-consuming facilities.

THE IMPLICATIONS OF A MORE ELECTRIC FUTURE

The likelihood of increasing electrification has been proposed a number of times in this chapter. From the standpoint of energy use, what does that imply? There is not room here to develop the concept very far, but the following changes appear likely and should be examined:

- In the long run, the manufacture of hydrogen with off-peak power might serve the dual purpose of supplying portable fuels and increasing the energy-efficiency of the electric power system. In fact, "hydrogen-electric future" is probably the best descriptive phrase. The hydrogen would be used in aircraft, in many industrial processes, and possibly in electric reconversion fuel cells.
- Substantially electrified mass ground transport, like that in present-day Europe, may play an increasing role.
- More efficient electric heat pumps might become the dominant means of heating and cooling buildings. These could run on an annual cycle, making ice in the winter for use in summer cooling.
- Cogeneration could become much more common in industry.
- Electricity might find wide use in automobiles, in either all-electric or partly electric cars. The former would require much improved batteries; in the latter, a small engine might run continuously at optimum efficiency to operate a small electric generator and electric motors at the wheels, with a battery charged by the generator to handle peak demands.

CONCLUSION

This assessment of our current energy situation might seem to imply that nothing but disaster lies ahead. Not so; the danger is there, but not the certainty. Neither the United States nor the world is running out of energy. The flow of oil and gas will not cease overnight, and a dedication to the conservation of energy as a social norm will relieve the strains on a changing supply system. Finally, the potential of solar energy, fusion, and new forms of fission are almost infinite if they can be developed and prove socially acceptable.

The difficulty ahead is in reconciling immediate problems with these opportunities. Of the problems the most apparent is our great reliance on gaseous and liquid fuels, both of which are in finite supply. The use of oil and gas is woven into the technical infrastructure of our society—transportation, industrial processes, heating, chemical manufacture, and so forth. The task ahead is not simply recognizing that these fuels will in time become scarce or in estimating when that will happen, but rather of preparing our society for a transition to new forms of fuels and a more careful use of them.

One complication is in the quite different perspectives of those who must participate in that transition; the difference, for example, between the time it takes scientists and engineers to develop a new energy technology to a commercial level and the usually shorter times when business expects to make a profit on its investments. Also, it is difficult for all to agree that a quite different energy future is possible; that, for example, liquid hydrocarbon fuels are not necessarily the essential base for advanced technological civilizations; that we have simply designed our society around those fuels most available, most convenient to use, and, until recently, quite cheap. The fuels are still convenient, but no longer as available nor as cheap; and, therefore, we must begin to think of other ways to fuel our technology. As an example, if we are technically clever about it, hydrogen can be used to fuel airplanes; indeed, much of our transportation could in time operate on energy from stationary sources, such as electric power plants. However, there will continue to be some need for liquid and gaseous fuels in some small amounts.

In the end, the discussion returns to questions of social purpose, economics, environmental costs, and public understanding. Our energy future will depend in part on whether energy is available, on its forms, and on our technological ingenuity, but even more on decisions made by society on what technological options to use, to what purposes, and with what safeguards.

OUTLOOK

The following outlook section on energy is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

Over the next five years, the United States will confront basic questions regarding its future energy pattern. How rapidly will oil and gas resources, domestic and global, decline? Can growth in energy demand be tempered sufficiently to allow the smooth substitution of new energy sources for oil and gas? What form will these replacements take and what problems may their use entail?

The answers to these questions are shaped in part by time. For the mid-term transition from oil and gas to other energy forms, the United States has in reality very little maneuverability, in both energy supply and use. Its major alternative sources will be coal and fissionable uranium; growth in energy demand, while it can with sufficient time be smoothly and slowly reduced with little harm, cannot be cut sharply without causing serious economic disruptions.

For the long term, here assuming beyond 2000, the United States has a number of potential options: various applications of solar energy; fission reactors using more plentiful uranium isotopes; possibly fusion (though somewhat later than the other options if at all); and much improved efficiency in energy use.

Most of the mid-term and long-term energy supply technologies are electrical, and this suggests a change in the character of energy supplies, from increasingly scarce liquid and gaseous fuels to electricity. Technological ingenuity can transform such resources as coal and oil shale into gases and liquids, but at considerable costs, both economic and in several cases environmental.

That this country's energy pattern will change in the coming decades is certain. Oil production and discovery are unlikely to keep up with rising demands. World reserves of oil in 1978 were only about 2 percent above 1972 levels, though meanwhile prices had quintupled and world demand for oil had risen more than 15 percent.

The United States is particularly vulnerable. Its oil reserves and production have declined over the past 10 years as its demand for oil has grown, and it will likely be forced in coming years to draw more heavily on an increasingly limited world market for oil. The domestic gas outlook is even worse. The world has a good deal of gas, which is potentially available, but, except for that in Mexico and Canada, it must be liquefied and transported to the United States at very low temperatures in special tankers. Its expense and hazards may make imported liquefied gas unattractive compared to alternatives, the strategic consequences of reliance on foreign sources aside.

These facts suggest that we are overdue in considering ways of limiting U.S. dependence on oil and natural gas, by substituting other forms of energy where possible and by instituting vigorous measures to increase the efficiency with which energy—oil and gas in particular—is used. At the same time, possible means of securing stable access to energy for the indefinite future must be examined. This will require new technologies and the political will to develop and use them responsibly. In principle both can be done; in practice each will be difficult.

Financial and political institutions are more responsive to the short-term view dictated by market rates of return and electoral intervals than to the longer and broader view determined by domestic resource depletion and the risks of depending heavily on imported energy. Only an educated public and enlightened public officials can temper this disparity by responsible efforts to avoid the difficulties to which short-range expediency can commit us.

The biggest difficulties will be not purely, or often largely, technological, but rather matters of social decision—what levels of pollution to accept, how best to protect national security, and even what, in broad terms, constitutes a desirable society.

CONSERVATION: THE HIGHEST PRIORITY

Conservation appears to deserve the highest priority in energy planning over the next several years. First it can provide some quick and fairly cheap gains through lowered thermostats, more careful thermal control in manufacturing processes, various kinds of insulation, and some simple waste heat recovery. Second, very substantial technological improvements indeed will be economical and more important in the longer term—as buildings, vehicles, and industrial equipment are replaced or rebuilt. Recent econometric and engineering analyses suggest, for example, that rising energy prices and conscious conservation incentives over the coming 35 years or so could as much as halve the ratio of energy consumed to economic output in the United States, with negligible impacts on economic growth. Again, this must be a gradual process, or it could result in grave economic losses. Furthermore, it would require continuing determined effort and cooperation between public and private sectors—a more-than-difficult task.

MID-TERM SUPPLY STRATEGIES

Next to planning conservation policies, the second important issue that must be settled in the next five years is how to begin the transition away from oil and gas and toward energy supplies that are essentially infinite. The first step is to determine what is available now. The choices are rather strictly limited: coal and nuclear fission, with some help from solar energy and perhaps a little geothermal energy in suitable locations.

The two major energy sources, coal and fission, both have serious problems. Coal, for example, is a prolific source of pollutants and other hazards, including the long-term climatic risks of CO_2 emissions; also, while plentiful compared to oil and natural gas, coal is not infinite. Nuclear fission in its current form, the light water reactor, depends on a quite limited resource: the naturally fissionable isotope of uranium; it is also under political and regulatory pressure that perils the reactor industry. Neither coal nor nuclear fission using light water reactors is appropriate for the long term.

However, the promising long-term technologies—solar energy in its more advanced forms, nuclear fission in breeder reactors, and controlled nuclear fusion—are far from ready for commercial deployment in the United States; none is likely to make a major contribution before the end of the century. Given the speed with which the world's oil and gas resources are being depleted, there is serious question about whether any of these will become timely replacements for coal combustion and nuclear fission, let alone serving as vehicles for a smooth transition directly from oil and gas. Thus, in the next several years the nation has no choice but to use what is available, improving where possible our ways of using coal and nuclear fuel, while doing the work necessary to ensure the availability of more appropriate long-term technologies. The difficulty of the consequent industrial and economic shifts reinforces the vital importance of energy conservation.

While there is intense interest in synthetic liquid and gaseous fuels from coal and oil shale and several large demonstration projects do exist, synthetic fuel production is still on a rather primitive basis. Developing it to the commercial stage will take another decade at least. Even then it is likely to be limited to a small fraction of present oil and gas use, notably by the availability of fresh water to meet the large requirements of the processes.

This apparent imbalance between likely demand and the supply of liquid and gaseous fuels presents a choice that will have to be seriously addressed in the next several years. We can either try to improve coal-based synthetic fuel technology to the point at which it can serve our energy needs in much the manner of oil and gas today or restructure the technological basis of civilization to reduce permanently the heavy dependence on such chemical fuels by turning more toward electricity. In practice, both approaches will be necessary, but the balance between them is complicated and ill-understood.

SOME OUTSTANDING ENVIRONMENTAL AND HEALTH EFFECTS OF ENERGY USE

Energy production in the past has had unforeseen impacts on the human environment. The costs of air and water pollution, waste-disposal problems, and many occupational hazards in energy production and use are only now beginning to be appreciated. It is not possible to rank the various energy supply options precisely according to the relative severity of their impacts. However, there are grave reasons for special concern about the impacts of certain energy sources given current energy projections.

Especially important are those of fossil-fuel combustion, and particularly those of coal with its enormous chemical variety and complexity. As it increases its reliance on coal, the nation in the next several years will have to address the possible climatic consequences of continued emissions of carbon dioxide, and air pollution in general. Also, water availability and competition for other water uses will impose severe restrictions on the siting of energy facilities, and will require major technological developments in water use as well as large water-related investments.

LONG-TERM ENERGY OPTIONS

Only two major classes of long-term energy supply options exist: solar power in various forms, and nuclear power, both fission and fusion. These can be, and are now, augmented in especially favorable locations by modest amounts of geothermal power, but the normally microscopic heat flow from inside the earth makes solar power a better bargain in most places. The potentials of tidal power, and probably also wave power, are quite overemphasized.

NUCLEAR FUSION

The prospects for fusion support what can be described as guarded optimism. It seems clear now that an acceptable fusion plasma can be produced and confined in a vacuum by strong magnetic fields—a plasma not touching any physical surface and kept together long enough for an interestingly large fraction to react. Thus, in a scientific sense, fusion power appears feasible. It has taken about 25 years and several billion dollars to reach this stage.

The most expensive and difficult parts—the technology and the practical engineering of a working reactor producing power—are still to come. The basic technological feasibility of controlled fusion will not become clear for at least a decade, and the engineering and economic feasibility much later. Fusion will come (if at all) later than breeder reactors and advanced solar power systems, both of which may be called for by the end of this century. Still, it remains a contender for the long term.

SOLAR ENERGY

Solar power requires investments embodying nonrenewable resources that cost money, though sunlight itself is free. The real attractiveness of solar power, besides its ubiquity, is the relative ease with which it can be put to various uses, such as water and space heating. This attraction, however, has been ignored by various high-technology schemes.

Solar power, by current convention, includes not only direct conversion of sunlight into useful forms, but also hydroelectric power, winds, and biomass (organic matter) considered as an energy source.

The most immediately promising solar application is in producing heat in the 70–200°C range, for residential hot water and space heating and a variety of commercial, agricultural, and industrial purposes. The simplest systems use flat-plate collectors. More advanced ones use mirrors or lenses to concentrate the solar heat and provide higher temperatures. All have benefited in recent years from advanced science and engineering. Another decade of rising oil prices plus improvements in commercial solar systems should put this technology in a strong economic position. The federal government has acted to provide some economic stimuli.

At the high-technology end of the possibilities are advanced photovoltaic conversion schemes, the so-called power tower concept, in which a vast array of steerable mirrors focuses sunlight onto a boiler atop a several-hundred-meter tower, and ocean thermal electric conversion systems, which exploit the temperature differences between the surface and the deeps of tropical ocean water. Photoelectric generation has two main branches: crystalline and amorphous photovoltaic materials. Both show promise and both warrant continued effort. The power tower does not look promising; it is expensive, and significant cost reductions cannot be expected. Ocean thermal energy conversion and proposed solar satellites beaming power to earth represent similar attempts to apply high technology inappropriately.

Windpower devices occupy an intermediate status—neither high technology nor low—and like most other solar technologies should be regarded as augmenting conventional power supply. Hydroelectric dams present small risks of dam failures. They also chop rivers into ecological bits and unnatural parts, with consequences disproportionate to the power produced. This may be true also of small-scale hydroelectric installations. In the field of biomass, the equivalent of 900-1800 million barrels of oil probably could be produced annually in the unlikely event that all the nation's farm, forest, and domestic wastes were exploited. "Energy farms" would not only present the ecological problems of any monocultured crop, but also might compete directly with food and wood production.

NUCLEAR FISSION FOR THE LONG TERM

The light water reactors that now provide almost all the nation's nuclear generating capacity exploit only about 0.6 percent of the energy potentially available in their fuel. According to one estimate, reasonably assured domestic uranium resources would suffice to fuel only about 300 thousand-megawatt power plants.[18] The fuel could be reprocessed to recover plutonium and some still usable uranium, but these additions would increase the total available energy by a factor of only about 1.6 or 1.7. Conventional methods of reprocessing plutonium for re-use in light water reactors lead to its brief appearance in a form suitable for nuclear explosives. If this course were taken, the proliferation resistance of the nuclear fuel cycle would be endangered for at best a moderate benefit.

Advanced converter reactors, the most frugal of which would use only about half as much uranium (and might also use thorium) to generate the same amount of energy, even without reprocessing, could help. Their contributions would be significant for the long-term, however, only under some national circumstances, especially of low growth in electric power demand.

If nuclear power is to have a long-term future, it must be with breeder reactors. These reactors as now envisaged convert the more common nonfissile uranium isotope, uranium-238, to plutonium, then fission the plutonium to provide heat. Breeder reactors can also be designed to convert thorium-232 to fissile uranium-233. Using breeders, more than 70 percent of the uranium that is mined can be effectively used, yielding about 70 times the energy from a given amount of uranium as can be obtained from light water reactors, even with fuel reprocessing. This means that we could then afford to pay very much more for the uranium than before; at that price, the reserve base becomes large enough to last millions of years. This much reduced uranium need is reflected in a similarly reduced need for mining and ore processing. Thus, the hazards of radiation exposure in these activities, now subjects of lively debate in connection with light water reactors, will be almost eliminated.

Breeder reactors require fuel reprocessing, again raising the question of nuclear weapons proliferation. New techniques of reprocessing, however, can be designed so that the availability of plutonium-239 or uranium-233 in a form usable for weapons is much reduced.[21] Nonetheless, breeder fuel in its present form represents a proliferation risk.

All fission reactors produce nuclear waste, and the total amount (measured by its radioactivity) depends hardly at all on the reactor type. An adequate science and technology base for entombing these wastes in geological formations exists. With proper site-specific engineering and geological knowledge, the long-term public hazards arising from waste depositories would consist of low-level radioactivity less than that from natural background radiation or other nuclear fuel cycle activities. What is lacking today is the detailed engineering and assessment for specific suitable repositories.

RESEARCH AND DEVELOPMENT NEEDS

Several important fields for energy research are discussed in the main body of this chapter, particularly under the heading "Some Research Needs." However, even the synopsis just presented reveals some particularly pressing problems:

- The relationship between energy use and economic growth is obviously vital. The economic impacts of constraining energy demand cannot be very precisely projected, but some good beginnings have been made in recent years.
- Energy conservation—in industry, in transportation, in buildings, and in appliances and services of many kinds—still receives inadequate attention.
- Because of the necessity of reducing global dependence on fossil fuels in response to world environmental hazards, and the need to increase use of some fossil fuels—especially coal—during the period of transition to more benign long-term energy sources, we must learn to understand the chemistry and combustion kinetics of these fuels better, in order to derive maximum benefit from limited use.

NOTES AND REFERENCES

1. Hayes, E.T. Energy Resources Available to the United States, 1985 to 2000. *Science* 203(4377):233–239. 1979. An excellent overview of the prospects for energy supply (although we do not agree with him on some of the economic and environmental assessments implicit therein).

2. CONAES Demand and Conservation Panel. Energy Demand: Some Low Energy Futures. *Science* 200(4338):142–152, 1978.

3. *Energy Modeling for an Uncertain Future, Supporting Paper 2* (NRC Committee on Nuclear and Alternative Energy Systems). Washington, D.C.: National Academy of Sciences, 1978.

4. Probstein, R.F., and H. Gold. *Water in Synthetic Fuel Production: The Technology and Alternatives*. Cambridge, Mass.: MIT Press, 1978, p. 151.

5. Burwell, C.C. Solar Biomass Energy: An Overview of U.S. Potential. *Science* 199(4333):1041–1048, 1978.

6. Harte, J., and M. El-Gasseir. Energy and Water. *Science* 199(4329):623–634, 1978.

7. At much higher prices, the alternatives (synthetic fuels, nuclear or solar power, etc.) would become relatively attractive and thus tend to drive oil out of many markets.

8. Moody, J.D., and R.E. Geiger. Petroleum Resources: How Much Oil and Where? *Technology Review*, March/April:39–45, 1975.

9. Pretreatment of high-sulfur coal, mainly by washing, has been underemphasized as a sulfur-control method.

10. There appears to be no practical way of scrubbing the CO_2 from any appreciable fraction of fossil-fuel emissions.

11. See G.M. Woodwell. The Biota and the World Carbon. *Science* 199(4325):141–146, 1978; Minze Stuiver. Atmospheric Carbon Dioxide and Carbon Reservoir Changes. *Science* 199(4326):253–258, 1978; U. Siegenthaler and H. Oeschger. Predicting Future Atmospheric Carbon Dioxide Levels. *Science* 199(4327):388–395, 1978; C.S. Wong. Atmospheric Input of Carbon Dioxide from Burning Wood. *Science* 200(4338):197–200, 1978.

12. *Air Quality and Automobile Emission Control*, Vol. 1: Summary Report (NRC Coordinating Committee on Air Quality Standards). Washington, D.C.: National Academy of Sciences, 1974, p. 11.

13. *Nitrates: An Environmental Assessment* (NRC Coordinating Committee for Scientific and Technical Assessments of Environmental Pollutants). Washington, D.C.: National Academy of Sciences, 1978.

14. *Principal Conclusions of the American Physical Society Study Group on Solar Photovoltaic Energy Conversion*. New York: American Physical Society, 1979.

15. Estimate of National Hydroelectric Power Potential at Existing Dams. U.S. Corps of Engineers, 1977. A better calculation goes roughly as follows: if every drop of rain falling onto New England passed through a 100 percent efficient hydropower turbine and no water flowed downhill otherwise, the total power generated would be about 18,000 megawatts (yearly average). From this, subtract 70 percent for evaporation and vegetative transpiration, to obtain 5,100 megawatts. The Corps of Engineers estimates that half this power could be generated, but that would require stopping the flow on half of every stream, no matter how large, small, accessible, or remote.

16. Rossin, A.D., and T.A. Rieck. Economics of Nuclear Power. *Science* 201(4356):582–589, 1978.

17. *Risks Associated with Nuclear Power: A Critical Review of the Literature*, Summary and Synthesis Chapter (NRC Committee on Science and Public Policy). Washington, D.C.: National Academy of Sciences, 1979.

18. *Problems of U.S. Uranium Resources and Supply to the Year 2010, Supporting Paper 1* (NRC Committee on Nuclear and Alternative Energy Systems). Washington, D.C.: National Academy of Sciences, 1978.

19. *Report of the Office of Science and Technology Policy Working Group on Basic Research in the Department of Energy*. Washington, D.C.: Office of Science and Technology Policy, 1978.

20. *The Department of Energy: Some Aspects of Basic Research in the Chemical Sciences* (NRC Committee on Chemical Sciences). Washington, D.C., National Academy of Sciences, 1979.

21. Levenson, M., and E. Zebroski. "Fast Breeder System Concept—A Diversion-Resistant Fuel Cycle," paper presented at the Fifth Energy Technology Conference, Washington, D.C. February 27, 1978.

6 Materials

INTRODUCTION

The near-term future for materials will be shaped largely by our responses to forces already well recognized. Some of these forces are long-standing and amount collectively to supply and demand. Others have emerged only recently and are related to national problems that include energy, environment, governmental regulation, and productivity and innovation. Still others relate to worldwide political and institutional changes and growing global interdependence. We shall touch on some of these materials-related pressures later. Together these forces are creating clear-cut challenges and opportunities in materials science and technology.

Americans consume annually about 10 tons per capita of materials and 15 tons of coal-equivalent energy. It is difficult to project the particulars of future materials needs either nationally or internationally. For this country, however, it can be said that the needs will increase incrementally in quantity, but that the kinds of materials used will not change much during the next five or ten years. Beyond that period, the national origins and the kinds and amounts of the materials used could change markedly. New technologies change both what is usable and how it is used.

Expenditures on materials research and development in this country grew rapidly beginning in World War II and peaked in the 1960's. In current dollars annual expenditures today are hovering at about five times the level of the 1950's, but are declining in constant dollars.

The results of the past 35 years of materials research and development have been spectacular. They have led to major new industries, including computers, plastics and synthetic fibers and rubbers, and nuclear power.

DEVELOPMENTS IN MATERIALS

The rapid evolution of new materials that began largely during World War II is still under way. Research and development has become more selective, however, and a variety of business factors are inhibiting the movement of new methods and products from the laboratory to everyday use. In short, we are seeing a decline in the innovative risk-taking that characterized the rapid materials progress of 1950–70. Still, advances should continue to be stimulated by the interaction of industrial technology with basic research in the universities, government, and industry.

Selected developments expected during the next five years are treated here in terms of eight classes of materials: metals and alloys, energy-related materials, information-related materials, polymers, ceramics and other inorganic materials, composite materials, renewable materials, and biomedical materials. These developments will illustrate the achievements and near-term potential of materials science and technology. They are limited largely to engineering materials and structures; we have not, for example, attempted to treat the extraction of raw materials. Except in unusual circumstances, technological developments that will mature during the next five years will rely on research already in hand.

METALS AND ALLOYS

In basic metals like iron and steel, aluminum, and copper, striking advances on a broad front cannot be anticipated. We are more likely to see incremental improvements in specific properties, reduction in cost through processing innovations, and better tailoring of properties to meet specific needs. Technological innovation and gains in productivity in the basic materials industry in the five years to come will depend less on technical progress than on the generation and availability of capital to modernize plants and improve technological capability.

Ability to control the properties of metals depends in part on knowledge of their internal structure and composition. Metals normally are composed of grains made up of microscopic crystals in which atoms are spaced in a pattern characteristic of the particular metal. Although the grains have a regular, crystalline structure internally, they have irregular shapes. Irregular surfaces are created when adjacent crystals interfere with each other as they form from the molten metal. Impurities in metals tend to segregate at the grain boundaries, the interfaces between grains.

The early 1950's saw the development of tools like the transmission electron microscope that permitted direct observation of the internal characteristics that control deformation and fracture in metals. With these tools, metallurgists were able to clarify in detail the relationships among the composition, microstructure, and properties of metals and how they determine performance in service.

Superalloys

More recently, technological progress has made it possible to raise significantly the permissible operating temperatures for nickel- and cobalt-base superalloys in gas turbines. Superalloys are alloys intended for service above 750°C and are used routinely in jet engines (Figure 21). The higher operating temperatures of around 950°C—which were achieved by designing microstructures that give the alloys unusual resistance to deformation and fracture—increase the efficiency of jet engines with significant improvements in performance and savings in fuel.

A further step has been the use of directional solidification to achieve unique high-temperature strength and performance in the nickel-base superalloys. Directional solidification of the molten alloy yields a structure that is strongly anisotropic—the properties of the metal are strikingly different in different directions. The part can be designed, therefore, so that it is strongest where it must bear the heaviest load. Further improvements have been made by the use of controlled solidification to produce single-crystal turbine blades and vanes of such alloys (Figure 25c). The absence of grain boundaries in these materials frees them of certain detrimental effects of grain boundaries on high-temperature behavior. Directionally solidified materials and single-crystal blades and vanes will be demonstrated in gas-turbine engines within the next five years. Both advances will increase the durability of turbines and may increase peak operating temperatures and thus efficiency.

Coatings with unusual resistance to oxidation have been developed to protect high-performance alloys in turbines and other hot environments. These coatings, together with the development of nickel-base alloys as mature engineering materials, have been key factors in the outstanding operating experience and long periods between overhaul of commercial jet aircraft engines.

High-Strength Steels

Within the past five years, steels have been introduced that combine relatively high strength with costs only minimally higher than those of conventional low-carbon steels, which have been the workhorse of the structural steel industry. These high-strength steels are being used increasingly where higher strength-to-weight ratios are advantageous. In automobiles, for example, they may prove to be the most cost-effective means of reducing weight and thus improving fuel economy in the near term. It has been estimated that as much as 500 pounds per car of the new steels will be used in U.S. vehicles by 1985. The weight of these steels required to perform a given function in a car is expected to be 10–30 percent less than the weight of the materials replaced.

These high-strength steels are made by coupling the effects of minor concentrations of alloying agents with control of the rolling and quenching operations in the steel mill. The resulting metal has a fine-grained microstructure and correspondingly higher strength. One class of high-strength steels, the dual-phase steels, combines the ductility and fabricability of the softer, low-carbon steels with the strength usually associated with higher-carbon, hard steel. This unusual combina-

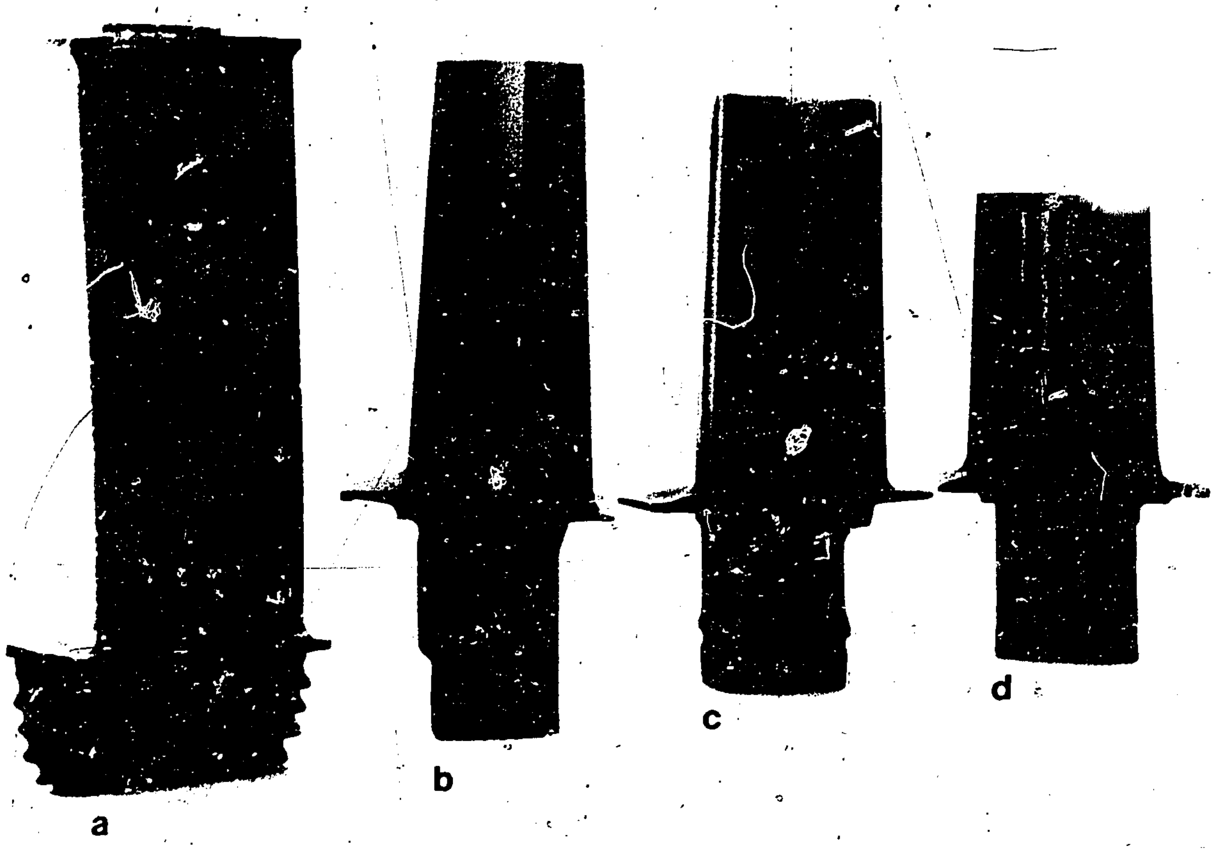


FIGURE 21 Gas-turbine blades etched to reveal grain structure. The solid cast blade of the 1950's (a) had a thrust/weight ratio of 1 and was used at 880°C. By the 1970's, the ratio was 1.9 and the hollow cast blade (b) was used at 950°C. For the 1980's, a single-crystal blade (c) will have a 2.2 thrust/weight ratio and operate at 970°C. The blade designed for the 1990's (d) will be made with directionally recrystallized columnar grains and will have a thrust/weight ratio of 2.9 and be able to operate at about 1,000°C. (A. R. Cox, Pratt & Whitney Aircraft Group, West Palm Beach, Fla.)

tion of properties is achieved at only modest increases in cost, and within the next decade the dual-phase steels should become a widely accepted class of structural steels. However, their availability from the U.S. steel industry currently is inhibited by economic factors, notably the cost of installing continuous annealing facilities.

Glassy Metals

A number of industrial organizations in this country and abroad are exploring commercial uses for a new class of materials, the glassy metals. Although metals are normally crystalline, certain alloys can be solidified in noncrystalline, or amorphous, form, like glass, by very fast cooling—100,000°C–1,000,000°C per second. These materials consist of metals like iron, cobalt, and nickel, alloyed with elements such as phosphorus, silicon, and boron. They can be solidified in glassy form in ribbons seven–eight centimeters wide and a fraction of a millimeter thick.

Glassy-metal magnets are very strong mechanically, and some show very low losses of energy of the kind that occur in other magnetic materials during magnetic cycling. This combination of properties makes the glassy metals good candidates for replacing iron–silicon alloy sheets in the cores of transformers used in the transmission of electrical power. It has been estimated that conversion to glassy-metal transformer cores could save the energy equivalent of 6 million barrels of oil per year.[1] The saving could be realized only gradually, however, because no more than about a tenth of the huge installed transformer capacity could be replaced annually.

Magnetic glassy metals in strips several centimeters wide can be purchased in development quantities today. The prospects for commercial use of the materials in the next five years, in applications such as magnetic shielding, are quite good. Glassy metals also resist corrosion exceptionally well. The hope of exploiting this characteristic undoubtedly will stimulate much research and development during the next five to ten years.

Corrosion

Corrosion of metals costs the nation billions of dollars annually. The problems range from rusting and deterioration of consumer products to failure of high-performance parts in hostile (e.g., hot, corrosive, erosive) environments. Corrosion is a problem particularly in the development of new energy-conversion devices in which hostile environments cause metal parts to deteriorate rapidly.

Many means of preventing corrosion have been provided by research on the mechanisms of corrosion; developments in protective coatings, paint systems, and surface treatments; and the availability of intrinsically corrosion-resistant materials, such as titanium. (It has been argued that the economic burden of corrosion could be reduced considerably by improved dissemination—and consequent wider application—of information already available on corrosion prevention technology.) The progress now being made in surface science and technology offers the hope of acquiring new knowledge of the mechanisms of corrosion and the reactivity of metals at the atomic level—knowledge that could lead to new approaches to corrosion control.

Magnet Alloys

The cobalt–samarium alloys introduced recently for use in permanent magnets are finding a growing market. The materials exhibit unusually high magnetic force per unit of weight and so are especially useful in small, permanent-magnet electric motors where they replace a great deal of copper and also save weight. These alloys have stimulated new concepts in the design of electric motors that would be significantly smaller and lighter than is common today. The full-scale commercial use of the materials depends on the cost and supply of cobalt and samarium and the development of cheaper ways to form the alloys into magnet shapes.

Zaire is a primary source of cobalt, and the possibility of interruptions in supply is spurring a search for lower-cobalt or cobalt-free alloys for permanent magnets of equal efficiency. Alloys of rare earths are being sought as less costly replacements for samarium.

Newer Engineering Metals

Newer metals whose uses have grown during the past three decades include the reactive metals—titanium, zirconium, and hafnium—and the refractory metals—niobium, tantalum, molybdenum, and tungsten.

Of the reactive metals, titanium has enjoyed the greatest growth. The metal was introduced commercially in the late 1940's, and its light weight, strength, and corrosion resistance soon assured its future. Titanium is now entrenched as an aircraft structural material and has a growing market as a corrosion-resistant material for use in chemical plants. Recently the metal has been used in condenser tubes in steam-power generation, and it is being introduced as a strong, corrosion-resistant material for the blades in low-pressure steam turbines.

Production of titanium mill products in this country in 1979 will total some 20,000 tons. Although titanium is the fourth most abundant metal in the earth's crust (after aluminum, iron, and magnesium), current demand for the metal could strain world production capacity.

Zirconium and hafnium have significant uses in nuclear reactors for generating electric power. Zirconium is used to clad the uranium oxide fuel elements in light water reactors. The metal's essential characteristic in this application is its very low absorption of the neutrons that drive the fission process; the metal also has good corrosion resistance in high-temperature water. Hafnium, on the other hand, is an excellent absorber of the neutrons that drive the fission process and also has excellent corrosion resistance in high-temperature water. These properties make hafnium ideal for its present use in the control rods of light water reactors.

The refractory metals have been used for years in small parts in electrical and chemical equipment. With the advent of missiles and space vehicles, it appeared that these metals would be needed in larger-sized sheets, bars, and forgings, and the necessary production facilities were developed in the 1960's. However, many of the missile and space-vehicle applications either never materialized, because alternative materials were used, or required less of the metals than anticipated at first. Consequently, the nation's production facilities for refractory metals are underutilized at present. The facilities have proved valuable, however, in producing materials for hot-working dies and for chemical equipment.

Metal Science

Metallurgical science now is benefiting from new microanalytical tools such as auger spectroscopy and scanning and transmission electron microscopy. These powerful tools—by-products of basic research in physics—permit detailed study of the segregation of impurities that is known to control many of the critical properties of alloys. Perhaps more importantly, the new tools allow the character of solids to be probed down to the atomic scale. The resulting knowledge is accumulating rapidly and will lead to major developments in the understanding and control of surfaces—so important in corrosion and catalysis. The new knowledge also may lead to the ability to control and manipulate the composition and properties of internal surfaces (grain boundaries) in alloys. This ability would permit the properties of alloys to be tailored to particular uses to a degree not possible today.

ENERGY-RELATED MATERIALS

Power Devices

Better performance at lower cost will continue to be achieved in semiconductor power devices for converting alternating to direct current. These rectifiers, therefore, will gradually replace most of the older electromechanical equipment. Power devices generally will be made from silicon except for operation at the highest frequencies, where gallium arsenide may be superior. An important materials problem in the power area is the dissipation of heat generated in various types of equipment. The use of electrical insulators of high thermal conductivity, such as beryllium oxide, will remain important for this purpose. Alternative materials that may find use in some applications include copper and aluminum nitride and perhaps even diamond.

Dielectrics

Dielectric materials, which include polymers and ceramics, are insulators or nonconductors of electricity. These materials are important in equipment for generating, transforming, and transmitting electric power. They are also critical in telecommunications and various electronic applications. In the electric power industry, new insulating materials are needed for high-voltage transformers and capacitors, where they would replace the polychlorinated biphenyls (PCB's), which are environmentally unacceptable. Work is also under way on new dielectric structures for storing high-density energy to meet the requirements of modern electrical circuitry. Here, the role of surfaces at the interface of the dielectric and the conductor must be better understood. Also needed are practical ways to measure the degradation of dielectrics so that their service life can be forecast accurately.

Solar Cells

A considerable effort is under way to produce efficient, low-cost solar cells for converting sunlight directly to electricity. Silicon is the standard cell material; encouraging progress is being made in growing ribbons of single-crystal silicon by a continuous process, which would significantly reduce the cost of the cells. Some progress has been reported in exploring amorphous silicon. Highly efficient gallium arsenide cells have been made for use in sunlight concentrated by reflectors. More novel compounds of two and three elements, such as aluminum gallium arsenide, are also being investigated for this application. The cost of producing solar cells should drop significantly in the next five years, but they are unlikely to become a major source of electrical energy during this century.

Lighting

The development of more efficient lighting is being spurred by the cost of energy and by improvements in fluorescent-lamp phosphors and lamp-envelope (bulb) materials. Fluorescent lights with improved efficiency and color will find greater use in home lighting in the next five years. High-pressure sodium- and metal-vapor discharge lamps with superior color rendition and efficiency will be replacing mercury discharge lamps used currently in many outdoor and commercial applications.

Superconductors

A continuing search is in progress for materials that become superconducting at higher temperatures than those now available. Wire made of a superconducting material does not resist the flow of electricity and so conducts current without the loss of energy caused by the electrical resistance of conventional conductors like copper and aluminum. A superconducting material loses its resistance to electricity at a characteristic temperature called the critical temperature, and the critical temperatures of all known superconductors are not far above absolute zero. To reach such temperatures, the materials must be refrigerated with liquefied gases, and the higher the critical temperature, the lower the refrigeration cost.

A niobium–germanium compound has the highest critical temperature—about 23°K (−250°C)—of any material found so far, but it is too brittle to be fabricated. Superconducting wire made of a niobium–tin compound, cooled by liquid helium, with a critical temperature of about 18°K carries the circulating current in superconducting magnets now in use.

Superconducting magnets are available in sufficiently large size—15 feet internal diameter—to operate subatomic-particle detectors in research in nuclear physics. Superconducting magnets also are used in advanced energy-conversion test facilities to contain hot, corrosive plasmas by means of a magnetic effect. This application may be important for thermonuclear fusion and magnetohydrodynamic power generation.

Superconducting wire shows promise for use in the electromagnets in electrical generators and motors. In this application it would permit major reductions in the size and operating costs of the equipment.

Work has also been done on superconducting power transmission lines. The problem here is the expense and difficulty of refrigerating long conducting lines.

INFORMATION-RELATED MATERIALS

Sensors

A variety of new sensor materials is on the horizon. Infrared detector arrays made from semiconductor crystals like indium antimonide, mercury telluride, and cadmium telluride will allow us routinely to "see" the world by the heat it emits, even in total darkness. Potential applications range from detecting tumors to locating sources of heat leakage from buildings and industrial operations. More recent development of the technology of thin-film infrared devices, as replacements for bulk-crystal devices, significantly extends the range of uses of infrared sensors. Also in prospect are new pressure-sensing materials, such as polyvinylidene fluoride and aluminum phosphate (berlinite). These piezoelectric materials, which generate small electric currents when stressed, will improve the performance of sound-wave sensors. This in turn will lead to developments such as microscopes that "see" by sound waves. Zirconium dioxide sensors have been developed and are being used to measure the oxygen content of molten steels, while zirconium dioxide and titanium dioxide sensors are now being used to measure and regulate the air/fuel ratio in advanced automotive engines. Control of the air/fuel ratio is a key requirement in the operation of vehicle exhaust-emission control systems.

Fiber Optics

Glass and quartz fibers have been developed through which information in the form of light signals can be transmitted several miles. Rigid control of composition and internal structure is essential in making the fibers. The sources of light will often be light-emitting diodes or lasers of high light-emission efficiency, and the light detectors will be made of silicon. The main advantage of these fiber-optic systems over conventional telephone lines is their much higher message capacity. Communication by fiber-optic transmission is well advanced and will see major growth in the next five years.

Electronic Displays

Electronic displays will continue to replace most of the mechanical devices used now to depict letters and numbers in cash registers, home appliances, instrument panels, and other equipment. Such displays will use light-emitting diodes (LED's), liquid crystals, and gas-discharge devices and will be driven by digital circuitry. LED's and gas-discharge devices emit light when stimulated electronically; liquid crystals change their reflectance of ambient light. Problems of performance of materials currently limit the utility of these devices. As the problems are solved, the devices will find many applications in addition to the already ubiquitous digital watches and calculators using liquid crystals and LED's.

Transistor and Computer Materials

Electronics is probably the most rapidly accelerating area in technology today, and silicon in the near term will remain the most emphasized electronic material. Our ability to control the composition, structure, and processing of silicon-base components has improved steadily over the years. As a result, almost annually for about two decades the number of electronic components on a single integrated-circuit silicon chip has doubled and the cost per component has fallen by nearly half. This momentum will continue for at least the next five years and probably longer. A very strong effort is under way in the development of new lithographic techniques employing electron beams and X-rays and of dry processing techniques such as plasma etching and laser-annealed ion implantation. These techniques will be used to provide increasingly complex circuits at lower costs. As circuits become more complex, continued improvement will be required in the quality of single-crystal silicon and in the control of process-induced defects.

Computer Memory

A technology is rapidly being developed for producing computer memory using magnetic substances such as gadolinium–iron–garnet. These bubble memories store information at very high density as microscopic magnetized domains. The production of these rather complex materials is demanding, but the opportunities they offer for low-cost, highly stable, mass memories will provide the driving force for their further development.

POLYMERS

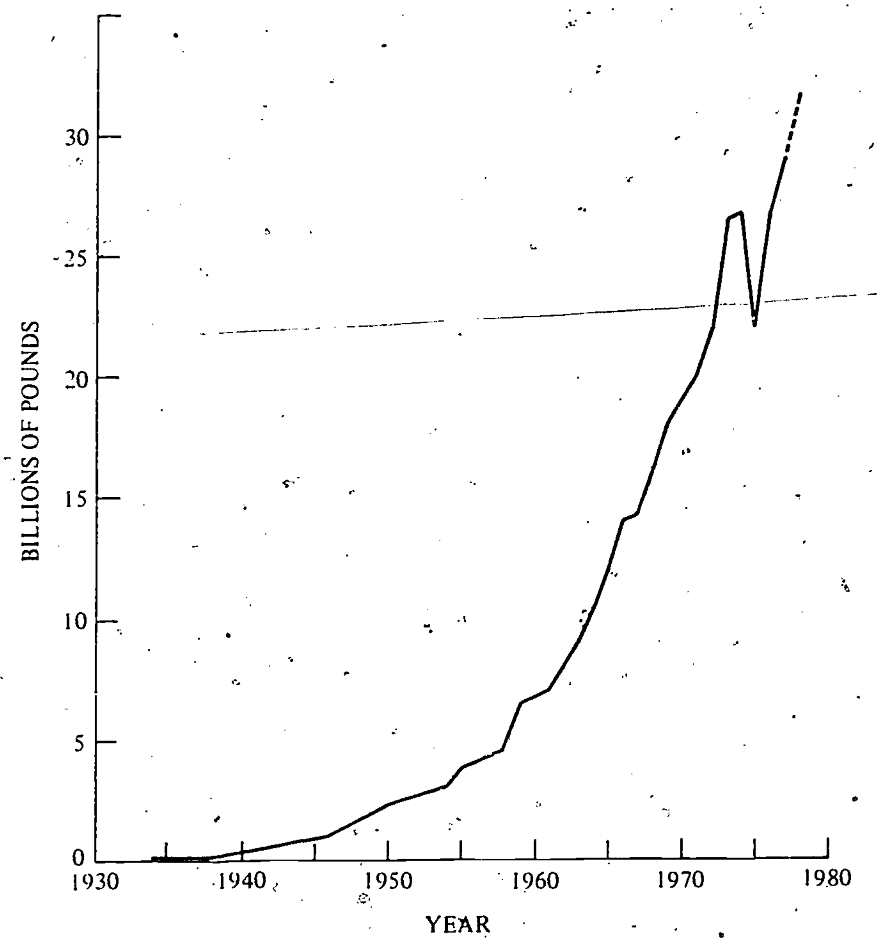
Synthetic polymers—plastics and rubbers—are the fastest-growing class of materials. Since about 1950, U.S. production of polymers has grown at a rate exceeding that for any other material and now tops the production of steel in volume but not in weight. U.S. production of plastics in 1977 was 29 billion pounds (Figure 22), and production of synthetic rubbers exceeded 5 billion pounds. This growth has been due not only to the development of new materials, but also to our steadily improving ability to couple materials, design, and processing so as to maximize properties and performance. In addition, plastics may offer manufacturing economies. One molded plastic part, for example, may replace a number of metal stampings at lower manufacturing cost and labor content.

Packaging, containers, and construction (siding and thermal insulation, for example) will remain major markets for plastics in the five years ahead. The auto industry is a growing market for plastics, which in part will be weight-saving replacements for steel. The plastics content of the typical U.S. car, currently about 200 pounds, may reach 300–350 pounds in 1985 models. High-performance polymers—engineering plastics, silicones, and specialty plastics—will remain the fastest-growing segment of the polymer industry. Already plastics are replacing die-cast aluminum and zinc because in some cases they offer better performance at lower cost.

New and modified polymeric materials will be developed to achieve specific combinations of properties. Processing methods will be improved as well. Demand is growing for plastics with greater resistance to heat, oxidation, and light and for plastics with better electric properties. A range of coatings and fire retardants is being used to ease problems with the flammability of plastics and the substances emitted by them upon combustion.

Petroleum will remain the principal source of raw

FIGURE 22 Growth of U.S. plastics production (1978 estimated 9 percent above 1977). (U.S. Plastics Production, *Modern Plastics*. January issue of years cited)



materials for polymers for many years, well beyond the time when its use as a fuel has begun to decline. Eventually, however, a shift is probable to other sources of carbon compounds, such as coal, shale oil, or renewable resources such as aquatic and terrestrial plants. For fossil sources of carbon compounds, the basic science and technology is largely in hand. For renewable resources, it is not.

CERAMICS AND OTHER INORGANIC MATERIALS

A major fraction of the materials and products used by an industrial society is inorganic and nonmetallic; ceramics are the best example of this class. Ceramics may be broadly defined as inorganic, nonmetallic materials processed or consolidated at high temperature. They are usually made from combinations of natural silicates, or compounds of silicon and oxygen with various metals, and oxides that are fused or sintered together. Ceramics include cement, bricks, tile, sanitary ware, china and dinnerware, glasses, porcelain enamel on metal, abrasives, and refractories. The technology of these materials evolves slowly, with progress relating principally to improvements in properties and production efficiency. There are, however, a number of scientific and technological challenges for the future, one example being the further development of the silicon carbide and nitride ceramics.

Silicon Ceramics

The silicon ceramics include silicon carbide, silicon nitride, and the SIALON's (compounds of silicon, aluminum, oxygen, and nitrogen). While these ceramics comprise one of the few classes of inorganic materials that do not occur in nature, recent research indicates that the chemistry of the silicon nitrides parallels that of the natural silicates. It should thus be possible to synthesize a large range of silicon nitride-type compounds with unique properties. For example, some of the silicon and nitrogen atoms in silicon nitride can be replaced by aluminum and oxygen, as in the SIALON's. The next five years will see intensive effort to synthesize such compounds and to delineate their thermochemistry and properties as well as their potential in uses such as high-temperature structural materials, optical and electronic devices, and refractories.

The silicon carbide and nitride ceramics are generally stronger and more stable than the normal oxide-type ceramics at higher temperatures (Figure 23). They resist corrosion, erosion, and thermal shock exceptionally well.

Intensive development indicates that the silicon carbide and nitride ceramics offer substantial promise

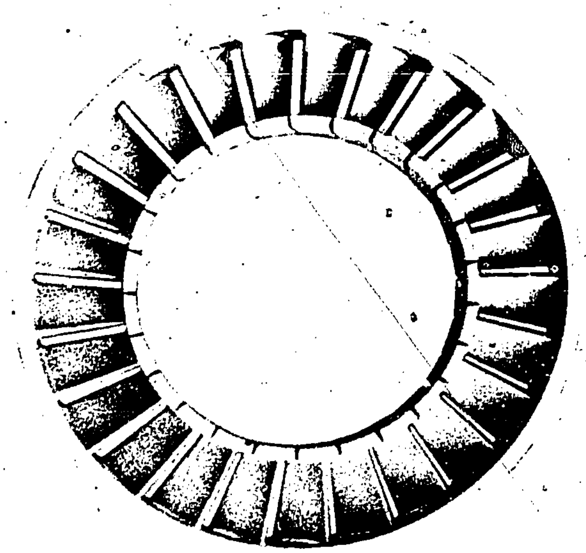


FIGURE 23 Turbine stator made of silicon nitride. (Ford Motor Company)

for replacing nickel- and cobalt-based superalloys in some high-temperature machinery, such as gas turbines, and would permit even higher operating temperatures than metals; the materials also are promising for use in ceramic heat exchangers. The higher operating temperatures that these materials would permit in energy-conversion processes would make possible higher efficiencies and savings in fuel. Success in this one area could yield far-reaching gains in power generation and in development of new generations of gas turbines. An all-ceramic turbine is under development as a power plant for automobiles, for example. Ceramics suffer from a tendency to fail unpredictably because of brittle fracture. Solutions to this problem depend in part on the development of improved, nondestructive means of detecting the minute flaws that lead to fracture. Reliability and predictability of service life are the key questions to be resolved if silicon carbide and nitride components are to be developed successfully. These materials and their development are an excellent example of the need for intimate interaction of materials, design, and processing in advancing the state of technology.

Optical Materials

New materials are being developed for military uses involving the acquisition and electronic processing of optical guidance and surveillance information. Two parallel needs—better sensors and better optical materials to protect them—comprise an extremely broad problem with nonmilitary implications in areas that include improved lighting and process-control equipment. Military equipment for acquiring optical information often must operate at high temperature and under severe mechanical pressures. Such equipment is limited at present by the need for optically transparent windows to protect the sensitive data-collection devices, or sensors, under these conditions.

COMPOSITE MATERIALS

A composite material generally consists of a matrix material through which a different, reinforcing material is distributed. The fiber composites originated with the glass-fiber reinforced plastics, in which the fibers have high strength, but relatively low stiffness. Glass-reinforced plastics have long been used in structural applications such as boat hulls, missile casings, and sporting equipment where a high strength-to-weight ratio is the predominant requirement.

In the 1960's, fibers of boron, graphite, silicon carbide, and other materials of high stiffness were developed. These newer fibers, used in both resin and metal matrices, promise fiber-composite structural materials of great strength and stiffness in a form capable of being fabricated into the most complex shapes. High-performance fiber composites already are used in aircraft. They have great potential for replacing metals in automobiles, where they may perform equivalent functions at weight savings of 50–70 percent. The key problem with high-performance composites in automobiles is cost. Much effort will be expended during the next five years to find new resin formulations and cheaper ways to prepare fibers and to develop methods of manufacturing composite parts in high volume at low cost.

RENEWABLE MATERIALS

Renewable resources are a potentially attractive source of materials, and a degree of attention is currently being devoted to the fabrication of new types of engineering or functional materials from resources like wood and paper. Certain wild plants also are being investigated as sources of materials. They include guayule, a source of rubber, and jojoba, a source of lubricating oil said to be the equivalent of sperm oil.[2,3] Both plants could be produced on desert scrublands. However, the economic feasibility of cultivating such plants as sources of materials is far from being established.

The use of renewable resources has certain broad limitations. Nonrenewable materials now account for some 90 percent of the tonnage of new supplies of nonfuel materials in this country. Attempts to reduce that percentage by shifting to renewable resources would begin at some point to create land-use conflicts with agriculture and other activities. Furthermore, intensive cultivation of renewable resources may require significant inputs of nonrenewable resources, such as fertilizers, pesticides, and mechanical energy.

BIOMEDICAL MATERIALS

Metals, polymers, ceramics, and composites should make further inroads in the next five years in replacing missing or defective limbs, joints, and other parts of the human body. The surgeon now has available materials of sufficient biocompatibility and durability for reasonably satisfactory use as soft-tissue implants in plastic surgery, hip and knuckle replacements, large artery replacements, hydrocephalus (cranial) drainage tubes, and implantable cardiac pacemakers. Materials for implantable teeth and bone sections and for replacing knee joints would appear to be nearing realization. This is not true, however, of materials for small-vessel (vein) prostheses and percutaneous (through the skin) electrical or fluid-conducting leads, as for constructing implantable lungs or kidneys.

Although substantial progress is being made in biomedical materials, they remain difficult to evaluate for use in humans. The problem would be eased by the development of laboratory tests that would predict the service life of implant materials more reliably than do present methods.

MATERIALS PROCESSING AND MANUFACTURING

Materials processing, broadly, is the conversion of raw materials into intermediate or finished products with useful shapes and properties. A distinction should be made, however, between primary processing and secondary processing or manufacturing. Examples of primary processing include the conversion of iron ore into steel (a bulk material) and mill processing of the steel into forms such as sheet and plate (engineering materials). Secondary processing includes the conversion of sheet, plate, or other semifinished steel into parts and finished products—machine tools, roller bearings, engines, aircraft.

PRIMARY MATERIALS PROCESSING

Productivity and product quality in the metals industries have been improved by important engineering advances during the past 50 years. Examples are continuous rolling of strip, automatic gauge (thickness) control, electronic inspection of bars and tubes, and computer control of rolling mills. The advent of the basic oxygen steelmaking process, the high-pressure blast furnace, and continuous casting initiated a revolution in steel productivity.

Full-scale adoption of technological advances in the primary processing of metals is extremely capital-intensive because of the scale of operations and the conditions and environment, such as high temperature and large mechanical forces, in which they must be conducted. The industry, therefore, approaches new technology cautiously and expects the corresponding new plant and equipment, once in place, to perform for a long time.

A number of forces lately have converged on primary metals processing to slow the rate of investment in new technology. Governmental regulations related to the environment and to occupational safety and health have required large capital investments that otherwise might have gone into production technology. As high-temperature processors, the industry has been affected significantly by the rising cost of energy and so has put major emphasis on energy conservation. Also, the growth of product-liability litigation has led to greater stress on nondestructive methods for detecting flaws in products. Shortages of raw materials like cobalt have led to price spirals in special metals. These forces have combined to depress the profitability and rate of innovation of most of the nation's basic metals-processing industries.

Improvement of Properties by Processing

The scientific basis of primary metals processing is sound in some areas, but less so in others. In melting and refining, a strong scientific base in thermodynamics, kinetics, and transport phenomena provides useful insights and guidance in processing operations. In deformation processing, such as rolling, the situation is less satisfactory. Macroscopic understanding of deformation processing is adequate as a guide to routine process planning and control, but a better scientific foundation is needed for future advances.

Metallurgists generally do not expect the discovery of broad new classes of alloys in the near term. Improvements in properties are more likely to come from learning to control and refine the structure of metals through more precise understanding and control of the steps in deformation processing. The high-strength steels mentioned earlier illustrate what can be achieved.

Improvement of the properties of metals through processing requires detailed knowledge of the degree of deformation and the temperature in all parts of the material in relation to elapsed time as the material deforms. This information must then be integrated with the kinetics of the development or change of internal structure in the material. Computer-based analytical methods for treating this problem are being developed.

In polymer processing, there are good possibilities for aromatic polyamides. Here, greater understanding of the structure and flow characteristics of the anisotropic fluid state and of the solidification of anisotropic fluids could lead to the production, through processing, of polymeric materials with stiffness comparable to that of metals.

SECONDARY MATERIALS PROCESSING

As we draw closer to the final manufactured product, we find a higher level of technological innovation in metals processing. Examples include casting and solidification, powder metallurgy, and microelectronics.

Casting and Solidification

In casting and solidification, great progress has been made, particularly in metals, by control of grain or crystal size and of the direction in which grains grow during cooling. This control is achieved by directional solidification: heat is removed selectively from the molten metal so that crystals grow in a given direction and parallel to each other to form a columnar structure. As a result, the boundaries of the grains in the metal are also approximately parallel, and the properties of the finished casting are significantly different in different directions. An even more recent advance is the casting and controlled solidification of single-crystal blades and vanes for aircraft gas turbines. As noted earlier, improved high-temperature durability is achieved in these materials by eliminating the grain boundaries entirely.

In a new process called rheocasting, the melt is cooled and agitated vigorously before pouring until it becomes mushy, with small crystals supported in the molten metal. The high solids content and lower casting temperature result in smaller, more uniform grains and denser, stronger parts. The lower temperatures involved also conserve energy, increase the life of dies, and make the entire process easier to control.

Other innovations in casting include the use of polystyrene patterns that vaporize when the molten metal engulfs them. These consumable patterns make it possible to produce complex shapes at low cost. Precision investment casting, in which the mold is first built up around a wax model of the part to be cast, has been refined to a high art, increasing both the size and quality of the castings that can be produced. Corresponding advances have been made in other foundry products through the use of molds of resin-bonded sand. This country leads the world in cast-iron technology.

Powder Metallurgy

Powder metallurgy—pressing and compacting of metal powders—has seen a revival with the advent of new consolidation methods and very fine metal powders, made by newly developed, rapid-solidification techniques. These powders are especially uniform in composition and microstructure and relatively free of otherwise embrittling microconstituents. The resulting products, after consolidation and heat treatment, have superior ductility and high-temperature properties. The new consolidation methods minimize porosity and thus yield solids of high density. With hot isostatic pressing, in which pressure is exerted uniformly from all directions, the most intractable metal or ceramic powders can be pressed into shapes with very nearly the desired final dimensions, minimizing finishing operations. A cheaper route to high-performance parts is provided by powder forging, in which preformed powder-based billets of about 20 percent porosity are forged to the final product. The new fine metal powders permit the production of metals sufficiently fine-grained to be superplastic—to undergo very large plastic deformation. Superplasticity, when built into metals like titanium that normally are difficult to form, allows them to be formed easily if relatively slowly. This property could be exploited to reduce the cost of making parts from such metals.

Microelectronics

A striking example of structure-oriented materials processing is microelectronics, particularly large-scale integrated circuits for computers. Probably in no other technology has processing been coupled so intimately with structure and has progress been so rapid and far-reaching. The early military and space applications of the technology moved quickly into the civilian economy. The transfer was possible because of the strong scientific base that already existed in solid-state chemistry and physics and because of the relatively modest investment required.

Computer-Aided Design and Manufacturing

Although this country is the birthplace of mass production, the vast majority of parts are made in numbers too small to justify major investment in assembly-line machinery. However, the manufacture of parts in relatively small runs is undergoing sweeping change. It began with the development of the numerically controlled machine tool in this country in the 1950's and accelerated as computers became more compact and inexpensive. Today, computer-aided design is exerting a major impact on manufacturing, and computer-aided manufacturing should begin to do so in the next five years.

Computer-aided design (CAD) employs powerful programs (software) for obtaining approximate solutions to problems inaccessible to precise mathematical analysis. CAD greatly extends the ability to analyze the design of a part for factors like stress and generation of heat. It permits parts to be designed with greater precision and reduces the likelihood of unforeseen failure. CAD equipment also can reduce designs to paper, bypassing the laborious efforts of draftsmen and thus increasing productivity in graphics and drafting.

An important aspect of CAD is that it is becoming possible to use its output directly, in combination with manufacturing data, to generate automatically the programs needed to optimize and automate the manufacture of the designed parts.

In full-scale use, computer-aided manufacturing (CAM) will be capable of providing: computer generation of the optimized production plans—selection of processes, equipment, tooling, operating conditions, etc.; optimized production control—dynamic scheduling of the work, maximizing and balancing the use of the manufacturing equipment, minimizing the time that parts in process lie waiting to be worked on, etc.; and automated machining of the parts by numerical control of machine tools. In numerical control, design data are combined with manufacturing data to produce a control program. The program is then used, via punched tape or small computer, to control one or more machine tools that produce the finished parts from raw stock.

The great capability of the computer is making it possible to design processing operations to a degree heretofore undreamed of. For example, precision forging to almost final shape, so as to minimize subsequent machining, is becoming possible with computer design of forging steps and dies. To achieve the full potential of the CAD/CAM approach in such operations, we must acquire new information on how materials flow, better understanding of friction and how to model it, and improved means of applying computer analysis to problems related to the plasticity of metals. Progress good enough to lead to widespread adoption of CAD/CAM methods in metalworking is possible in the next 5–10 years.

Full-scale use of CAD/CAM will stimulate advances not only in process modeling and design, but also in other areas: materials management and control, cost estimation, and inventory control. CAM also involves robotlike equipment, whose use to perform unpleasant or routine assembly or inspection operations is already a reality; current research is increasing this capability.

Laser Processing

The laser is an economical tool for manufacturing processes such as cutting, welding, and drilling. It can also be used to produce very fine-grained or amorphous surfaces by rapid melting and solidification—laser glazing. This technique is potentially effective because structural failures usually originate at a surface flaw. Laser modification of surfaces is being developed for valve seats and other components in the automobile industry. The use of lasers in metal processing can be expected to expand rapidly as the cost of the devices declines and as more engineers become aware of the unique capabilities of high-powered lasers.

Metal Removal

The development of high-speed metal removal processes—5,000–50,000 surface feet per minute—could result in major increases in productivity. With machine tools properly designed for high-speed cutting, it becomes practical to machine wrought-aluminum alloys at speeds considerably higher than are used today, since the rate of tool wear in machining these materials is not excessive up to speeds on the order of 12,000 surface feet per minute. However, to make such operations really economical, materials handling and tool changing must be automated so as to reduce noncutting time. With the constant improvement in cutting-tool materials that is taking place, practical speeds for machining cast iron and steel are evolving toward 2,000 to 3,000 surface feet per minute. Again, however, to make operation at such speeds economic, noncutting time must be reduced.

In addition, better understanding is needed of chip-segmentation mechanisms, tool-workpiece interface reactions, cutting-tool wear behavior, and lubrication effects. Such knowledge, coupled with a development program for high-speed machining equipment, could produce significant progress in materials processing in 5–10 years.

BARRIERS TO PROGRESS IN MATERIALS PROCESSING

While the United States continues to lead the world in computer technology, it is not at all clear that we will capitalize on this position to take the lead in computer-aided manufacturing. A stronger effort is needed throughout industry. The situation here contrasts sharply with that in West Germany and Japan, which have well-established national research programs in CAD/CAM with joint industry–government funding.

The requirement for large capital investment is inhibiting the full-scale adoption of new technology in materials processing in this country. For example, 60 percent of our machine-tool base is more than 10 years old, as compared with only 30 percent in Japan and West Germany. Moreover, the economic risk and long development time tend to inhibit sustained research and development in materials processing. An essential condition for real progress in materials processing and manufacturing is an economic climate that will prompt industry on its own to modernize, develop and adopt new technology, and improve its productivity.

Materials-processing technology also suffers from insufficient attention in our engineering colleges. Fewer than 10 percent of the materials faculty (who themselves comprise only a small fraction of the engineering faculty) are expert in materials processing and manufacturing. These fields do not enjoy the status accorded some other academic disciplines, and little current research in the schools is relevant to major developments in materials processing. The near absence in our universities of research in materials-processing and manufacturing technology denies the country a potential source of new ideas and innovation. Furthermore, it means that the universities are not exposing young people to current advances in the field.

RECENT CONCEPTS IN MATERIALS

The research and development and other materials activities of the past few decades have helped to crystallize two important concepts: the total materials cycle and materials science and engineering.[4] The materials cycle is a physical concept—materials flow from the earth through various useful forms and back to the earth in a closed cycle that is global in extent. Materials science and engineering is an intellectual concept—a coherent system of scientific and engineering disciplines that combines the search for insights into matter with the use of the resulting knowledge to satisfy society's needs for materials.

TOTAL MATERIALS CYCLE

The total materials cycle is driven by societal demand, and materials move within it in five stages:

- Extraction of raw materials: ores and minerals, rock, sand, timber, crude rubber.
- Processing of raw materials into bulk materials: metals, chemicals, cement, lumber, fibers, pulp, rubber, electronic crystals.

- Processing of bulk materials into engineering materials: alloys, ceramics and glass, dielectrics and semiconductors, plastics and rubbers, concrete, building board, paper, composites.
- Fabrication of engineering materials into structures, machines, devices, and other products.
- Recycling discarded materials or products to the system or returning them permanently to the earth.

The materials cycle provides a framework for dealing with a system of interacting parts. The flow of materials at a given point can be sensitive to economic, political, and social decisions made at other points. Materials shortages usually are found to be due not to worldwide scarcity, but to dislocations in the cycle that interfere with the arrival of materials at a given point in the usual amounts and at reasonable prices. A shortage may arise at one point, for example, because of inadequate processing capacity at another point; but countermeasures can be taken within the cycle, including stockpiling, recycling, and substitution of one material for another.

Materials, energy, and the environment interact strongly at virtually every point in the materials cycle. About one-half of the energy consumed by all manufacturing industries in the United States goes into the value added to materials in producing and fabricating them to the point of becoming engineering materials.[5] But materials also are crucial to making energy available in the first place. In fact, inadequacies in the performance of materials currently are the primary constraint on the efficiency, reliability and safety, cost-effectiveness, or in some cases actual realization of our advanced energy-conversion technologies—gas turbines, nuclear reactors, high energy-density batteries, fuel cells, magnetohydrodynamic generators, coal conversion, and solar-energy conversion.

The materials cycle offers exciting opportunities in materials research and development. Objectives for each stage of the cycle might be:

- *Extraction and processing* Reduce energy consumption and pollution
- *Fabrication* Designs that improve efficiency of construction
- *Assembly.* Designs that improve recyclability
- *Operating life* Greater durability
- *Recycle* Designs that ease disassembly and sorting

In this manner, functional design could be coordinated with materials selection and the other operations of the materials cycle so as to generally facilitate recycling, substitution, and conservation.

MATERIALS SCIENCE AND ENGINEERING

The central purposes of materials science and engineering (MSE) (Figure 24) are to probe the relationships of the internal structure and composition of materials to their properties and performance; and to use the resulting knowledge in producing, shaping, and otherwise processing materials so as to control their properties and achieve the desired performance in the finished product. MSE links fundamental understanding of the behavior of electrons, atoms, and molecules to the performance of products.

It is remarkable that almost all of the technical advances in metals and alloys described earlier were based on science developed in the nineteenth century (classical thermodynamics—especially as elaborated by J. W. Gibbs—and chemistry) in conjunction with some more recently developed experimental techniques (also based largely on old science). It may be that these applications awaited the formation of a large cadre of technologists with enough training to understand thoroughly, and so to apply effectively, the old science.

MSE promotes the application of basic science to the development, processing, and use of materials by establishing a two-way flow of information between the basic scientists at one extreme and the user/consumer of materials at the other. We have been accumulating empirical knowledge of materials for a long time. But only in this century—and at an accelerating pace during the past few decades—have scientists made substantial progress in acquiring the corresponding basic insights. The microstructure of materials has been revealed by optical microscopy, substructure by electron microscopy, crystal and molecular structure by X-ray diffraction, atomic structure by various spectroscopies, and nuclear structure by high-energy atom smashers.

With this new knowledge it has become possible to

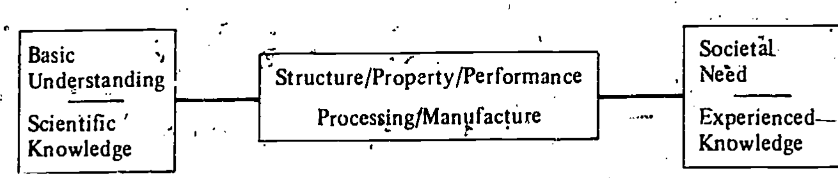


FIGURE 24 Materials science and engineering. Scientific knowledge flows to the right, while experienced knowledge flows to the left.

exploit the linkage of structure, properties, and performance. The strength and dimensional stability of polymers, for example, can be upgraded through methods of synthesis that yield highly ordered molecules that cluster into crystalline arrays. Transistors are made by manipulating the electronic structure of silicon and other semiconductors; they are produced on a large scale by methods that achieve exceptionally precise control of composition and internal structure.

Studies of the MSE process at work indicate that the two-way flow of information is most productive when basic understanding of a materials problem and the empirical need to solve it are mixed so intimately that it becomes difficult to tell which provided the initial impetus toward a solution. In the main, however, the initial impetus in the MSE process seems to arise more often from "technological pull" than from "scientific push."

The successes of the MSE approach should not be construed to mean that properties can necessarily be predicted from structure alone, nor performance from properties alone, except in a general sense. As a rule, the structure–properties–performance linkage must be worked out through the reciprocal flow of information—scientific and empirical—that characterizes the field of materials science and engineering.

NEAR-TERM ISSUES IN MATERIALS

The basic concerns of the United States about materials supply and demand have changed very little from decade to decade. They commonly involve adequacy of supply, prices, and national security. Long-term world supplies of the basic resources rarely have stirred alarm. Possible shortages of more exotic or less abundant materials have been recognized, but not as serious threats to world health and welfare. Concerns about supply and demand will persist, but any consideration of the immediate future for materials must take account of certain newer issues.

THE MOVE OVERSEAS

In basic materials, certain locations overseas offer advantages over this country, including richer ores, cheaper energy, cheaper labor, and in some instances tax advantages and readier access to capital. Regulation of pollution and workplace health and safety is generally less strict than in the United States, although it will not necessarily remain so. Except for copper, many of the overseas resources are owned and used by U.S. corporations, which are importing into the American market as well as into Europe and Japan. Overseas processing is expanding, and trends will develop toward primary processing of shapes at the source. The coming decade could see significant reductions in the domestic capacity of U.S. metals-producing industries relative to domestic demand. American companies also are manufacturing components and finished products abroad and importing them into the United States.

Companies in the United States and other nations will tend to shift basic materials-processing operations abroad in the years to come. The incentive in part is the hope of assuring supplies by strengthening local economies and thus cementing relations with governments that control basic resources. The beginning of this trend can be seen already. Our imports of aluminum increased 310,000 tons, or more than 50 percent, during 1976–77; imports of refined copper rose 218,000 tons, or more than 140 percent; and imports of iron and steel rose 5.5 million tons, or 44 percent.[6] During the same period, our imports of raw and processed materials increased in value to $20 billion annually, and the excess of imports over exports increased from close to zero to $5 billion.[7]

There is a world surplus of aluminum, copper, and iron. Companies that use those materials naturally seek them at the lowest price. The consequent rise in imports has several negative impacts. Employment in the primary aluminum industry in this country declined by 1,000 during 1975–77, in the primary copper industry by 5,000, and in the primary iron and steel industry by 80,000. These declines in employment have become a serious problem and led to the President's Nonfuel Minerals Policy Study. The study is being conducted jointly by the White House Office of Science and Technology Policy and the Department of the Interior, and the findings are due in the fall of 1979. Other negative impacts of rising materials imports include pressure on the balance of payments and uncertainties about reliability of supplies.

Loss of Self-Sufficiency

A new element in the situation is the gradual decline in the nation's ability to supply itself from domestic resources (Table 4). The United States historically has been accustomed to a large measure of self-sufficiency. However, most nations rely on resources abroad, and national economies clearly can thrive on imported materials. Notable examples are West Germany and Japan.

IMPACT OF REGULATION

Federal regulation has worthy social objectives, but it should be systematically reexamined to ensure that a reasonable balance is struck between costs and benefits. The problem is analytically difficult. None-

TABLE 4 U.S. Net Imports of Selected Metals and Minerals as a Percent of Apparent Consumption[a]

Minerals and Metals	1950	1955	1960	1965	1970	1973	1974	1975	1976	1977[b]	Major Foreign Sources[c] (1973 to 1976)
Columbium	100	100	100	100	100	100	100	100	100	100	Brazil, Thailand, Nigeria, Malaysia
Mica (sheet)	98	95	94	94	100	100	100	100	100	100	India, Brazil, Madagascar
Strontium	100	98	100	100	100	100	100	100	100	100	Mexico, Spain
Manganese	77	79	89	94	95	98	98	98	98	98	Brazil, Gabon, South Africa
Cobalt	90	68	66	92	98	98	99	98	98	97	Zaire, Belgium and Luxembourg, Norway, Finland
Tantalum	99	100	94	95	96	87	87	81	96	97	Thailand, Canada, Australia, Brazil
Platinum group metals	74	91	82	87	78	87	87	83	90	92	South Africa, U.S.S.R., United Kingdom
Bauxite and alumina	55	73	74	85	88	92	92	91	91	91	Jamaica, Australia, Surinam, Guinea
Chromium	95	83	85	92	89	91	90	91	89	89	South Africa, U.S.S.R., Turkey, Rhodesia
Tin	82	80	82	80	81	84	84	84	85	86	Malaysia, Thailand, Bolivia, Indonesia
Asbestos	94	94	94	85	83	82	87	82	85	85	Canada, South Africa
Fluorine	33	55	48	77	80	79	81	85	79	80	Mexico, Spain, Italy, South Africa
Nickel	90	84	72	73	71	69	72	72	70	70	Canada, Norway, New Caledonia, Dominican Republic
Potassium	9	0	E	7	42	53	58	51	61	66	Canada, Israel
Gold	25	34	56	72	59	48	63	52	76	60	Canada, Switzerland, U.S.S.R.
Zinc	41	51	46	53	54	64	59	61	59	58	Canada, Mexico, Australia, Peru
Antimony	33	32	43	36	40	50	44	49	54	52	South Africa, People's Republic of China, Bolivia
Cadmium	17	20	13	20	7	41	46	41	64	51	Canada, Australia, Belgium and Luxembourg
Selenium	53	18	25	44	11	57	59	66	59	47	Canada, Japan, Mexico, Yugoslavia
Mercury	87	20	25	49	41	78	86	69	62	46	Spain, Algeria, Mexico, Yugoslavia
Silver	66	58	43	16	26	66	55	30	50	42	Canada, Mexico, Peru, United Kingdom
Barium	8	25	45	46	45	37	38	32	42	40	Peru, Ireland, Mexico
Tungsten	80	NA	32	57	50	66	68	55	54	38	Canada, Bolivia, Peru, Thailand
Titanium (ilmenite)	33	40	22	9	24	28	33	25	29	38	Canada, Australia
Vanadium	4	E	E	15	21	43	36	38	37	37	South Africa, Chile, U.S.S.R.
Gypsum	28	27	35	37	39	35	37	34	35	35	Canada, Mexico, Jamaica, Dominican Republic
Iron ore	11	18	18	32	30	35	37	30	29	33	Canada, Venezuela, Brazil, Liberia
Copper	31	17	E	15	E	8	20	E	12	17	Canada, Chile, Peru, Zambia
Lead	40	39	33	31	22	29	19	11	15	14	Canada, Peru, Mexico, Australia
Iron and steel products	E	E	0	7	4	10	7	9	7	13	Japan, Europe, Canada
Salt	E	E	2	5	6	6	7	4	7	8	Canada, Bahamas, Mexico, Netherlands Antilles
Aluminum	17	E	E	4	E	18	4	E	9	8	Canada
Pumice and volcanic cinder	3	2	3	5	11	8	7	4	2	5	Greece, Italy
Cement	E	1	0	3	3	7	4	5	4	4	Canada, Spain, Norway, Bahamas
Iron and steel scrap	2	−14	−24	−17	−25	−21	−19	−27	−22	−11	—

E = net exports; NA = not available.

[a] Apparent consumption equals the U.S. primary plus the secondary production plus net imports. Based on net imports (imports minus exports plus or minus Government stockpile and industry stock changes of metals, minerals, ores, and concentrates.

[b] Estimate.

[c] Major foreign sources listed in descending order of amount supplied.

SOURCES: *Minerals Yearbook*, U.S. Bureau of Mines, various years; import and export data from U.S. Bureau of the Census. (From *Mining and Minerals Policy, 1977 Annual Report of the Secretary of the Interior Under the Mining and Minerals Policy Act of 1970*, p. 60. Washington, Government Printing Office, 1977.)

theless, there is preliminary evidence suggesting that the cumulative impact of regulation is quite significant for some industries.[8]

ENERGY

The recent rise in the cost of energy is another new element in the materials situation. It imposes a new constraint on our ability to make economic gains merely by substituting low-cost external energy for human energy. This constraint is particularly troublesome in materials. The materials industries tend to be energy-intensive, and the declining quality of resources, the need to control pollution, and the adoption of new, high technology all call for large inputs of energy. For example, to refine U.S. copper sulfide ores of 0.8 percent copper content requires about 20 percent more energy per ton of copper produced than to refine a Chilean copper sulfide ore of 1.5 percent copper content.[9]

Because of the cost of energy, technology that reduces its consumption is emerging in the materials industries. Examples include the chloride cell for refining aluminum, hydrometallurgical refining of copper, and electric furnace/scrap recycle for making steel. Materials fabricators and product manufacturers are selecting materials and processes so as to reduce energy consumption and increase productivity. In some instances, the materials and applications are new. Automobile makers, for example, are using more plastics, aluminum, and high-strength steels to reduce the weight of their cars.

CONSERVATION

Conservation of materials may be defined as the reduction of losses from the materials cycle or of materials going through the cycle. Losses for common metals have been estimated to range from half to three-quarters of the amounts that enter the cycle.[10] Materials are lost at all stages of the materials cycle, from mining and production (tailings, slags), through product use (wear, corrosion), to ultimate disposal. Of the means available for reducing such losses, two of the potentially most effective are recycling and substitution.

Recycling

Recycling of materials, where it reduces costs, will grow in the next five years. Cost advantages will be evaluated on the basis of total-cycle costs. Manufacturers already recycle significant amounts of in-plant scrap. Industry also recycles large amounts of materials from discarded products, including automobiles and telephones. In addition, worn but intact products are overhauled and reused extensively where the practice is profitable, a good example being automobile parts. Recycling of products, as opposed to the materials they contain, offers marked potential for conserving materials, recovering the energy invested originally in manufacturing the products, and reducing the pollution resulting from manufacturing.

The Resource Conservation and Recovery Act of 1976 is designed particularly to stimulate the recovery of materials and energy from municipal waste. Recycling of municipal wastes will be stimulated by technology that leads to design for recycle. It is essential, for example, that recyclable material be easily identified. The point is demonstrated by the relatively high recycle rate of the all-aluminum beverage can. Stimulated by high resale value, collections of cans in 1977 totaled a record 6 billion, about one in every four sold.

The concept of a waste dump as a man-made ore body is intriguing.[11] Most municipal waste is a poor source of iron and aluminum—it is leaner in those materials than are useful ores. However, such waste may be a better source of other metals, once the burnable constituents are consumed and the iron, aluminum, and glass removed. Work has been done on recovery of materials from incinerator residue, and the approach warrants greater consideration.

The level of recycling to be expected in the next five years will depend on a number of factors, of which the most important, as implied earlier, is cost. Old scrap metal and wastepaper, for example, must compete with in-plant wastes of known quality. The reworking of discarded products is labor-intensive and may offer a relatively low rate of return. We have much of the technology required to recover materials from municipal waste, but problems remain in separating these materials—sorting glass by color, for example—so they can be recycled in the most useful forms. Difficulties with institutional arrangements and waste collection must also be resolved.

A most important problem requiring research and development during the next 5–10 years is the recycling of plastic scrap (Figure 25). Landfill is rapidly becoming expensive and unavailable, so the scrap must be disposed of in some other way. Processes are needed to separate waste plastics satisfactorily for reuse or conversion to chemical materials or, alternatively, to convert them pyrolytically—at high temperature without burning—to a petroleum equivalent for use as fuel.

Substitution

Substitution of one material for another historically has been motivated by cost reduction, specific functional advantages, or supply considerations. Economic

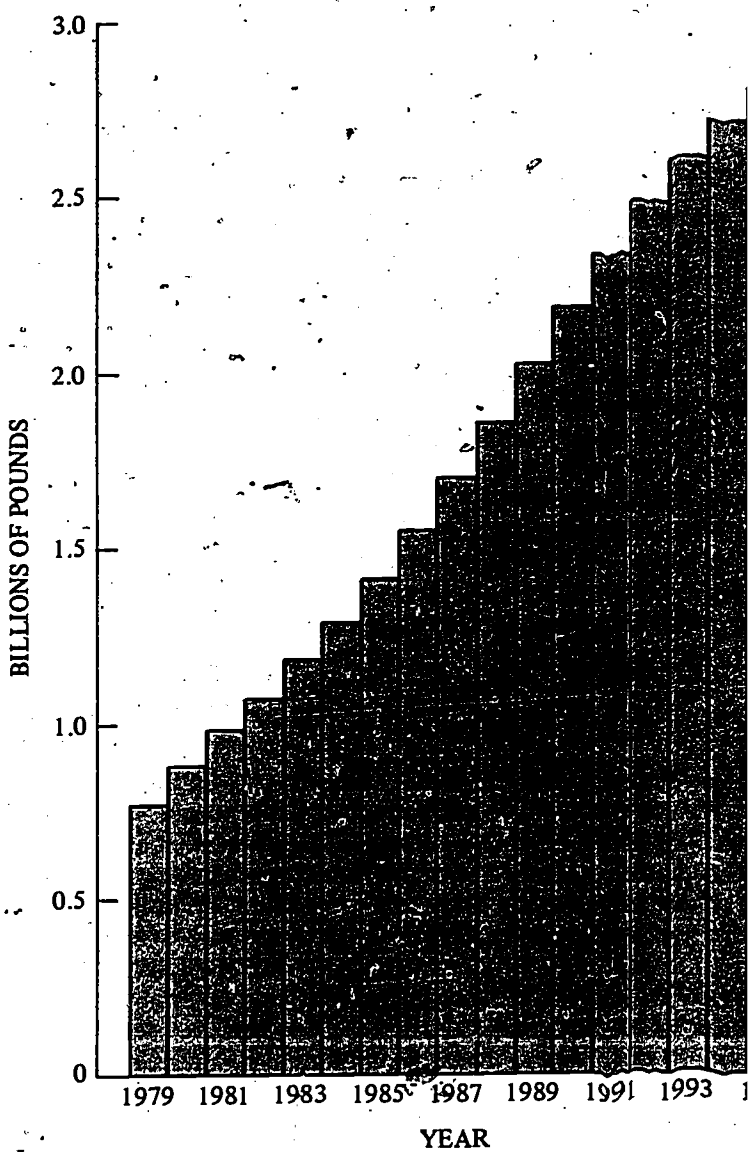


FIGURE 25 Projection of scrap plastics generated annually from junk cars. (L. R. Mahoney, J. Braslaw, and J. J. Harwood, "The Effect of Changing Automobile Materials on the Junk Car of the Future," paper 790299. Warrendale, Penn., Society of Automotive Engineers, 1979.)

incentives certainly have been a main driving force. The rapidly increasing substitution of plastics for metals and glass in many consumer and industrial applications provides a model for characterizing the substitution process.

In general, substitution tends to be an evolutionary process. It results not only from the advent of new materials, but also from improvements and tailoring of older materials for specific uses. Indeed, substitutions stimulate competitive development and innovation in the materials they threaten.

The chief characteristic of a sound climate for materials substitution is a stockpile of materials technology to draw on. In all phases of the materials cycle, research and development generate the knowledge required to expand the range of alternatives and options for materials and process selection proced
In that vein, it is important to restress the ea
observation that changes in any one phase of
materials cycle have impacts on the entire c
Major shifts in minerals or other materials mus
meshed with capacity aspects of other phases of
cycle and can have, for example, major conseque
in the recycling phase.

POLITICAL AND INSTITUTIONAL FACTORS

Shifts in the flow of mineral supplies can stem f
political and institutional changes as much as f
exhaustion of reserves. World consumption of in
trial raw materials has been growing at record r
since World War II, with the United States joi

Western Europe and Japan in their increasing dependence on imports. Canada, Australia, South Africa, and the Soviet Union remain large suppliers of world markets, but a large proportion of the remaining higher-grade mineral resources is located in the developing countries of Africa and Latin America.

Coupled with the actual or impending shifts of the centers of production has been the disappearance of the traditional forms of long-term mineral concessions granted to companies based in the United States or Europe. The arrangements that have replaced those concessions are not necessarily stable. In addition, nations have claimed unrestricted sovereignty over the development of their resources, including the right to revise conditions unilaterally without penalty or recourse. Developing countries aspire to a new economic order; they wish to improve the terms of trade for primary products through joint actions among suppliers or through commodity agreements executed by both producers and consumers.

Technology Transfer

Technology is viewed widely by the developing nations as a key to industrial development. Their goal has been to gain greater access to and control over technology so they can play a larger role in the development of their materials and minerals resources and manufacture of products. Technology transfer has become a point of serious contention between the industrialized and developing nations. Our industry and our government are being pressed continually to establish new policies and programs involving the transfer of technology. Industry is encountering increasing difficulties in maintaining established private and public technology-transfer processes in the light of the new governmental systems, codes, and institutions proposed by the developing nations.

Where these changes in political and institutional arrangements will lead the world in the longer term remains uncertain. But it is apparent that they create a potential for serious dislocation of the orderly process of exploration and the development of new reserves.

DIFFICULTIES OF PROJECTION

We noted at the outset that it is difficult to project future materials needs, and the light metals provide classic examples of the perils of doing so. Aluminum, titanium, magnesium, and beryllium were each at one time or another touted as a "metal of the future." But only aluminum has become a large-volume, low-cost metal. Titanium and magnesium have become more important as metals, but the largest tonnages of them are sold in nonmetallic form: titanium dioxide as a pigment and magnesium oxide as a refractory material. Beryllium has remained an expensive specialty metal used in alloys and nuclear reactors.

The patterns of supply for these metals have also been diverse. Aluminum's emergence as a major metal depended totally on one ore—bauxite—and the metal could be produced only so long as natural cryolite for the refining process was available from Greenland. Today we have every reason to expect that clays and anorthosite eventually could supplant bauxite, and we have synthetic cryolite as well. Titanium is an abundant element, available commercially in two mineral forms. Magnesium is produced primarily from seawater rather than from terrestrial sources. The imported beryllium ore, beryl, has been augmented by a domestic mineral that was not even known as a commercial source of beryllium 30 years ago.

The light metals also illustrate the many potential paths that the materials cycle provides for meeting societal needs. The economic goal is not to supply society with aluminum, for example, but with a lightweight, low-cost material with excellent corrosion resistance and other useful properties. Rarely is a material uniquely able to meet a given functional need. The question, rather, is which material best meets the need at a given time with due regard to the entire cycle from extraction to recycle or disposal.

MATERIALS AND THE NATION

We must conclude that developed resources always appear to be so limited that concern is ever present for their depletion or exhaustion. But our total unexploited physical endowment remains large contrasted with current needs. Ingenuity remains the driving force required to transform these resources from crustal promises into economic realities. Ingenuity must be aided, however, by better information, careful analysis, innovative technology, and functioning political and economic institutions attuned to support technology, innovation, and industrial productivity. Universities, industry, and the government must all have the perception to support materials research and development to the extent necessary to turn basic ingenuity to best advantage.

Exploration and development of mineral and materials resources are no longer left solely to private industry. Government has intervened because of concern about accessibility of foreign supplies and the assurance of continued domestic sources and because minerals development has large potential environmental impacts to which the public is increasingly sensitive. The importance of our basic minerals and materials industries to the economic welfare and security of the United States makes the creation of a healthy climate—social, political, economic—for these industries a matter of public policy that warrants serious attention. The issue of continued investment in minerals and materials development both domestically and abroad will be more critical during the next five years than the question of physical adequacy of resources.

Continuing concerns about materials availability, supply, and costs have focused new attention on the role of materials in industrial operations and in national affairs. The recognized interaction and interdependence of materials, energy, and the environment are a needed catalyst for the understanding of the pervasive force of materials technology throughout our world. Certainly, the trends in materials supply, availability, and costs will bring additional pressures and intensify others. Research and development programs on materials, with emphasis on conservation, recycling, substitution, and the management of materials, can provide opportunities to offset some of these pressures. The systems approach to materials, in the context of a materials cycle, and research on materials processing and manufacturing may be key elements in the response of research and development to the materials–resources challenges of the future.

OUTLOOK

The following outlook section on materials is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

New pressures and issues have begun to affect the outlook for materials. Energy, environment, and transportation have become matters of national priority. New governmental agencies are involved in research and development and in regulatory activities in these fields. Uncertainties about industrial productivity and innovation and the business climate for risk-taking permeate materials technology and other fields. Decisions based on relative and near-term payoffs are shaping the research environment.

The health of materials processing and manufacturing technology in the United States is arousing concern. This is particularly so with respect to foreign competition and trends toward moving basic materials operations overseas to take advantage of richer ores, cheaper energy and labor, and less stringent environmental requirements.

In short, the forces that mold the materials enterprise are shifting to an uncommon degree from the purely technological to social and economic factors.

DEVELOPMENTS IN MATERIALS

It is difficult to forecast materials needs either nationally or internationally. For this country, however, it can be said that the needs will increase incrementally in quantity but that the kinds of materials used will not change much during the next 5–10 years. Subsequently, both the national origins and the kinds and amounts of the materials used could change markedly. New technologies change both what is usable and how it is used.

The rapid evolution of new materials that began during World War II is still under way. Research and development has become more selective in recent years, however, and business factors are inhibiting the movement of new methods and products from the laboratory into everyday use.

METALS AND ALLOYS

Striking advances cannot be expected on a broad front in basic metals such as iron and steel, aluminum, and copper. More likely are incremental improvements in specific properties, reduction in cost through innovations in processing, and better tailoring of properties to meet specific needs.

Unique high-temperature strength and performance can be achieved in cobalt- and nickel-base superalloys by directional solidification of the molten alloy. Further improvements in high-temperature behavior have been demonstrated with single-crystal turbine blades and vanes of such alloys. Both types of materials will be demonstrated in gas-turbine engines within the next five years. They will permit longer performance life and perhaps higher peak operating temperatures and consequent higher efficiencies in the engines. Rapidly solidified superalloys, now under active development, hold even greater potential for the longer term.

Certain alloys, when cooled at very high rates, solidify in noncrystalline, or amorphous, form, as opposed to the crystalline structure found normally in metals. The properties of these amorphous metals make them good candidates for use in power-transmission transformer cores, where they could yield significant savings in energy. Amorphous metals have good prospects for commercial use in the next five years in other applications, such as magnetic shielding.

ENERGY- AND INFORMATION-RELATED MATERIALS

The cost of producing solar cells for converting sunlight directly to electricity should drop significantly in the next five years. However, solar cells are unlikely to become a major source of electrical power during this century.

Infrared detectors made of semiconducting materials will allow us routinely to "see" objects by the heat they emit, even in total darkness. Potential uses range from detecting tumors to locating sources of heat leakages from buildings and industrial operations.

Communication by fiber-optic transmission will see major growth in the next five years. Light signals from light-emitting diodes or lasers already can be transmitted several miles through glass or quartz fibers and detected by silicon-based devices. The main advantage of fiber-optic equipment over conventional telephone lines is its very high message capacity.

The number of electronic components on a single integrated-circuit silicon chip will continue to increase in the next five years, and the cost per component will continue to fall. New processing methods will provide increasingly complex circuits at lower cost.

The so-called bubble memories for computers, based on magnetic materials like gadolinium–iron–garnet, are being developed rapidly. The materials are difficult to produce but they offer opportunities for low-cost, highly stable, mass memories.

POLYMERS

Production of synthetic polymers—plastics and rubbers—will continue to grow rapidly, with emphasis on the development of new materials to achieve specific combinations of properties. The already large market for polymers in automobiles is likely to grow 50 percent in the next five years as manufacturers substitute plastics for metals to reduce the weight and thus the fuel consumption of cars. Petroleum will remain the principal source of raw materials for polymers well beyond the time when its use as a fuel has begun to decline. In the longer run, however, a shift is likely to other sources of carbon compounds, such as coal, shale oil, and, ultimately, renewable resources.

CERAMICS

Silicon nitride and carbide ceramics offer promise for replacing superalloys in gas turbines and also may be used in ceramic heat exchangers. In such uses these materials permit higher operating temperatures than metals and thus even higher efficiencies than now are common in energy-conversion devices. The major problem to be solved with these ceramics is their tendency to fail unpredictably by brittle fracture, but much progress should be made in the next five years in improving their properties and reliability.

Recent research indicates that it should be possible to synthesize a large range of silicon nitride-type compounds with unique properties. The next five years will see intensive studies of such materials. Potential uses include high-temperature structural components, optical and electronic devices, and refractories.

COMPOSITE MATERIALS

Considerable progress should be made with high-performance composites—resin and metal matrices reinforced with fibers of boron, graphite, and other materials of high stiffness. Such composites have great potential for aircraft, where they are already used to some extent. They are beginning to be used in automobiles to save weight.

RENEWABLE MATERIALS

Plants that could be raised on scrublands and in the coastal oceans are being studied as sources of materials such as rubber and lubricating oil. The economic feasibility of this approach is far from established, however. Also, a strong shift to renewable resources would begin at some point to create land-use conflicts.

MATERIALS PROCESSING AND MANUFACTURING

In the processing of metals, great progress has been made in casting and directional solidification. Precision investment casting has been raised to a high art, and corresponding advances have been made in other foundry techniques. This country leads the world in cast-iron technology.

The properties of metals can be improved by modifying their internal structures through more precise understanding and control of the steps in deformation processing. An example of what can be achieved is the high-strength microalloy steels produced by controlling the conditions of rolling and other processing operations. Automobiles of the 1985 model year may contain as much as 500 pounds of these weight-saving steels per car.

Powder metallurgy is seeing a revival because of new consolidation techniques that produce parts of high density and nearly the desired shape, thus minimizing machining and scrap generation. The cost of fabricating high-performance structural parts will be reduced by compressing very fine metal powders, made by new techniques, in large, hot isostatic presses. A more recent development is atomized powders that are especially uniform in composition and microstructure. These newly available powders permit the production of metals that are superplastic and thus more readily formable than they would be normally.

Laser treatment of surfaces to improve the wear-resistance of metal parts can be expected to expand rapidly as the cost of lasers decreases. The technique is being developed for automotive valve seats and other components.

The development of high-speed metal removal processes promises major increases in manufacturing productivity. Basic studies, coupled with a development program on high-speed machining equipment, could produce significant progress in materials processing in 5–10 years.

COMPUTER-AIDED DESIGN AND MANUFACTURING

The manufacture of parts in relatively small numbers is undergoing sweeping changes based on computer-aided design (CAD) and computer-aided manufacturing (CAM). Full development is yet to come, especially with CAM. An important aspect of CAD is that it is becoming possible to use the output of that process directly, in combination with manufacturing data, to generate automatically the programs needed to optimize and automate the manufacture of the designed parts by CAM.

NEAR-TERM ISSUES IN MATERIALS

The political and institutional factors that affect the worldwide flow of raw materials have been changing markedly. The long-term results are uncertain. It is apparent, however, that these changes potentially could seriously disrupt the exploration and development of new reserves.

American and other companies will tend to shift basic materials processing operations overseas in the next decade. The aim in part is to assure supply by improving relations with governments that control basic resources. Also, certain locations overseas offer advantages that include richer ores and cheaper energy and labor. This trend can have negative impacts: loss of jobs in the domestic materials industries, greater pressure on the balance of payments, and uncertainties in deliveries of materials.

The rising cost of energy could be especially troublesome to the materials industries, which tend to be energy-intensive. Also, large inputs of energy are required to offset the declining quality of domestic ores, to control pollution, and to use new, high technology. In some cases the materials industries are easing the problem by adopting technologies that reduce energy consumption. Examples include the chloride cell for refining aluminum and hydrometallurgical refining of copper.

CONSERVATION, RECYCLING, SUBSTITUTION

Conservation of materials may be defined as the reduction of the amounts of materials that flow through the materials cycle or the reduction of losses from the cycle. Such losses occur in all stages of the cycle, from mining and production (tailings, slags), through product use (wear, corrosion), to ultimate disposal. Of the available means of conservation, two of the potentially most effective are recycling and substitution.

Recycling of materials, where it reduces costs, will grow in the next five years. Industry already recycles large amounts of in-plant scrap and of materials from discarded products, including automobiles and telephones. Various products are now recycled by overhaul and reuse. This practice offers marked potential for conserving materials, recovering the energy invested originally in manufacturing the products, and reducing the pollution resulting from manufacturing.

Substitution of one material for another historically has been motivated by cost reduction, specific functional advantages, or supply considerations. The chief characteristic of a sound climate for substitution is a stockpile of materials technology to draw on. The rapidly growing substitution of plastics for metals and glass in many consumer and industrial applications provides a model for characterizing the substitution process.

MATERIALS AND THE NATION

We must continue to examine the physical adequacy of resources and potential shifts among materials through time, but crises do not seem imminent in the near future. A more important issue may be the need to create an atmosphere—economic, social, and political—that will assure a strong future for the country in minerals and materials resources and a competitive industrial base in a rapidly changing materials world.

REFERENCES

1. Norton, E., Electric Power Research Institute, private communication.

2. *Guayule: An Alternative Source of Natural Rubber* (NRC Board on Science and Technology for International Development). Washington, D.C.: National Academy of Sciences, 1977.

3. *Products from Jojoba: A Promising New Crop for Arid Lands* (NRC Committee on Jojoba Utilization). Washington, D.C.: National Academy of Sciences, 1975.

4. *Materials and Man's Needs* (NRC Committee on the Survey of Materials Science and Engineering). Washington, D.C.: National Academy of Sciences, 1974.

5. Pick, H.J., and P.E. Becker. Direct and Indirect Uses of Energy and Materials in Engineering and Construction *Applied Energy* 1:31–51, 1975.

6. *Mineral Commodity Summaries 1978*. Washington, D.C.: U.S. Bureau of Mines, pp. 4, 46, 84, 1978.

7. Andrus, C.D. *Annual Report of the Secretary of the Interior under the Mining and Minerals Policy Act of 1970*. Washington, D.C., 1978, p. 32.

8. *Proceedings*, Council of Economics—AIME, 1978. Denver, Colo.

9. Page, N.J., and S.C. Creasey. Ore Grade, Metal Production, and Energy. *Journal of Research—U.S. Geological Survey* 3:9, 1975.

10. *Metal Losses and Conservation Options in Processing, Design, Manufacture, Use, and Disposal of Products*. Office of Technology Assessment, Congress of the United States, draft report, February 1979.

11. Blum, S.L. Tapping Resources in Municipal Solid Waste. *Science* 191(4227):669–675, 1976.

III SCIENCE AND THE UNITED STATES

7 • Demography

INTRODUCTION

DEFINITION, METHODS, AND TOOLS

Demography is the quantitative study of human populations. Its basic materials are derived from censuses, vital statistics, and sample surveys. Its methods are observational, empirical, and statistical. Its techniques include the use of advanced mathematics to refine raw, often incomplete data into interpretable trends and comparisons.

The field can be divided roughly into formal demography and the broader area of social demography, or population studies. Formal demography is concerned principally with analysis of the three major determinants of population change: births, deaths, and migration. Social demography studies population change in social, economic, and physical settings, and, therefore, considers other variables such as occupational structure according to sex and age, income levels for different population groups, and housing patterns.[1]

DEMOGRAPHIC DETERMINANTS

Demographic patterns such as those of fertility, marriage, and migration originate in millions of individual actions. These in turn are triggered by such factors as individual aspirations, education, kinship, personal finances, and health.

The individual actions reflected in demographic statistics affect the economic and social environments in which people spend their lives. A sharp increase in the birthrate, for example, concentrates large numbers at the critical thresholds of the life cycle—entering and leaving school, entering the job market, establishing families, renting and buying houses, and eventually retiring and living on income from sources other than current earnings.

Employment and business conditions may affect the age of marriage and of childbearing. Changing lifestyles affect family structures. The desire for independence by the elderly, along with the provision of some basic financial security, has in part stimulated the growth of one-person households.

Changing industrial and transportation technology, climate, tax, and labor laws, and business practices are among the major factors affecting where people live. New settlement patterns, in turn, have created new demands for public and private services and have led to competition among cities, states, and regions—between the "sunbelt" and the "frostbelt," for example, and between central cities and suburbs. A recent and unprecedented reversal of migration patterns away from large central cities is forcing reassessment of community planning and affecting other economic, social, and governmental activities.

Demographic trends are composites of individual actions, and national trends are composites of regional and local events. The national trend toward empty classrooms and unemployed elementary and high school teachers also subsumes rapidly growing communities and regions with increased loads on schools and other public services for the young.

Similarly, improvement in some of the highly desired aspects of life have not been equally distributed. The differences between men and women in expected longevity have widened, although life expectancy for both men and women has increased.

Finally, the form in which information is presented can lead to different interpretations. It may matter, for example, whether actual numbers of people or percentages are discussed. For example, between March 1975 and March 1976 more people moved into nonmetropolitan areas than out. However, a larger percentage of the nonmetropolitan residents moved to metropolitan areas than the reverse—3.0 and 1.8, respectively.[2]

USES OF DEMOGRAPHY

Population size and composition largely determine how funds are apportioned by federal and state governments to counties, municipalities, and other government units. They determine representation in national, state, and local legislative bodies. Furthermore, the rights, duties, and powers of municipalities depend in part on the sizes of their total populations and the sizes of defined segments, such as minorities, unemployed, children, the elderly, rural residents, and others.

SOURCES OF INFORMATION

The major sources of data about the American population are the decennial census; the registration of births, deaths, marriages, and divorces; the registration of immigrants; and sample surveys sponsored by government or by private organizations.

A census of population has been taken by the federal government at 10-year intervals since 1790. After 1980 a census will be taken once every 5 years.

The census is the only single source of information concerning population changes and characteristics for every state, county, municipality, township, neighborhood, and city block. But it primarily reports characteristics of the population at the time of the enumeration. Information concerning intercensal changes is derived by using previous census data as benchmarks and updating them with current data from other public records.

Registration of births, deaths, marriages, and divorces is done by state or local agencies. Although the primary purpose of registration is to meet legal requirements and to provide official records for individuals, these data are compiled on a monthly and annual basis.

Sample surveys by both government and private agencies are important sources of information. For instance, the Current Population Survey, conducted monthly by the Bureau of the Census and based on interviews of 60,000 households, provides monthly estimates of employment and unemployment rates and other information about important population developments. The increased use of computers for tabulation of data has greatly extended the usefulness of raw data.

PROBLEMS IN DATA COLLECTION AND ANALYSIS

Although individual responses to the census questions are compulsory, errors are possible in the information supplied or recorded. Methods have been developed to measure and to reduce the effects of such errors, but concerns about privacy and costs limit the degree to which such errors can be completely eliminated.

Error estimates themselves are subject to error, because they depend to some degree upon assumptions. For example, registration of births and deaths is presumed to be virtually complete, but data on marriages and divorces less complete. Finally, registration of immigrants is beset with a number of difficulties aside from the problems of counting undocumented aliens.

A major problem is that of missing data. Despite efforts to reach every person in the United States, it has been estimated that about 2.5 percent of the population, 5.3 million persons, were missed in the 1970 census.

Data from sample surveys also are subject to sampling, response, and recording errors, as well as errors caused by incomplete coverage of sampling units.

To say that data are subject to error is not to deny their utility for many purposes. Good practice calls for allowances for error along with results of a survey. While techniques for minimizing errors and for estimating the size and nature of those that remain have been developed, they continue to deserve the fullest attention.

Because of the political importance of demographic information, it must be reliable, current, and comprehensive. Many significant demographic shifts occur at the regional, state, or local levels. The capacity of state and municipal agencies to collect, analyze, interpret, and disseminate demographic data often falls short of their requirements. Furthermore, future censuses will require more assistance from local agencies than formerly, and it is important that these agencies be effective and efficient.

DEMOGRAPHIC PROGRAMS OF THE FEDERAL GOVERNMENT

Analytical work of high quality is being done by the federal government, through the Bureau of the Census, the National Institute for Child Health and Human Development, the National Center for Health Statistics, and other agencies. These agencies address questions related to their special missions; but no centralized federal agency has responsibility for monitoring and analyzing overall demographic trends in the United States. Congress, however, is examining the effects of population change; recently, for example, through the Select Committee on Population of the House of Representatives.

DEMOGRAPHIC TRENDS IN THE UNITED STATES

Birthrates have been generally declining for the past 100 years. Americans have been leaving their central cities since the 1920's, but the present rate of migration out of some metropolitan areas is unprecedented. The westward drift of the U.S. population center goes back to the founding of the nation, and in the past 200 years this center has moved from Baltimore to St. Louis. For the past 40 years there has been a slight movement of that center toward the South.

These changes can be interpreted in various ways. For example, if low birthrates and increasing life expectancies continue, the age structure of the American population will be transformed, raising the proportion of the elderly and lowering that of the young. In 1977 the largest 5-year age group was the 15–19-year-olds; in 2000 it will probably be the 40–44-year-olds. This change will affect social and economic structure in ways that are still not certain.

Migration patterns of the 1970's will affect cities and suburbs, the places left behind, and destinations. Many large cities, as well as some suburban areas, are losing young professionals and small towns are gaining them. Many middle-class black and white families are moving out of large cities.

There are exceptions to every aggregate demographic trend. While fertility rates are declining overall, they are higher for some groups within the population than for others. There is outward migration from metropolitan regions, but a large number of people are also moving in. Migration patterns leave some large cities in the north and north-central regions with a larger proportion of poorer, older, and minority populations; but some of these same cities are also experiencing urban renovation efforts by middle-class young families.

RATE OF GROWTH: FERTILITY, MORTALITY, AND IMMIGRATION

The annual population growth rate of the United States is calculated on the basis of gains and losses from births and deaths and international immigration and emigration. This growth rate has steadily declined from more than 3 percent before the Civil War to less than 2 percent between 1920 and 1960 and less than 1 percent in 1978. In numbers, the U.S. population grew by 1.7 million in 1978, or about 500,000 less than the growth in 1970. In 1970 there were 203 million Americans; in 1979 about 220 million. The decline in population growth has been steady but not smooth, marked by sudden drops and surges, such as the sharp increase in the number of births between 1947 and 1964, the years of the baby boom.

The gain or loss from annual births and deaths has and continues to be the largest factor in national population change. People in the United States, and in most major industrialized countries, are having fewer children and living longer. With the recent decline in fertility, immigration is becoming more important; it now accounts for nearly 20 percent of the total annual increase in the number of people in the United States, compared with about 11 percent in the early 1950's.

BIRTHRATE AND FERTILITY

The Baby Boom

After decades of decline the U.S. birthrate rose sharply after World War II, peaking in 1957 and declining below 1947 levels in 1964. Since then, the rate of decline has been rapid and has outdistanced recent declines experienced by other industrial countries. Between 1957 and 1978, the rate of childbearing among all women of childbearing age was halved and the annual number of births declined from 4.3 to 3.3 million.

The very rapid rise in birthrates after World War II and the equally rapid decline in the 1960's were unpredicted. The effects of these fluctuations remain, however, in the form of the very large number of babies born during the baby boom and the much smaller number following them. In all, the 42 million babies born between 1955 and 1964 set a 10-year record which has not been equaled. These babies comprise the nearly 41.5 million 14–23-year-olds in 1979 and will be the 41.3 million persons 35–44 years old in 2000. By 2030, when they will be 65–74 years old, their numbers will have been reduced to about 32 million—more than twice the number of 65–74-year-olds in 1978.

Current Fertility

Women now in the midst of their childbearing period are likely to have two children as compared with three for their mothers. In 1964 slightly more than half of all newborns were first or second births. In 1977, almost three-fourths of all births were first or second children.[3] Women are also likely to complete their families in 7 years as compared with 10 for their mothers. As marriages have been delayed, the mother's age at birth of the first child has been increased, while the age at which she has her last child has declined. Finally, the birthrate for women 35–44 years old declined by two-thirds between 1960 and 1976.

There are other indications that the birthrate will remain low or decline. One such indication is the increased number of fertile couples who seek sterilization as a form of birth control. Among one-fourth of the couples of reproductive age, either the husband or the wife has been sterilized.[4] Moreover, some 11 percent of women in their early twenties expect to remain childless.

Teenage Fertility

Birthrates differ considerably among various groups and fluctuate within those groups. For example, the fertility rates among women aged 15 to 17 are different from those aged 18 and 19, and both these rates are different from the fertility rates for women over 20 and under 15. Furthermore, the rates differ among ethnic groups.

In 1966 there were 629,554 births to mothers under 20, and in 1977 fewer than [illegible]00,000. Of the 1977 births, approximately 214,000 were to mothers 15–17 years old.[5]

Currently the birthrate for all teenagers appears to be declining. The birthrates among 18- and 19-year-olds declined from 121.2 per thousand in 1966 to 85.7 per thousand in 1975. This decline more nearly parallels the declining birthrate among women in their early twenties, the prime childbearing ages. Among 15–17-year-old mothers the birthrates per 1,000 women were 35.8 in 1966, 39.2 in 1972, 36.6 in 1975, and 34.6 in 1976. The rate peaked in 1972 and has declined since then. Birthrates for girls under 15 account for less than 0.5 percent of the total number of births with little fluctuation from one period to the next.[6]

Despite this decline in the number of births to teenage mothers between 1966 and 1975, the teenagers' share of births rose from 17 to 19 percent of the total number of births between 1960 and 1976. The 15–19 age-group increased considerably in numbers in the latter part of the period as the children born between 1957 and 1961 matured. The increased proportion of births to teenagers was further accentuated by the decline in the birthrate among older women.

The birthrate among teenage black women tended in the past to be higher than that for teenage whites. Between 1966 and 1975 the gap in birthrates between black and white 15–17-year-olds narrowed as a result of a decline among blacks and a slight increase among whites. There is some recent indication that birthrates among 15–17-year-old white women are now declining. Among the 18- and 19-year-olds, both blacks and whites, the birthrate had declined considerably. In sum, birthrates among blacks and whites of both age-groups have declined, and the rate of decline has been greater among black teenagers (Table 5).

There has been an increase in sexual activity among teenagers of all groups, but at the same time there has been an increase in the use of contraception and abortion. The latter is believed to account for the declining birthrates among teenagers.

The lower on the economic scale the teenager, the more likely she is to become pregnant, the lower her chances of completing her education, and the less likely she is to have proper medical attention. Also, the younger the mother, the less likely she is to be married. Furthermore, a higher degree of risk is associated with childbirth among the very young, and the younger the mother, the more likely her child will suffer birth defects.

MORTALITY

Mortality rates continue a steady decline among infants, women in childbirth, and the elderly. There has been a decline of 13 percent from 1950–75 in death rates for those 65 and over. This decline at older ages, signifying a new trend, is due mainly to lower death rates from cardiovascular diseases. The result of lower death rates from these diseases is that between 1950 and 1975 life expectancies after age 65 rose from 12.8 to 13.7 years for males and from 15.0 to 18.0 years for females. These changes in mortality rates at older ages have, of course, increased the projected number of the aged in future populations. The most recent estimate by the Bureau of the Census is 31.8 million people 65

TABLE 5 U.S. Birthrates for Women under 20, by Age of Mother and Race (live births per 1,000 women)

Age of Mother and Race	1977	1976	1974	1972	1970	1968	1966
10–14 Years							
White	0.6	0.6	0.6	0.5	0.5	0.4	0.3
Black	4.7	4.7	5.0	5.1	5.2	4.7	4.3
TOTAL	1.2	1.2	1.2	1.2	1.2	1.0	0.9
15–17 Years							
White	26.5	26.7	29.0	29.4	29.2	25.7	26.6
Black	82.2	81.5	91.0	99.9	101.4	98.9	97.9
TOTAL	34.5	34.6	37.7	39.2	38.8	35.2	35.8
18–19 Years							
White	71.1	70.7	77.7	84.5	101.5	102.0	109.6
Black	147.6	146.8	162.0	181.7	204.9	201.3	209.9
TOTAL	81.8	83.1	89.3	97.3	114.7	114.9	121.2

SOURCE: *Monthly Vital Statistics Report*, Vol. 26 (No. 5 supplement), p. 9, 1977; and "Final Natality Statistics, 1976 and 1977," Washington, National Center for Health Statistics.

and over by 2000, or 4–11 percent above projections made between 1964 and 1975.

There are differences in life expectancies between men and women, among residents of different states and different sections of a state, and among races. The difference in life expectancies at birth of males and females has steadily widened, from 4.5 more years for females in 1940 to 7.8 years in 1975. Similarly, there is a difference of nearly 8 years in the life expectancy at birth between those living in the most favored state (Hawaii) and in the least favored (the District of Columbia). The highest state average for women of "all other" races is equal to the lowest state average for white women. A study in Chicago showed that there are differences of as much as 10 years in the mortality rates of small areas within the city. Despite substantial improvements in recent years for blacks and whites, a black female baby born today with a life expectancy of 72 years has the same average life expectancy as her white counterpart born in 1950. Even for those 65 years old and over, the average number of remaining years differs by as much as 2 years between the highest and the lowest state averages.

IMMIGRATION

Net legal immigration into the United States has averaged about 400,000 per year since 1965. To that must be added an unknown number of undocumented aliens. The law provides for a maximum quota of 170,000 immigrants from the Eastern Hemisphere and 120,000 from the Western Hemisphere, and for specified groups of nonquota immigrants, such as the recent admission of refugees from southeast Asia.

While net immigration into the United States now accounts for an increasing fraction of total annual growth of the population, data on immigration are the least developed of our population statistics. There are several problems. The legal definitions of an immigrant are complex: Not all immigrants can be identified as such when they enter the country, and little is known about the persons who leave the country. The movements of undocumented aliens across national boundaries add other complications. There is not now, nor has there been, a set of immigration statistics commensurate with their importance to the United States.

AGE STRUCTURE

Declining mortality rates for the elderly and the concomitant increase in their numbers combined with current low fertility rates and fewer children have increased the proportion of the elderly in the United States. This has resulted in an aging population. By 2000, the population 65 years and older will have increased some 36 percent since 1977. The population of the very old, those 80 and over, will increase especially rapidly, with their numbers almost doubling, from 4.8 million in 1978 to 8 million in 2000.

Countries with high fertility rates, where each age-group is followed by successively larger ones, have a pyramidal age structure with a large number of young at the bottom and a small number of the elderly at the top. As the sizes of the different age-groups in the

United States gradually are becoming more equal, the age structure is taking a more rectangular shape with a small number of very young at the bottom and proportionately more elderly at the top (Figure 26).

EFFECTS OF CHANGING AGE STRUCTURE

Just as "boom" babies crowded schoolrooms and universities throughout the 1950's and 1960's and are now putting pressure on the labor market, they could, when they reach retirement ages, strain pension systems and other services for the elderly.

The changing age structure affects the dependency ratio, the index for measuring the relative size of the working-age population (20–64 years old) in relation to the nonworking-age population (under 20 and over 64 years old). The 1978 dependency ratio was 78.7 nonworking age per 100 persons of working age. In 1960 it reached a high of 91.5/100, reflecting the effects of the baby boom.

The U.S. Social Security system is currently being affected by increasingly large numbers of the elderly in relation to the working population. In 1975 the number of persons 65 and over was 18.9 per 100 persons of working age. By 2010 it is predicted that the ratio will be 20.9/100 and by 2030, as the baby boom group reaches retirement age, the ratio will increase sharply, by 60 percent, and change the elderly/working-age ratio to 33.6/100.[7]

Because of the sharp increase in the size of the working-age population, dependency ratios should improve by 1984. In the long term, however, current low fertility rates will reduce the number of people paying into social security and the number receiving benefits will greatly increase. Therefore costs to the workers will go up.

INTERNAL MIGRATION

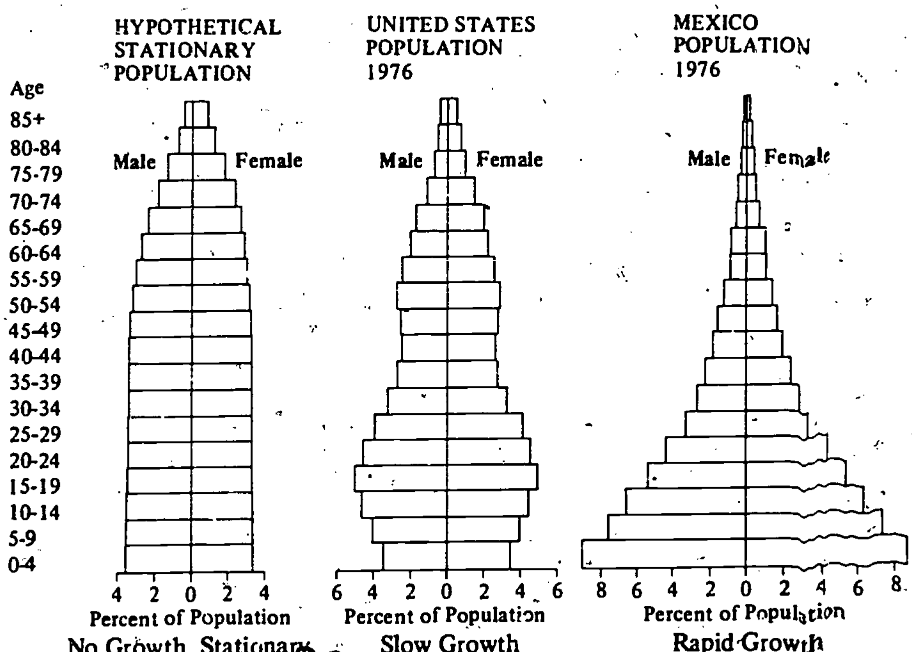
The average 20-year-old American moves nine times in his lifetime.[8] Many of these moves are and will continue to be outward from the cities and westward, continuing an American tradition. What is new is the current movement toward the South, for historically net migration was away from the South to the North and West. In general, 15 states, 8 of them in the South, which had a loss of population to other states in the 1960's, are gaining population in the 1970's. The states growing most by migration are Florida, Texas, Colorado, Arizona, and California. The largest losses were in Ohio, New York, and Illinois.

The historic movement of blacks from the South to other parts of the country also has been reversed in recent years, and there is now a movement of blacks to Maryland, Florida, Texas, and Virginia. However, at the same time, blacks continue to move out of Alabama, Mississippi, Arkansas, Louisiana, and the District of Columbia.

CHARACTERISTICS

Those who move usually differ from those who stay behind in socioeconomic status, age, and education. The movers tend to be young, usually in their

FIGURE 26 Stationary population age-sex structure and contrasts. (Population age-sex structures for Mexico and the United States reproduced from Arthur Haupt and Thomas T. Kane, *Population Handbook*, Population Reference Bureau, 1978, p. 14. Hypothetical stationary population age-sex structure calculated from New Zealand Life Tables, 190–72. Wellington, N.Z. Department of Statistics, June 1976, Tables 1 and 2 for the non-Maori population. [From *Population Bulletin*, Vol. 33, No. 3, 1978. Courtesy of Population Reference Bureau, Washington, D.C.)]



twenties, finished with formal education, beginning their careers, and newly married but with no children. However, there is also a large number of relatively well-off retired people who move to the warmer climates of Florida and the Southwest.

METROPOLITAN EXODUS

The net flow of people from metropolitan areas to small towns and rural areas has received a great deal of recent attention. It eased the "urban explosion" concern of the 1960's best illustrated by the fear of a "Boswash" conurbation. In the 1970's the comparatively higher population growth rate for nonmetropolitan areas reversed the situation of the 1960's, when the growth rates for metropolitan areas were four times as high as for nonmetropolitan areas (Figure 27).

Such aggregate comparisons may be a bit deceptive. Three-fourths of metropolitan population growth in the 1960's came from an excess of births over deaths and only one-ninth from the arrival of individuals from nonmetropolitan areas.[9] Other factors involved were immigration into the United States and expansion of metropolitan land areas as more people owned cars and as the interstate highway system spread. Regional migrations were a minor part of metropolitan growth during the 1960's.[10]

Moreover, loss of metropolitan populations through migration is not new to American society. In the 1960's, for example, 40 percent of metropolitan areas had more people leaving than arriving. But the effects of these migrations were cloaked by the much larger population gains from excess of births over deaths.

Seen in that light, the recent trend toward migration beyond metropolitan areas was foreshadowed in the 1960's and made evident in the 1970's by the continuing decline in fertility rates. The trend is ubiquitous. Most nonmetropolitan areas are gaining migrants, with the major exceptions being the tobacco and cotton belts from North Carolina to the Mississippi Delta and some of the Great Plains areas. Nearly two-thirds of all nonmetropolitan counties gained migrants in the 1970's, compared with one-quarter in the 1960's and one-tenth in the 1950's.[11]

About three-fifths of the nonmetropolitan gain is occurring in counties that are contiguous to the

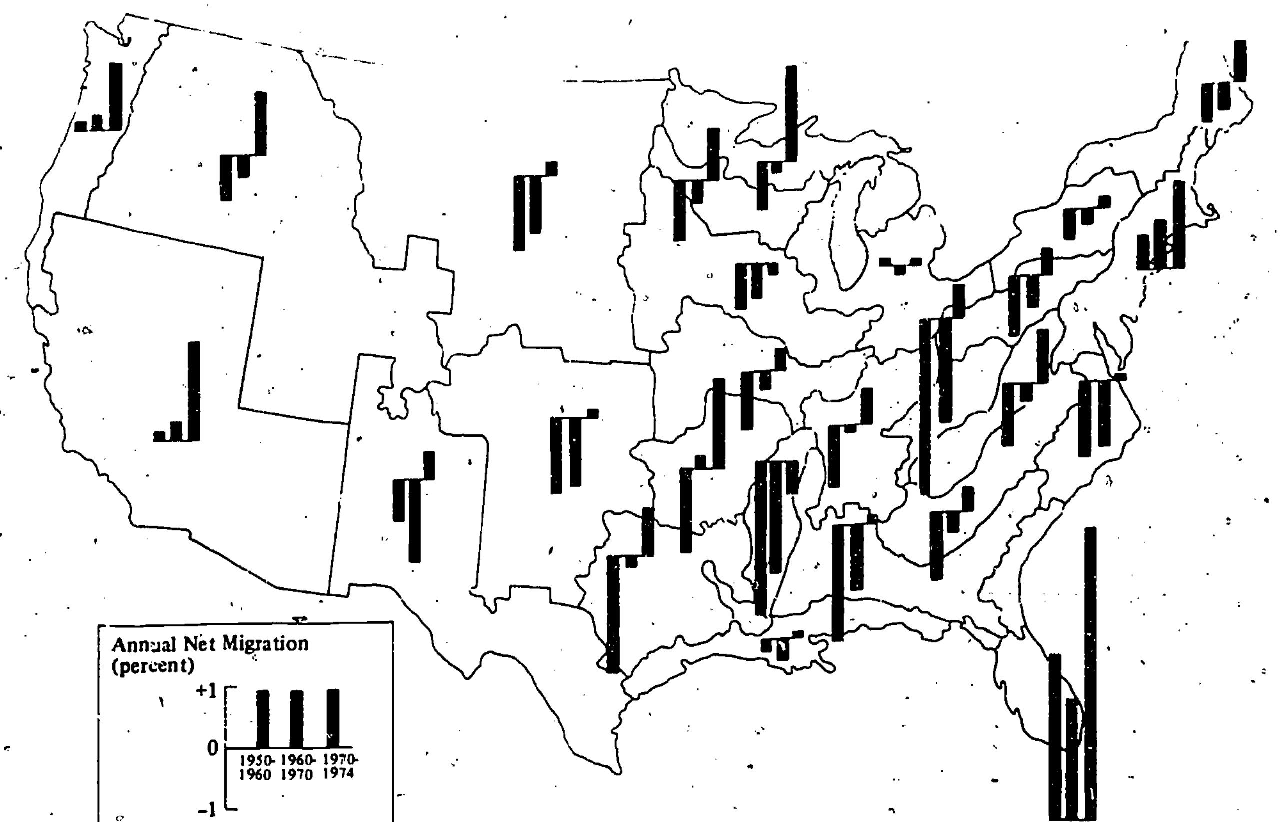


FIGURE 27 Annual rates of net nonmetropolitan migration for 26 U.S. regions. (Calvin L. Beale and Glenn V. Fuguitt, *The New Pattern of Nonmetropolitan Population Change*, CDE Working Paper 75-22. Madison, Wisc.: Center for Demography and Ecology, University of Wisconsin, 1975. [From *Population Bulletin*, Vol. 30, No. 3, 1976. Courtesy of the Population Reference Bureau, Washington, D.C.])

metropolitan areas.[12] At the same time there are over 500 nonmetropolitan counties losing population. Most of these are heavily agricultural and/or predominantly black.

Population losses in the large cities are mirrored to an extent by population losses for entire metropolitan areas, that is, a central city and its suburbs. During the 1950's, only one metropolitan area, St. Joseph, Missouri, had lost population. In this decade, eight metropolitan areas have already lost population. Nationally, metropolitan areas continue to increase their population, but at a slower rate than nonmetropolitan areas. Moreover the 15 fastest growing metropolitan areas from 1970 to 1975 were all in the South and West, including 9 in Florida.

Migration into nonmetropolitan areas and the general movement toward the South and West, where population densities have been generally lower, are resulting in a geographic shift in population densities.

CITY AND SUBURB

The proportion of Americans living in central cities relative to the suburbs has been declining for several decades. In 1920, two-thirds of metropolitan area residents lived in central cities and in 1975 less than half did so. A decline in the total number of people living in the nation's central cities is new, stemming in large part from the continued migration of whites out of the cities. The loss is now national and no longer limited to the cities of the north-central and northeastern regions of the country. However, the largest losses still occur in these latter regions. In contrast, smaller cities, those with 1 million or fewer residents, are either gaining population or suffering losses proportionately smaller than the larger cities.

Measures of Change

The interpretation of the changes occurring in central cities and suburbs depends on what measure one uses; whether income levels, racial and age composition, or family structure.

The differences in median incomes between suburban and central city families appear to be widening: Suburban median incomes were 17.8 percent higher than central city median incomes in 1970 and 23.8 percent higher in 1977. Or, to cite a different measure, over 10 percent of all central city families now receive some sort of public assistance income, but only 3.9 percent of suburban families do so.

The majority of black Americans live in central cities. However, the suburban growth rate during the 1960's and 1970's was higher for blacks than whites; still, suburban populations remain largely white.

The median age of central city residents has historically been higher than that of suburbanites, but that gap is narrowing as the general decline in fertility rates reduces the number of young children in suburbs. Finally, the proportion of all families headed by females is higher in the central cities than in the suburbs, with these differences largely due to black families, 36 percent of which are headed by females compared with 11 percent of white families.

SOME QUESTIONS

These are some of the signs of major changes since 1970 in the movement of population within metropolitan areas and between metropolitan and nonmetropolitan areas. We lack precise information on the patterns and effects of regional migrations—in and out of metropolitan areas, suburbs, and central cities, and geographical regions. Differences in the volume of such migrations and in the characteristics of the migrants are critical elements in determining the economic and social effects of migration into and out of a region, town, or city.

As noted earlier, rural areas, which historically lost population to the cities, have recently begun to keep more of their population and to attract more newcomers. Questions remain: Are former rural residents returning? Are urban-reared and educated persons opting for rural life? Are jobs, climate, recreation, or some combination of these the major attractions? Are the people left behind increasingly the disadvantaged? Do changes in family organization and in settlement patterns affect the needs for public services by specific groups, such as the elderly?

HOUSEHOLDS AND FAMILY FORMATION

Recently there has been a noticeable postponement of the age of first marriage. The median age of marriage for both men and women increased by about 1.5 years between 1956 and 1977, and the proportion of women 20–24 years old who have not married increased by one-half between 1960 and 1970.

Combining such demographic trends with other social factors is useful in analyzing social patterns. For example, the postponement of marriage is associated with higher college enrollment among women, increased number of women in the labor force, and increased unemployment among young adults. The increased number of young unmarried women of marriageable age may also be related to the fact that there are 5–10 percent more women at the usual age of first marriage (18–24) than there are young men (20–26), because the older men were born during a time of low birthrates, while the women were born during the baby boom. The same effect might work in reverse in the next five years, as the men born during the baby boom seek marriages among women born in the 1960's, when birthrates began their sharp decline.

DIVORCE

Divorce rates historically tend to rise and fall with marriage rates, with the former lagging behind the latter by seven years for first marriages and three years for remarriages. Since marriage rates peaked in 1972 and have now declined by about one-tenth from that level, it can be expected that the rate of divorce will begin to decline about 1980.

ONE-PARENT FAMILIES

In 1977, 17 percent of families with children under 18 living at home were one-parent families and 90 percent of these parents were women. Seventeen percent of all children under 18 were living with one parent, usually the mother, in 1977. Ten percent of children living with two parents live with a stepparent. Based on current trends, five years from now 20–25 percent of all households with children are likely to be one-parent families.

Although 90 percent of all families in the United States are headed by men, half of all families below the poverty level with dependent children are headed by women. Although the percentage of families headed by women is greater among blacks (33.1 percent in 1976) than among whites (11 percent), the number of white families headed by women is increasing and in absolute figures is well above that of blacks. In 1975 the Census Bureau reported 5,380,000 white families headed by women and 2,102,000 black and "other" families headed by women.

HOUSEHOLDS

Although population growth is slowing, the number of households is growing three times as fast as the total population—an increase of 20 percent between 1970 and 1978, as compared with an increase of only 7 percent in the total population. This is a result of the trend toward smaller families, more single parents, fewer or delayed marriages by young adults, more elderly couples, widows, and widowers living alone, and less doubling up because of greater affluence. The effect is that while the population of a particular city, suburb, or metropolitan area may be declining, the number of households may be increasing.[12]

The growth rate was led by one-person households, in particular those maintained by persons either divorced or never married. Half of all divorced persons who have not remarried and half of those who have never married maintain households alone. Among divorced parents, a father usually lives alone while the mother maintains a one-parent family.

The average size of households declined from 3.3 persons in 1960 to 2.8 in 1978, and may be expected to decline further to 2.6 persons by 1983.

Elderly persons will contribute substantially in the next five years to the continued growth in the number of small households. In 1977, 33 percent of women over 65 but only 9 percent of the men were living alone. If current trends continue, the number of households maintained by persons 65 years old and over is expected to increase by 10 percent, or double the expected rate of household formation for the total population.

If present trends continue, the rate of household formation in the next five years will continue to exceed by far the rate of population growth.

LABOR FORCE

CURRENT LABOR FORCE PARTICIPATION

In May 1978 the total civilian labor force exceeded 100 million persons for the first time. During the preceding five years, while the population had increased by about 8 percent, the total number of employed persons had grown by about 11 percent. Moreover, the proportion of the total population in the work force continued to grow. Unemployment, which fluctuates more rapidly than total employment rates, also increased, with the result that in 1978 there were both more employed and unemployed workers than in the previous five years.

These recent changes continue a long-term trend of increasing proportions of the population working for pay or profit. Also, the ratio of workers to nonworkers by sex and age has fluctuated widely, primarily as a reflection of the increasing proportions of women in the work force and a shorter work span for men. Reductions in the numbers of the self-employed have increased the proportions of older men who drop out of the labor force abruptly at retirement rather than continuing as self-employed persons who can adjust their work to declining ability. Also, a growing trend of basing retirement benefits on length of service, instead of on a fixed retirement age, may have led to some increase in earlier retirements. An increased number of years devoted to education has delayed entrance into the labor force. However, by age 25–29, at least 95 percent of men are in the labor force. Young women are returning to the labor market much more quickly after the birth of children, and more men are retiring earlier than the customary retirement age.

A much higher proportion of men and women in the age range 25–54 are working than was true in the past. As a result, the work life patterns of men and women are becoming increasingly similar.

Earlier Retirements

Smaller families and completion of families at younger ages may contribute to earlier retirement because of reduced economic responsibility as the youngest child leaves home while the parents are relatively young. The increasing labor force participation of women may also contribute to early retirements because of additional income provided by the wife after retirement of the husband. Increased numbers of young workers may lead to continued pressures on older workers who find it more difficult to adjust to changing technologies. However, the recent change in legislation raising the age of mandatory retirement may slow these trends.

Prospective changes in the age composition of the population and of the labor force imply a substantial decline in the ratio of nonworkers to workers. (Not to be confused with "dependency ratios," which are based on working-age population, whether working or not.) This ratio has already declined from 152:100 in 1965 to 125 in 1975 and is expected to continue to decline to about 111 by 1985. By that time a growing proportion of nonworkers will be older persons.

Major changes have occurred in the kinds of jobs that are available. Significant increases have occurred in the proportion of workers employed in professional, technical, clerical, sales, and service occupations. Correspondingly jobs for unskilled workers and farm workers have declined rapidly. These changes have been shared by men and women; but women have also made major changes from household work to clerical and sales occupations.

Black workers in general have experienced the same changes in their occupational distribution, but at a much more rapid rate. The rates of labor force participation among men have been dropping somewhat more rapidly among blacks than whites. The rates for white women have been increasing more rapidly than those for black women but the rates for black women continue to be higher than those for their white counterparts.

CHANGES IN NUMBER OF WORKING WOMEN

In 1977, about half of all women in the United States were in the labor force, compared with about a third in 1950. Between 1965 and 1977 the proportion of young women (25–34) in the labor force increased from 39 to 60 percent. This increase is particularly remarkable because the majority of the women in this age-group are married and many have children at home, factors that traditionally have kept women out of the labor force. Indeed, the largest increase in labor force participation in recent years has been among mothers with preschool children.

Earnings are the sole income for most of the 16 million women in the labor force who are single, separated, divorced, or widowed. Within husband–wife families, the majority of wives now work and account for roughly 30 percent of the families' incomes. This additional income has been an important factor in lifting some families above the poverty level. Where the husband's earnings are more substantial, the income from the wife's work may cover the children's college expenses, help maintain living standards during inflation, and permit the husband to consider other work or early retirement. It is possible that the gap between poor and well-to-do families will widen with the increase in the number of families with two high-salaried earners.

UNEMPLOYMENT

There is currently a high degree of unemployment, with an average of 6 million unemployed persons in the United States in 1978—or 6 percent of the civilian labor force. To be considered unemployed a person must be out of work, looking for work, and available for work. Persons looking for their first job are included, along with others who have been in the labor force for long periods of time. Fully retired persons are not considered unemployed.

The National Commission on Employment and Unemployment Statistics currently is evaluating the concepts and measures of labor participation and unemployment.

About half the unemployed persons lost their last job and about one-fourth were coming back into the job market after having been out for some time. The remainder were almost equally divided among those seeking their first job and those who had left a previous one. More women than men were new entrants or re-entrants, and more women than men had quit their jobs.

Unemployment rates vary considerably by age, sex, race, residence area, and occupation. They are particularly high among those born during the baby boom who are now beginning to enter the labor force (Table 66).

Unemployment by Age

One-third of the unemployed are in the prime working ages 25 to 44. In recent years 150,000 people over 65

TABLE 6 1978 Annual Average Employment Status of the Population in Metropolitan and Nonmetropolitan Areas by Sex, Age, and Race (thousands)

Employment Status	Metropolitan Areas Total	Metropolitan Areas Central Cities	Metropolitan Areas Suburbs	Non-metropolitan Total
Total				
Civilian noninstitutional population	107,391	45,323	62,068	51,550
Civilian labor force	68,738	28,108	40,630	31,682
Percent of population	64.0	62.0	65.5	61.5
Employed	64,529	26,029	38,499	29,844
Unemployed	4,210	2,079	2,131	1,837
Unemployment rate	6.1	7.4	5.2	5.8
Not in labor force	38,653	17,215	21,437	19,869
Males, 20 years and over				
Civilian noninstitutional population	45,158	18,660	26,499	21,848
Civilian labor force	36,459	14,484	21,975	17,005
Percent of population	80.7	77.6	82.9	77.8
Employed	34,880	13,658	21,223	16,332
Unemployed	1,579	825	753	673
Unemployment rate	4.3	5.7	3.4	4.0
Not in labor force	8,699	4,176	4,523	4,842
Females, 20 years and over				
Civilian noninstitutional population	51,219	22,267	28,953	23,269
Civilian labor force	25,897	11,268	14,629	11,518
Percent of population	50.6	50.6	50.5	47.5
Employed	24,360	10,511	13,849	10,820
Unemployed	1,537	757	781	699
Unemployment rate	5.9	6.7	5.3	6.1
Not in labor force	25,322	10,999	14,323	12,751
Both sexes, 16–19 years				
Civilian noninstitutional population	11,014	4,397	6,617	5,433
Civilian labor force	6,382	2,356	4,026	3,158
Percent of population	57.9	53.6	60.8	58.1
Employed	5,289	1,860	3,428	2,692
Unemployed	1,093	496	597	466
Unemployment rate	17.1	21.1	14.8	14.7
Not in labor force	4,632	2,039	2,591	2,275
White				
Civilian noninstitutional population	92,782	34,856	57,925	46,798
Civilian labor force	59,566	21,732	37,834	28,890
Percent of population	64.2	62.3	65.3	61.7
Employed	56,464	20,491	35,973	27,372
Unemployed	3,102	1,241	1,861	1,518
Unemployment rate	5.2	5.7	4.9	5.3
Not in labor force	33,216	13,124	20,091	17,908
Black and other				
Civilian noninstitutional population	14,609	10,467	4,142	4,752
Civilian labor force	9,172	6,376	2,796	2,792
Percent of population	62.8	60.9	67.5	58.7
Employed	8,065	5,538	2,527	2,472
Unemployed	1,108	839	269	319
Unemployment rate	12.1	13.2	9.6	11.4
Not in labor force	5,437	4,091	1,346	1,961

SOURCE: *Employment and Earnings*, Vol. 26, No. 1, 1979. Washington, U.S. Bureau of Labor Statistics.

Danny Lyon, EPA—Documerica

have been reported as unemployed. One-fifth of the unemployed are between 15 and 19 years old.

Teenage Unemployment In 1977 unemployment rates for teenage blacks were more than double those for whites, with unemployment rates for black teenagers of 40 percent in many places and possibly higher in others. Youth unemployment may be undercounted, and this may be particularly true for black unemployment rates in central cities.

In the past, high unemployment rates among teenagers have been a transitory phenomenon, an inevitable consequence of the school-to-work transition, since these rates diminished to adult levels as the youths entered the adult labor market. However, there is some evidence that those who were unemployed or underemployed in their early years continue to have difficulty in later years and that their earnings remain relatively low.[13] This is particularly true for black and white females and for black males.

By Race and Sex

Unemployed men exceed women in numbers in each age-group, and the number of unemployed whites exceeds the number of unemployed blacks in each age-group. But, while there are nine times more whites than blacks in the United States, the number of unemployed whites is only about four times as great as that of blacks. Women are more likely than men to experience unemployment. Unemployment rates among male family heads have been considerably below and female family heads above that for the labor force as a whole.

By Residence Area

In 1978, rates of unemployment for men over 19 years old were highest in central cities, and the rates in the suburbs were somewhat lower than those in nonmetropolitan areas. The rates for women over 19 were highest in the central cities and lowest in the suburbs. In the third quarter of 1978 unemployment rates for men and women 16–19 years old were 12.1 percent for metropolitan residents and nearly as high, 11.4 percent, for those living in nonmetropolitan areas. Unemployment rates for white teenagers were highest in central cities, but the rates for suburbs and nonmetropolitan areas were identical. Although suburban areas account for more of the population than any other, the numbers of unemployed persons were almost evenly distributed among the three residence areas—central cities, suburbs, and nonmetropolitan areas.

By Occupation

In general, blue-collar workers are about twice as likely to be unemployed as white-collar and farm workers, with unemployment rates for service workers falling between these groups. Service workers include those involved with protection, food, health support, and cleaning.

Duration of Unemployment

In recent years, about one-fifth of the unemployed have been so for 4 weeks or less. One-sixth to one-fifth have been unemployed for more than 26 weeks. The average duration of unemployment has been 14–16 weeks or between 3 and 4 months. Fewer than half the unemployed were drawing unemployment insurance benefits, some because they are new workers, others because they have exhausted their benefits, or because they were not covered in that program when working.

LABOR MARKET OUTLOOK

The large increase in the number of adults aged 14–22 due to the high fertility rates from 1957 to 1964 will probably put severe pressure on the employment market and for a time may narrow their employment opportunities. Also, some career opportunities will lag behind others. Falling birthrates have already decreased job opportunities in elementary and secondary school education and are beginning similarly to affect employment in colleges and universities.

The pressure on the youth labor market from the baby boom will continue during the next five years but will begin to ease by 1981. However, this easing, with its attendant decline in teenage unemployment, will be manifest mainly for white youths, as the number of black teenagers reaching working age will decrease only slightly. Also, other demographic forces will continue to exert pressures on the youth labor market, including shifts of population and industry out of the central cities and increased competition among the older unemployed for the low-skill jobs traditionally held by teenagers.

SOCIAL CONSEQUENCES OF CHANGED LABOR FORCE PARTICIPATION

The new working patterns for women come at a time of recent decrease in fertility, increases in marriage and divorce rates, and later age of marriage. These trends make it more likely that women will continue to work at present or higher rates. Conversely, as women are assimilated more into the job market, they are likely to marry later, have fewer children, and divorce more readily.

Employment Issues of Women

These relationships between patterns of work and changing family organization in turn highlight several social problems. For example, women are still predominantly employed in "female" occupations, and the median income from their full-time work is less than 60 percent of the median income of males.

These average lower earnings particularly affect women who live alone and must depend solely on their own earnings, even more so for women with children. Unmarried mothers may have difficulty simply providing basic needs, given both low earnings and the complications of caring for a child during working hours. Divorced women who have not remarried—the fastest growing group of female family heads—typically also do poorly economically when they enter the labor market. Also, divorced women with children usually assume child-care responsibilities. Even though support payments from fathers are often specified in divorce and separation cases, they are frequently unpaid and uncollectable. Most divorced women eventually remarry, but usually only after a financially precarious period. Those who do not remarry often do not have the job skills needed to earn an adequate living.

Day Care of Children with Working Mothers

Half of all children under 18, and 37 percent of preschool children, have working mothers. There are obvious pressures on these mothers, many living alone, who must both keep a home and arrange for supervisory care of their children during working hours. In about two-thirds of families with working mothers, the parents, living alone or not, care for their children themselves. For the remainder, child care is provided in a variety of ways: in their own homes by relatives or others, in the homes of others, and least frequently (3 percent) in day-care centers.

Equity in Taxes and Benefits

The increase in the number of married women who work has raised problems of equity in federal income tax rates and social security benefits, both of which are oriented toward the family as the basic economic unit. Given progressive income tax rates, the two-earner married couple may pay more taxes than two single persons who live together and earn the same total income, thus incurring a "marriage tax." Social security taxes are levied on individuals, but many of the benefits are family oriented. For example, a working wife may find at retirement that her contributions to social security have earned her nothing above the retirement benefits she would have received through her husband's contribution. There probably will be increased pressure for some changes in the rate structure of income taxes and in social security benefits to allow for changes in family income patterns associated with the increased participation of women in the labor force.

SCHOOL AND COLLEGE ENROLLMENT

Sixty million persons were enrolled in schools or colleges in both 1973 and 1978. But the mix by type of school changed considerably during these years. The number of nursery school children increased by half a million from 1.3 million to 1.8 million; kindergarten and elementary school enrollments went down by 3 million from 34.5 million to 31.5 million; high-school enrollments were virtually constant at 15.5 million; and college attendees went up by 2 million from 9 million to 11 million. Increases at the nursery school and college level resulted largely from rising enrollment rates, whereas the other enrollment figures reflect primarily the population level of the relevant age-group at the two dates.

During the next five years elementary school enrollment may decline another million. High-school enrollment is now at its peak and is expected to decline nearly 2 million in the next five years.

College enrollment increased between 1973 and 1977 but fell by 400,000 between the fall of 1977 and the fall of 1978. The population of 18–24-year-olds will not peak until 1981. But the number who will actually be entering college is difficult to predict, and may actually decline gradually before another five years pass. In brief, the total number of college students may have come close to a plateau with little change expected in total attendance for a few years. The number of graduate-school students in 1978 (1.7 million) was somewhat larger than the number of seniors (1.4 million) and has been increasing somewhat faster than that of the rate of growth of the seniors (22 percent versus 17 percent since 1973).

White college students were 16 percent more numerous in 1978 than in 1973, whereas black college students were 49 percent more numerous. In 1978, black persons 18–24 years of age were 12 percent of the total population, and black college students were 11 percent of all college students. Corresponding proportions for persons of Hispanic origin were 6 percent and 3 percent, respectively.

In 1978 white women, white men, and black women at ages 18–21 years had somewhat similar enrollment rates—32, 35, and 29 percent, respectively. Black men, however, had a much lower rate, 22 percent.

If present trends continue, black college enrollment rates may be expected to continue to approach those of whites, but those for persons of Hispanic origin seem likely to continue to lag behind.

BIRTHRATE PROJECTIONS

Projecting the number of births for the next decade seems on its face a simple matter, since the number of potential mothers is known, their attitudes toward family size have been surveyed, and it is reasonable to assume that low fertility rates will continue. However, the assumption is tempered by the fact that the baby boom came as a surprise to most social scientists and also by the current disagreements as to what birthrates will be in the early 1980's, less than five years from now.

In 1977 the Bureau of the Census made projections based on a range of fertility rates including an intermediate rate of 2.1 births per woman (the long-term replacement rate) and a low rate of 1.7 close to the actual current rate of 1.8.

If the current fertility rate of 1.8 continues, there will be some 250 million Americans by 2000 compared with 220 million in 1979. Should the fertility rate rise to 2.1 children per woman or decline further to 1.7 children per woman, assuming no change in immigration rates, the population of the United States will rise by the year 2000 to 259 or 245 million, respectively. In all, if there were no change in the timing of childbearing and fertility were to attain exact replacement level in 1980 to 1985 and remain there indefinitely, the population of the United States would increase some 29 percent between 1975 and the year 2030, and decline slightly until the year 2100.[14]

One analysis predicts that fertility will increase in the next few years as the proportion of young men in the labor force declines. This smaller number of men will command higher incomes relative to the incomes of older men, and young men will, therefore, be able to marry at a younger age. They will also begin their families sooner and be able to support larger families. The analysis predicts increasing fertility rates about 1984 and, with increased numbers of young children, a significant change in the age structure of the population.[15]

However, another view argues that with the perfection of contraceptive methods we are fast approaching an era of few, if any, unwanted pregnancies. This, combined with transient marriage patterns, should lower fertility rates and lead to further declines in population growth rates.[16]

Regardless of the fertility rate, however, births will be the major component of the country's population growth for the next several decades. Even if the unprecedented 20 million American women now approaching childbearing ages have children merely at the replacement rate, they will add some 3.4 million babies to the population each year throughout the 1980's—an echo baby boom. Thereafter, the number of births should begin to decline as much smaller numbers in that age-group reach childbearing age and as the baby boom men and women leave it.

IMPLICATIONS OF GEOGRAPHIC REDISTRIBUTION

The changes in distribution of population will lead to changes in the need for community services—ranging from schools to highways to chronic care facilities for the elderly—since they are shaped by the size and composition of the population. Because migrants tend to be young, better educated, and therefore more able to secure well-paying jobs, they may alter the service requirements in both the areas where they settle and those they leave.

An important issue is the effect of growing and declining populations on the ability of communities to finance services. In regard to growth, an increase in population in sparsely settled rural areas may result in lower per capita costs for public services, since, in many cases, a moderate population increase enables a more efficient use of already existing facilities such as roads, schools, and hospitals.[17] However, in other cases significant increases in population may raise the per capita cost of services if existing facilities are used beyond capacity, requiring increased capital expenditures for additional facilities. Furthermore, newcomers often require services before they make commensurate contributions to tax revenues. A rapidly growing area, for example, may require an immediate expansion of school facilities, with contributions of the newcomers to taxes coming much more slowly.

Recent population changes include rapid growth in nonmetropolitan areas adjacent to existing metropolitan areas. Like the earlier rapid growth of the suburbs, this has placed additional strains on the multiplicity of local governments. There have already been a number of governmental efforts to meet the need for areawide planning.

When population declines in urban areas with established facilities, the remaining poor populations often do not generate sufficient taxes to support the social services that they need most. Also, the demand for public services in relation to changing composition of the population is unclear. Which services, such as crime protection, criminal justice, education, and income maintenance, are in greater or lesser demand as a consequence of lower population? Or does the remaining population require different services?

The costs of urban public services are difficult to control or reduce despite declines in population. Often the costs, such as municipal debt servicing, were incurred before the decline set in, and usually infrastructural systems and services, such as schools and utilities, were set up to support a much larger and wealthier population.[18]

In time, however, it is possible that declining populations may reduce the demand for many services thus reducing costs.

SCHOOLS

The falling birthrates beginning in the early 1960's have resulted in the yearly declines of about 1.5 percent of the elementary school population. Secondary school enrollments will certainly decline through the 1980's. That population will drop from almost 16 million in 1970 to about 14 million in 1985 and 13 million in 1990, a 20 percent decline in 20 years.[19]

Declines in numbers and the migration of school-age children will have different effects in different parts of the country. While the northern school systems and cities in particular are losing population and may face greater than average declines in secondary school enrollments, systems in the new settlement areas will probably need to build more schools to meet the demands of young migrants with school-age children.

In the first half of the 1970's enrollments grew by 5 percent or more in Arizona, Florida, Alaska, Nevada, and New Hampshire. However, enrollments declined in some of the local school systems within those states. Impacts will generally parallel those of gross population movements. For example, the net migration of school-age children 5–13 years old tends to flow from central cities to suburbs to nonmetropolitan areas.

Growth and decline pose different problems for school systems.[19] Growth, unless anticipated, may result in severely crowded classrooms and hasty and inadequate financing. School systems losing population are facing personnel problems related to excess numbers of teachers and school administrators, including tenure, job protection rules, and other limits on personnel reduction. The closing of neighborhood schools on economic grounds because they serve too small a student population has also encountered strong resistance from area parents.

SOCIAL SERVICES

Social services for the poor, ill, aged, unemployed, or disabled require special facilities and trained personnel. Migration has altered the proportion of these vulnerable populations in many places, changing the level, nature, and location of pension, welfare, and rehabilitation programs and hospitals and other facilities.

Between March 1975 and March 1976, the South had a net gain and the northeastern and north-central regions a net loss in populations of the old and the very young, unemployed women, and those below the poverty line[20] (Table 7).

However, the actual fiscal consequences for the South and West of gaining populations in need of social services are problematical, because these same regions are also gaining young and relatively well-educated populations. In effect, additional costs to meet increased need for social services in these areas may be offset by increased ability to pay for them.

The future social services needs of populations in central cities, suburbs, and nonmetropolitan regions are more difficult to categorize and should be studied, analyzed, and planned for. Overall, central cities and nonmetropolitan areas in 1974 had the highest proportion of the elderly, the poor, and one-parent families led by women.[2] Countrywide, the nonmetropolitan areas had the largest drop in the proportion of the population below the poverty line. At the same time the proportion of elderly populations and female-headed families increased most rapidly outside central cities.

General inferences from these and other patterns suggest that the demand and need for social services will continue at the same or higher levels in the central cities even though they may be losing populations, that the influx of the young and the old into nonmetropolitan areas and the South will increase the demand for services related to their particular needs, with or without public financing, and that the need for programs of income support for the elderly will possibly increase in sun belt regions.

JOBS

Economic patterns tend to mirror those of populations and vice versa. As with populations, there is a continuing shift in manufacturing industries from metropolitan areas to small cities and rural areas. Since 1970 the growth in nonmetropolitan areas of wages and salaries and nonagricultural jobs has been twice that of metropolitan regions. By and large, the shifts to nonmetropolitan areas by manufacturing have been in labor-intensive industries, while capital-intensive industries have tended to stay in metropolitan regions.[21]

While there was little increase in the total number of jobs in industry between 1960 and 1975 (about 1.5 million jobs, or an 8.8 percent increase), regionally the changes were considerable. The South accounted for almost all of the job gains at the same time that the Northeast lost almost 14 percent of its 1960 total.[22]

However, while it is possible to discern relationships between regional shifts in employment and shifts in population, there is no direct quantitative link between increase in employment and population. Relationships between migration patterns and employment change with time and with the particular region. Employment in an area may be determined by the state of the national economy, the nature of the industry, or the socioeconomic status of the population.

That there is a positive but rough link between employment and population is illustrated by the fact that while population growth in the South and West has been similar only since 1970, both regions have shared a similar rate of employment growth since 1960. Growth in local employment is also coupled with the number and characteristics of migrants. Increases in population in a region generate employment through additional demands for goods and services at the same time that the migrants are employed to provide them.

MIGRATION AND RESOURCES

Land

Population density in the United States is remarkably low compared with other industrialized countries.[23] The dispersion of the population from central cities to suburbs to beyond the metropolitan zones will continue to reduce the population density in some areas of the United States while raising it in others.

The shift of people to the South and West is a shift from high to low density and from high-value to low-value land.[17] Both the low densities and low prices imply that a resident moving from Baltimore to Houston may sharply increase the amount of land he uses. The lower costs of land in settlement areas also imply future increases in low-density homes and businesses until land values rise and force more intensive use of the land.

Energy

Shifts in population between the North and the South and from high-density to low-density areas will create shifts in energy usage. Because residential heating is an important consumer of energy, the population gains by the South and Southwest from the colder parts of the country may in time result in significant national savings of fuel for heating purposes. These savings may be offset, however, by the greater use of the car and higher gasoline consumption in the South and West because of lower residential and business

TABLE 7 Net Changes in Selected Characteristics of Immigrants and Outmigrants by Region, 1975–1976 (thousands)

	Northeast	North Central	South	West
Under 5 years of age	−29	−16	+21	+22
65 Years or over	−18	−18	+27	+8
Below poverty level	−40	−100	+78	+63
Unemployed male	+1	−7	−20	+30
Unemployed female	−13	−10	+19	+4
4 Years of college	−42	−18	+17	+41
Professional, technical workers				
Male	−7	−34	+16	+25
Female	−4	−15	+7	+13

SOURCE: *Current Population Reports*, P-20, No. 305, p. 108, 1977. Washington, U. S. Bureau of the Census. (From "Social Services and Population Redistribution in the 1970s." *Implications of Population Redistribution in the United States in the 1970s*. Washington, National Research Council, February 1978.)

densities. Rural residents drive 70 percent more miles than those living in cities with populations of 100,000 or more. As rural areas continue to gain population, their consumption of gasoline will continue to mount and they may become proportionately more vulnerable to fuel shortages.

Conversely, a decline in the populations of large metropolitan areas implies a reduction in the demand for mass transit because such transportation, to be efficient and cost-effective, must rely on high population densities.

CONCLUSION

ZERO POPULATION GROWTH

A slowing rate of population growth, including a total fertility rate currently below the long-term replacement level, does not portend an early arrival of zero population growth. With 3 million births and 2 million deaths per year, the population continues to grow, even without immigration. Also fertility rates may decline further.

FUTURE AGE STRUCTURE

Whatever the overall rate of growth may be, whether increasing or leveling off, major population changes will occur in the near future. Baby boom babies are passing into adulthood, establishing new families and households, and creating new demands. As they move from teenage to adult status and into middle age, numerous adjustments in living patterns, work relations, and public policy are to be expected. The aging of the population, especially the rapid increase in the number of persons who have passed their sixty-fifth birthday, will be significant in the country's social and economic future. This will be the case even before the survivors of the baby boom become 65, sometime after 2012.

Between 1978 and 1983 the increase in the number of persons who are over 65 almost balances the expected decline in the number under 20, while the number of persons who are between 20 and 64 years of age increases by about 10 million persons. But between 1983 and 1993, the population of the elderly and those under 20 are both expected to increase by about 5 million each. This 10 million increase in these two groups is barely met by an increase of 12 million in the age-group they sandwich—the 20–64-year-olds.

Lower fertility increases the likelihood that more women will enter and thereby enlarge the labor force. Changes in retirement legislation could extend the working life of the elderly, also increasing the labor force. If both developments occur, the burden of support for persons in dependent ages would be reduced relative to the numbers in the labor force.

Consumer demand will be affected by the increase of the young middle-aged. Between 1978 and 1983 the increase in the number of persons between 25 and 44 years of age accounts for about 90 percent of the total projected population growth. This may be good news for consumer markets providing goods and services for relatively young households.

POPULATION REDISTRIBUTION

The future effects of population redistribution are immediately visible. There will be significant differences in population growth through migration among regions, states, urban and rural areas, and between metropolitan and nonmetropolitan areas. There will be differences in fertility and migration by age, sex, and color. The South and the West will be confronted with issues arising from relatively rapid growth, while the northeastern and the north-central states will face problems of managing older and declining populations.

Current information about migrants, their motives, and their adjustments to new locations is very limited, and additional information must await the 1980 census. There is evidence that trends reported through 1976 continued into 1978.

ELEMENTS OF CHANGE

Trends revealed by demographic analysis can be mechanically extrapolated to form a picture of the future of 1984 or 2000. However, such projections should not be taken as forecasts, for they do not allow for the possibility of major shifts in the underlying forces that cause demographic change.

Further understanding of these forces is essential to answer the questions raised by demographic analysis: Why are more Americans moving to smaller metropolitan areas and rural regions? Does the decision by more Americans to marry later or not at all signal a deep change in the structure of the American family? Or is the phenomenon a temporary interruption of traditional patterns just as the baby boom interrupted a long-term decline in fertility rates? What is the effect now and what will be the effects in the future of the decisions of young Americans to have fewer children than their mothers had?

Answers to such questions are complicated by the weaving in of individual motives and circumstances, but also by the fact that demographic trends themselves have changed very rapidly. Population projections into the future have had to be revised frequently. Not many years ago it seemed reasonable to project a population of about 300 million by the year 2000. The Census Bureau's preferred current projection now is 260 million, but this may also be an overestimate.

Neither the baby boom nor the speed of the recent declines in fertility were accurately predicted. It is not known what the course of fertility will be during the next five years, for changes can be and indeed are so rapid that only current observations are dependable.

Historical relationships and trends should be used cautiously as a guide to the future. It cannot be assumed, for example, that marriage and fertility rates will correlate in the future as they did in the past. To illustrate, the baby boom occurred partly because more Americans decided to marry at younger ages and thereby increased their chances of having more children. However, the fact that many Americans now are putting off marriage does not mean that when and if they do marry they will choose current patterns in family size. They might have three or four or no children instead of the "standard" one or two. Nor, if marriage rates go up, can one assume that the "usual" proportion of couples will choose to have children.

Demographic information alone is not sufficient to make intuitive judgments. For example, while the number of children three–five years old in the United States has dropped considerably in the last decade, enrollments in nursery schools and kindergartens have increased sharply. The reasons are not embedded in demography, but in other social influences, such as, perhaps, the increased number of working mothers.

It should also be kept in mind that global assessments of trends often fail to reveal highly significant developments at the subnational level. For example, while from 1970–75, 11 of the fastest growing counties were in the South, 16 of 25 counties losing the greatest percent of population were also in the South.

Whatever the difficulties in coping with the rapidity of demographic change and the uncertain insights to the future offered by past events, population trends do have long-lasting and deep effects. Since the size of current population groups is known and since mortality rates are unlikely to change drastically, barring catastrophe, changing dependency ratios can be predicted and their effects anticipated. Post-World War II increases in fertility rates will affect the American social structure well into the next century. Schools and colleges, the labor market, the types and magnitude of public services demanded, and, eventually, the needs of the elderly will all be affected by current demographic change.

OUTLOOK

The following outlook section on demography of the American population is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

Two significant recent demographic developments in the United States are the changes in age structure, the results of the sharp increase in birthrates after 1947 and the subsequent sharp decline after 1957; and changes in the patterns of internal migration, signified by net movement to nonmetropolitan areas.

AGE PROFILE

As a result of the 1947–64 baby boom, by 1984 the population of the United States will have an unprecedented number of young adults of working age (20–37 years old). Because of the continuing decline in the birthrate, there will be relatively fewer school-aged children and university-aged men and women. At the same time, there will be a significantly higher number of elderly people. These combined factors contribute to the aging of the population. However, the early 1980's may see the beginning of an upswing in the numbers of the very young as the large young-adult population reaches traditional marriage and childbearing ages.

GEOGRAPHIC DISTRIBUTION

Current migration and settlement patterns in the United States have changed population densities in many areas. These patterns are characterized by a decline in the populations of the urban centers of the North and Northeast and an increase in the populations of the Southwest and South. Although many young couples have moved into city centers and upgraded housing and other facilities, the overall trend is away from large cities. If current trends continue, in

1984 central cities, especially in the northern and north-central states, will have a disproportionately large number of the elderly and impoverished.

FAMILY SIZE AND STRUCTURE

Trends toward high female employment rates, high divorce rates, increased life expectancy, and later age at marriage will probably increase the number of one-person and one-parent households as well as small complete families. There will be an increased number of small households comprised of elderly people living alone and elderly couples. The number of households headed by females will increase among the urban poor.

EMPLOYMENT AND THE LABOR MARKET

The proportion of men in the labor force has declined recently because of the increased number of working women and earlier retirement among men. The age profile of the labor market in 1985 will be characterized by the large number of young adults, male and female, either just entering the market or already established. Older women will probably continue to be a significant proportion of the labor market, but there is some speculation of a significant decline in the rates of labor force participation of young women as they reach childbearing age. Because there are proportionately fewer teenagers following the baby boom, thus putting less pressure on the teenage labor market, teenage unemployment should begin to decline by 1984.

COMMERCE AND INDUSTRY

Manufacturing is expected to continue to expand in the South and Southwest along with the labor force as young adults continue to move to these areas. It is further anticipated that a demand among young adults for consumer products, housing, and services associated with marriage and family formation will stimulate economic growth nationwide.

SOCIAL AND ECONOMIC IMPLICATIONS OF CHANGING AGE STRUCTURE AND POPULATION REDISTRIBUTION

Changes in the age profile of the population and shifting patterns of migration will require major changes in infrastructural services and facilities, changes in industry, and changes in the allocation of local, state, and federal resources.

Energy demands may shift as more people move to warmer climates and require less fuel for heating homes, but at the same time the newly dispersed populations will use more automotive products, gasoline, and roads.

Nationwide there will be less demand for goods, services, and public programs for the very young and a greater demand for services for the elderly. However, in the Southwest and South there will be a greater need for youth-oriented activities. At the same time northern cities will require more services geared to the poor and elderly.

In the next five years increases in the working-age population in relation to the size of nonworking-age population should lessen the burden on individual workers and temporarily ease the pressures on funding of social security pension systems.

If migration patterns continue, there will be an increase in the tax base of the nonmetropolitan areas and small cities gaining population and a decline in the tax base of the old cities. Supplying the goods and services needed in rapidly growing areas should prove less of a burden on public resources than meeting the needs of declining cities.

REFERENCES

1. Kirk, D. Population. *International Encyclopedia of the Social Sciences* 12:342–349, 1968.

2. Alonso, W. Metropolis Without Growth. *Public Interest* 53:68–86, 1978.

3. Taeuber, C. "Some Current Population Trends in the United States." Testimony before the House Select Committee on Population (February 9, 1978), p. 2.

4. Ford, K. Contraceptive Use in the United States, 1973–1976. *Family Planning Perspectives* 10(5):264, 1978.

5. *Monthly Vital Statistics Report*, National Center for Health Statistics, DHEW Pub 79-1120, 27(11 supplement):12, 1979.

6. Baldwin, W. Adolescent Pregnancy and Childbearing—Growing Concerns for Americans. *Population Bulletin* 31(2):36, 1977.

7. Ball, R.M. *Social Security Today and Tomorrow*. New York: Columbia University Press, 1978.

8. Long, L.H., and C.G. Boertlein. *Geographical Mobility of Americans*. Washington, D.C: Bureau of the Census, 1976, p. 14–15.

9. Alonso, *op. cit.* p. 73.

10. Berry, B.J.L., and D.C. Dahmann. *Population Redistribution in the United States in the 1970's* (NRC Assembly of Behavioral and Social Sciences). Washington, D.C.: National Academy of Sciences, 1977, p. 7.

11. Morrison, P.A. "Overview of Demographic Trends Shaping the Nation's Future." Testimony before the Joint Economic Committee, (May 31, 1978), p. 7.

12. *Ibid.* p. 18.

13. Adams, A.V., and G.L. Mangum. *The Lingering Crisis of Youth Unemployment*. Kalamazoo: W.E. Upjohn Institute for Employment Research, 1978.

14. Day, L.H. What Will a ZPG Society Be Like? *Population Bulletin* 33(3):8–14, 1978.

15. Easterlin, R.A. What Will 1984 Be Like? Socioeconomic Implications of Recent Twists in Age Structure. *Demography* 15(4):397–432, 1978.

16. Westoff, C.F. Marriage and Fertility in the Developed Countries. *Scientific American* 239(6):51–57, 1978.

17. Beale, C.L. "Recent U.S. Rural Population Trends and Selected Economic Implications." Testimony before the Joint Economic Committee (May 31, 1978).

18. Bahl, R. "The Fiscal Problems of Declining Areas." Testimony prepared for the House Select Committee on Population (June 6, 1978), p. 3.

19. Katzman, M.T. Implications of Population Redistribution: Education. *The Implications of Population Redistribution in the United States in the 1970's* (NRC Assembly of Behavioral and Social Sciences). Washington, D.C.: National Academy of Sciences (in press).

20. Perlman, R. Social Services and Population Redistribution in the 1970's. *The Implications of Population Redistribution in the United States in the 1970's* (NRC Assembly of Behavioral and Social Sciences). Washington, D.C.: National Academy of Sciences (in press).

21. Sternlieb, G., and J.W. Hughes. New Regional and Metropolitan Realities of America. *American Institute of Planners Journal* 43(3):230, 1977.

22. Greenwood, M.J. The Employment Policy Implications of Population Redistribution in the United States. *The Implications of Population Redistribution in the United States in the 1970's* (NRC Assembly of Behavioral and Social Sciences). Washington, D.C.: National Academy of Sciences (in press).

23. Mills, E.S. Population Redistribution and the Use of Land and Energy Resources. *The Implications of Population Redistribution in the United States in the 1970's* (NRC Assembly of Behavioral and Social Sciences). Washington, D.C.: National Academy of Sciences (in press).

8 Health of the American People

INTRODUCTION

Major progress has been made in public health in the twentieth century in the United States and other industrialized nations. Advances in sanitation—particularly improvement of water quality and less human contact with excreta—have improved health profoundly, especially in terms of infectious and parasitic diseases. Similarly, advances in food production and distribution have led to unprecedented quality in nutrition in the industrialized world, thus further diminishing susceptibility to infectious and other diseases.

Such improvements have been paralleled by advances in the practice and technology of medicine. Thus, the health care system can offer antibiotics for many infections, immunizations for viral and bacterial infectious diseases, and the detection of serious genetic defects in the human fetus through prenatal diagnosis. Encouraging progress has been made in developing pharmacological and other therapies for common chronic diseases such as hypertension and schizophrenia. Gross nutritional deficiency diseases have been virtually eliminated; for example, pellagra was a major cause of severe mental illness until it was found to be associated with a deficiency of niacin and protein and combated successfully on that basis. Important anemias have been brought under control. Many hormonal disorders can now be effectively treated.

ASSESSING THE NATION'S HEALTH

Progress in public health is evident in the statistics used to assess the health of the U.S. population. These vital statistics include mortality rates, life expectancy, and data on the incidence of serious illness.

Mortality Rates

Mortality rates in this country, excluding war deaths, declined steadily during the first half of this century. In 1900 the rate was almost 20 1,000 inhabitants. By 1940 the rate had fallen to 10.8 1,000 and in 1950 to 9.6 1,000. In 1975, the overall mortality rate dipped below 9 1,000 for the first time. In both 1975 and 1976, the U.S. mortality rate was 8.9 1,000. Even taking into account the continuing aging of the American population, the age-adjusted death rates declined 10 percent between 1950 and 1975.

The decline in infant mortality during the first year of life that was so marked in the first half of the century halted abruptly around mid-century:

The relatively poor progress in reducing infant mortality since the early 1950s has been a source of increasing concern in the United States. The subject has been examined previously in the context of international, national, and local changes in pregnancy loss rates; but it is clear that continued discussion based on the analysis of old and new data is very much the order of the day. . . .

In 1950 the assessment of the performance in the immediate past could well have led to an expectation of additional impressive gains in the future. Today the mood is quite different. For over a decade there has been no sizable decrease in the infant mortality rate. In fact, during the 1950s there were years in which the rate increased—a most unusual occurrence in half a century of vital statistics reporting in the United States. Events in the last few years give the definite impression that while the infant mortality rate will not remain stationary, its downward movement will be slow indeed.[1]

In 1975, the mortality rate for nonwhite male infants was 30.1,000, almost twice the rate of 15.9 for white male infants. Similarly, among female infants the rates were 25.2/1,000 for nonwhites and 12.2 for whites. The higher rates for nonwhite infants is made more telling by the fact that several industrialized nations have infant death rates even lower than those for white infants in the United States.

Major contributors to total mortality rates today are cardiovascular disease, cancer (mostly lung cancer), and accidents. Cardiovascular disease and cancer will be covered in more detail later. In recent decades, automobile accidents in the United States accounted for more than 50,000 deaths per year, a disproportionate number of them involving males between 15 and 24. In 1975, accidents, particularly automobile accidents, were by far the largest cause of death among young males. At 96.7 deaths per 100,000 young males, accidents far outpaced the second and third causes of death among this group. Accidents were also the chief cause of death among females between 15 and 24, amounting to 23.7 fatalities per 100,000.

Two final points concerning mortality rates should be mentioned. One is that they are higher for males than for females in every age group. The lower mortality rates for infant girls, as well as the lower rates of female deaths from accidents, are notable examples. There is also a sex difference in mortality from cancer and cardiovascular disease. Although sex differences in death rates are virtually universal and probably have a biological basis, they are especially evident in modern technological societies.

The second point about mortality rates is that at all ages the rates for nonwhite Americans are higher than the rates for white Americans. This is apparent in the rates for cardiovascular disease and cancer, as well as in the infant mortality figures noted earlier. Differences in mortality rates by race are strongly related to differences in socioeconomic status and can be viewed as partly reflecting the limitations of the health care system. Where socioeconomic differences are deliberately counteracted—as the U.S. Public Health Service has done through a highly organized and sustained effort in prenatal and perinatal care for American Indian mothers and their infants—both infant and maternal mortality rates among nonwhite citizens drop substantially.

Life Expectancy

Life expectancy is the average number of additional years that any one person can expect to live; typically it means life expectancy at birth. A female born in 1900 could expect to live, on the average, 51 years; a male born in the same year could expect to live 48 years. But a female born in 1975 could expect to live for almost 75 years, or 24 years more than the female born in 1900. Similarly, a male born in the United States in 1975 could expect to live to age 66, or 18 years more than the male born in 1900.

Improvements in life expectancy do not apply equally at all ages. For example, although an infant born in 1975 could expect to live about 20 years longer than an infant born in 1900, a person of 40 in 1975 could expect to live, on the average, only 6 years longer than the person who was 40 in 1900.

The increase in life expectancy in the United States during this century has been quite remarkable. By far the largest part of the increase occurred during the first half of the century. In the first two decades of this century, life expectancy at birth, regardless of race or sex, increased 4.6 years. Between 1920 and 1930 the increase was even larger, 5.6 years. Improvement during the next decade was smaller, 3.2 years by 1940; but larger by 1950, 5.3 years.

As with mortality rates, life expectancy showed relatively small improvement between 1950 and 1975. Similarly, there are significant differences in life expectancy at birth between male and female and between white and nonwhite Americans. Both white and nonwhite females can expect to live longer than white and nonwhite males. Finally, whites of both sexes can expect to live longer than nonwhites.

Male life expectancy in the United States is not outstanding compared with other countries. Male life expectancies in the Scandinavian countries (apart from Finland) surpass those in this country. As of 1975, U.S. male life expectancies also lagged behind those of other parts of Europe—including the Netherlands, Italy, Spain, Switzerland, England, Wales, Ireland, East Germany, Bulgaria, and France—and also of some countries outside Europe—Japan, Israel, Canada, New Zealand, and Cuba. However, the incidence of cardiovascular disease and its main known risk factors are declining in the United States, but are stable or increasing in most other industrialized nations. This almost unique trend promises to improve the relative standing of the United States in regard to male life expectancy.

American female life expectancies compare more favorably with those of other countries. The life expectancy for female infants born in this country in 1975 was surpassed in only eight other nations: Norway, Sweden, the Netherlands, France, Canada, Japan, Denmark, and Switzerland, in that order. One reason the United States lags behind other nations is the lower life expectancies among nonwhite Americans.

CHANGING PATTERNS OF ILLNESS

Since 1900, the pattern of fatal illnesses in this country has changed significantly (Figure 28). While this may reflect changes in the way we live, other factors can affect the ranking of specific diseases as causes of death. For example, if one disease is reduced in incidence, the next in line rises in order, without necessarily rising in incidence. Age is also a factor. If fewer people die of pneumonia, more will live to die of cancer.

Most noteworthy has been the rise of arteriosclerotic heart disease and cancer. In 1900, the leading causes of death in the United States were pneumonia, influenza, and tuberculosis—all infectious diseases. By 1940, heart disease had become the leading cause of death, with cancer second.

Other changes in the pattern during this century are also significant. Tuberculosis, once one of the most dreaded diseases and the seventh-ranking cause of death in 1940, was no longer among the 10 most important diseases by 1960. Two other leaders in 1900—diarrhea and other intestinal infections and diphtheria—were no longer among the 10 leading causes of death by 1940. Nephritis (a kidney disease), the fifth leading cause of death in 1940, had disappeared from the top 10 by 1960.[2]

However, two other causes of death—strokes and accidents—have maintained a fairly steady position among the top 10 throughout the century, although the incidence of death from strokes is decreasing. Strokes were the fifth leading cause of death in 1900 and the third in 1940, 1960, and 1970. In 1900, accidents as a single category were the seventh leading cause of death; in 1940, when motor vehicle accidents were listed separately from other accidents, the rates for the two combined made them fourth, the position maintained by accidents of all types through 1970.

Finally, certain diseases only recently appeared among the leading causes of death. These include diabetes mellitus, eighth in 1940 and seventh in 1970; arteriosclerosis, seventh in 1960 and eighth in 1970; cirrhosis of the liver, tenth in 1960 and ninth in 1970;

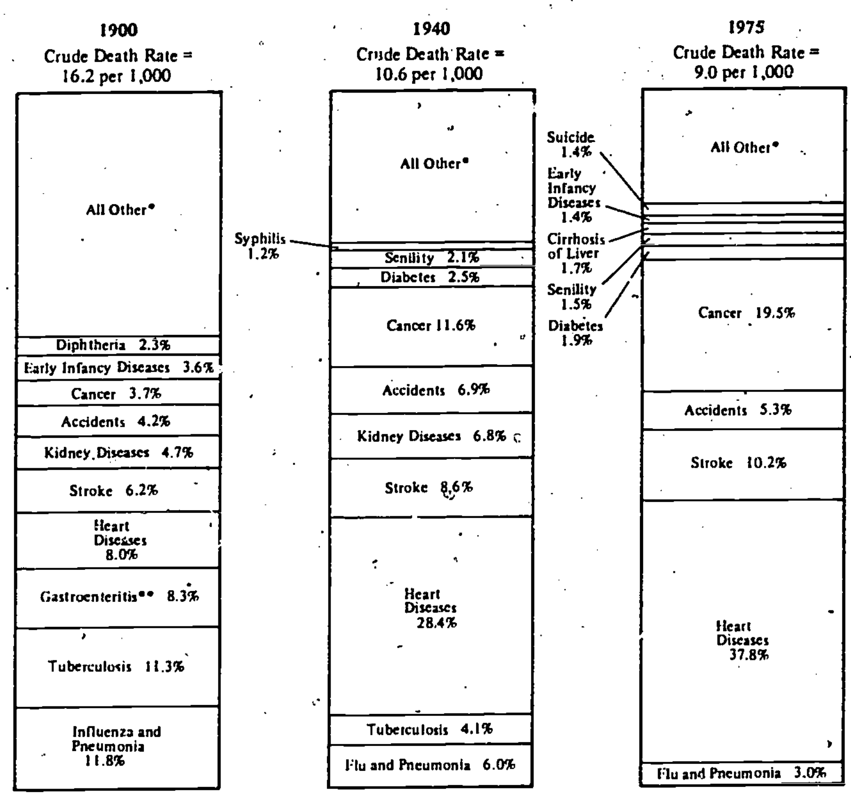


FIGURE 28 Major causes of death in the United States (1900-1975). *No disease in this category represents more than 2 percent of all deaths. **Inflammation of the stomach and intestines. (Courtesy of the Population Reference Bureau, 1337 Connecticut Ave., N.W., Washington, D.C.)

and bronchitis, emphysema, and asthma, tenth in 1970.

In very general terms, then, the pattern among the 10 leading causes of death in the United States during the first eight decades of the twentieth century has been marked by the ascendancy of cardiovascular disease (heart disease, strokes, arteriosclerosis) and cancer, the upsurge of diabetes, cirrhosis, emphysema, and asthma, and a substantial decline in infectious diseases. The rise of cardiovascular disease and cancer to their present positions has been a long one. Even in 1900, diseases of the heart were the fourth leading cause of death in this country and cancer eighth. Moreover, while cardiovascular disease is the leading cause of death, it is now declining; this development will be considered later in the chapter. And while infectious diseases have declined in incidence, they have not disappeared from the top 10. As late as 1970, influenza and pneumonia together were still the fifth leading cause of American deaths.

The point is that the patterns of serious illness are far from immutable. Medical attacks on the principal health problems of the past have been remarkably successful. There is no inherent reason why the incidence of such current afflictions as heart disease and cancer cannot be diminished through sound basic research and the application of the resulting knowledge. For example, advances in the treatment of cardiovascular disease, combined with changing public behavior with respect to risk factors, have contributed to the average decline of 2 percent a year in the overall cardiovascular mortality rate for American men between the ages of 55 and 64 during the past decade. There are indications that this trend is extending into older age groups.

BURDEN OF ILLNESS

The burden of any illness depends on both its clinical and its economic impact.[3,4] Clinical impact is determined by how many people suffer and die; in clinical terms, heart disease or cancer are clearly more important than allergies. Economic impact has two components: the cost of providing diagnosis, treatment, and care; and the productive work lost through illness. Respiratory diseases have a particularly heavy economic impact. Ranging from pneumonia and flu to the common cold and sinusitis, they occasion more visits to doctors and more days lost from work than any other illness.

While we have long known in general terms the total cost of providing medical care for serious diseases, little effort has been made until recently to distinguish the clinical importance of diseases from their economic importance. The burden of illness is most commonly measured in mortality rates, which describe only clinical impact. In addition, mortality rates, life expectancies, and statistics on the incidence (number of new cases per unit of time, usually a year) and prevalence (number of cases in the population at a given time) of life-threatening illnesses tell us nothing of the cost and clinical impact about such common but usually nonfatal problems as dental caries, arthritis, hay fever, the common cold, blindness, deafness, schizophrenia, ulcers, and others. We must also understand the cost of these problems.

The National Center for Health Statistics, of the Department of Health, Education, and Welfare, has been paying increasing attention to standardized assessments of illness in terms of both clinical and economic burden. Specific annual indices developed for this purpose include potential years of life lost (age at death compared with average life expectancy), inpatient (hospital) days, outpatient (primary-care) visits, and work-lost days associated with various categories of illness. Such new indices of the burden of illness demonstrate the heavy impact of respiratory disease. Similarly, mental illness and emotional distress are high on several indices of burden, although they are not prominent in mortality.

The next decade is likely to see considerable improvement in the reliability of clinical and economic measures of burden of illness. These measures can help to make more rational allocations of resources for research, education, and services.

CARDIOVASCULAR DISEASES

As infectious diseases came under increasing control during the middle third of the twentieth century, diseases of the heart and blood vessels became more prominent. These cardiovascular diseases are now the leading cause of death in the United States, accounting for more than half of all deaths in 1976. They are also responsible for a heavy burden of illness and economic loss, ranking first as a cause of limited activity, Social Security disability, and hospitalization.

Cardiovascular disease and the availability of research tools have stimulated basic research and clinical investigation into the cardiovascular system and its disorders, including congenital heart disease, rheumatic heart disease, hypertension, stroke, and coronary artery disease.[5] New screening, diagnostic, and monitoring techniques have emerged, along with effective drugs, procedures for repair or replacement of diseased blood vessels, surgical repair of the heart, antibiotics for the prevention of rheumatic heart disease, and public education programs on cardiovascular risk factors and their management.

Taken together, such advances have contributed to the recent decline in mortality from cardiovascular diseases. While they are still the number one killer in this country, the rates of death from cardiovascular diseases have fallen more than 30 percent since 1950, with the decline most rapid in the past 10 years. Cardiovascular mortality has diminished for men and women of all ages and races. Between 1970 and 1976, the age-adjusted decrease was 22 percent for stroke and 16 percent for coronary artery disease. In 1975, deaths from all cardiovascular diseases fell below 1 million for the first time since 1963.

Great gains have been made in understanding the causes and treatment of these diseases, and promising medical and technological developments will be pursued for further gains. However, we know that many cardiovascular diseases are "silent." They develop slowly for many years, perhaps even from childhood, until manifested as an acute episode—a heart attack or stroke. In the long run, we need early warning signals and effective preventive measures for these diseases.

CONGENITAL HEART DISEASE

About 25,000 infants with congenital heart defects are born each year; 3,000 die before their first birthday. Each year about 6,000 persons of all ages die of congenital heart disease, a 38 percent decline in mortality since 1948. Congenital heart problems can result from either genetic or environmental factors, the latter including a preventable disease, German measles.

Until quite recently, babies born with deformed hearts were largely beyond the reach of effective treatment. But developments in biomedical engineering, physiology, and surgery, from 1950 onward, made effective operations possible. The heart–lung machine, cooling and other life-support techniques, artificial heart valves, and improvements in anesthesia all contributed to this growing competence. Moreover, sensitive diagnostic procedures gave clear guidance for corrective surgery. Follow-up studies indicate that such surgery has become steadily more effective while the risks have diminished.[6]

Nonsurgical techniques are being explored for conditions previously amenable only to surgery. One type of congenital heart defect may be modified in newborns (simplifying later surgery) by a medication that promotes normal development of the heart. This therapy works by inhibiting a prostaglandin, one of a family of biologically active organic acids that occur naturally in the body. The biochemistry of the prostaglandins has been elucidated in the past two decades, and they are emerging as a family of compounds with great functional significance for human health.

RHEUMATIC HEART DISEASE

Painstaking clinical research has clarified the sequence of events in which a childhood streptococcal infection, if not treated, may be followed shortly by rheumatic fever, which, in turn, can damage the heart valves. One mechanism involves immune reactions to streptococcal antigens. When the underlying processes are more deeply understood, another opportunity for preventive medicine may appear.

Rheumatic heart disease was a significant source of disability and death among young adults in the first half of this century. While it is far less prevalent and rarely fatal today, there are still 100,000 new cases each year, and more than 1 1/2 million adults have rheumatic heart disease. Penicillin is administered to patients who have had rheumatic fever to block recurrence of streptococcal infection; the antibiotic may be given for years or even a lifetime. This practice has sharply reduced disability and death from the disease. Also heart valves damaged by rheumatic fever can be replaced through a remarkable surgical development.

Recent studies have shown that well-organized health programs for children with rheumatic heart disease can decrease days of illness and hospital time, and prevent later complications of childhood diseases such as streptococcal infections. In one such program, prompt and appropriate use of antibiotics reduced the incidence of rheumatic fever by 60 percent.

HYPERTENSION

Hypertension, or high blood pressure, develops insidiously over many years before damage becomes apparent in such forms as stroke, heart attack, heart failure, or kidney failure. It affects an estimated 35 million adults.

Basic research in physiology, biochemistry, pharmacology, genetics, and behavior—applied to the renal, cardiovascular, endocrine, and nervous systems—has increased our knowledge of hypertension. We have learned much about its clinical signs, consequences, prognosis, and treatment. However, in more than 80 percent of the cases we still don't know its cause and call it "essential" hypertension. We need to know more, for example, of the role of dietary factors such as weight and salt intake. Moreover, hypertension is more frequent among blacks than among whites in this country, and we do not know why. Elucidation of the specific genetic and environmental factors implicated in this major public health problem remains a major scientific challenge.

Drugs for Hypertension

Drugs now available for high blood pressure can greatly reduce the risk of stroke and kidney and heart failure. However, long-term follow-up studies indicate that diuretics, the drugs used most often to control blood pressure, may also raise cholesterol and sugar levels in the blood of most individuals. Thus, while the diuretics are undoubtedly effective in lowering blood pressure and reducing its complications, especially strokes, they may increase the risk of other disorders. This preliminary finding deserves further investigation. It also demonstrates the value of long-term follow-up in clinical research generally and in clinical pharmacology specifically.[7]

The awareness of the limitations of diuretics provides a powerful stimulus for creating and testing new drugs, especially in the light of advancing knowledge pertinent to control of blood pressure. One active and promising line of inquiry involves drugs that diminish the activity of the sympathetic nervous system. This part of the nervous system can constrict the small arteries and thereby elevate blood pressure. It can be "slowed down" either by drugs that act directly on the sympathetic nervous system throughout the body or by drugs that act on its central controls in the brain. Some drugs that inhibit the sympathetic nervous system are now in regular clinical use; others are being tested experimentally. Improved drug therapy for hypertension should become available in the next five years. It will take longer to determine exactly which therapeutic regimens most effectively prevent such complications as stroke or heart failure.

Biobehavioral Approaches

Recent work on behavioral approaches to controlling blood pressure has also been promising. The Stanford three-community study, discussed below, showed that people can learn to reduce their consumption of salt, which the study found to be the best predictor of blood-pressure change among the several behavioral factors, such as weight reduction and relaxation, that were investigated. Use of biofeedback learning techniques to control blood pressure is also receiving attention.

Unknowns in Hypertension

The risk from high blood pressure occurs over a broad range of pressures—there is no sharp cutoff between "normal" and "high." More information is needed about the effects of gray-zone pressures—between the clearly elevated and those accepted as normal. Primary preventive measures are needed to stop or slow the frequent increase in blood pressure with aging. There is no reason to assume that such an increase is inevitable. Evaluation of the benefits of medication in elderly and young people is needed, as is evaluation of the benefits and costs of medication of mild hypertension. More must be known about the basic processes of hypertension to improve both pharmacological and behavioral prevention of the disease and its complications.

CORONARY ARTERY DISEASE

Coronary artery disease involves damage to blood vessels leading to the heart muscle and subsequent angina, myocardial infarction or heart attack (death of heart-muscle cells), heart failure, or arrhythmia (irregularity) of heartbeat. Coronary artery disease usually can be traced to arteriosclerosis, narrowing of the channels of the blood vessels that restricts blood flow. Coronary artery disease, usually manifested as a heart attack, accounts for 650,000 deaths annually, or two-thirds of the total deaths from cardiovascular disease.

The prevalence of coronary artery disease has stimulated much clinical innovation, apparent especially in the nationwide development of hospital coronary-care units. In such units, sophisticated electronic monitoring permits early detection of heart-rhythm abnormalities; also, drug therapies and modern equipment for electrical stimulation and resuscitation are immediately available. Artificial pacemakers are used for immediate or long-term control of heart rhythm. Techniques for assessing the extent and location of damage to heart muscle have been improved. Surgical bypass of seriously diseased coronary blood vessels has become common and in certain cases can relieve pain.

Despite the technical sophistication and intrinsic plausibility of these approaches, we are still not certain whether extended treatment in a coronary-care unit gives better overall results than careful treatment at home after initial hospital treatment. Nor do we know whether coronary bypass surgery has any effect in addition to the highly valuable one of relieving pain. In view of the gravity and cost of these procedures, we must learn to determine more precisely which coronary disease patients are appropriate candidates for given therapeutic approaches.

PREVENTION OF CARDIOVASCULAR DISEASE

Prevention of cardiovascular disease is being increasingly emphasized. The contribution of basic research to prevention probably will increase over the next decade, particularly in regard to hypertension and arteriosclerosis, the two main causes of damage to the heart and blood vessels.

More than half of the Americans with high blood pressure are unaware of their condition. And of those who are aware, about half simply do not adhere to the prescribed treatment. It is important to learn, through health-services research, why patients do not take medication that is good for their health. Such research may produce effective ways to help patients modify their behavior, not only in regard to medication, but also in regard to such health-related habits as eating, smoking, and drinking.[8]

We already know that aside from the cost of drugs, failure to take medication depends partly on such factors as waiting time for the doctor and at the clinic; inadequate follow-up by doctors and clinics; overcomplicated and confusing dosage schedules; unsatisfactory doctor-patient relationships; and inadequate explanation of possible side effects. The last problem is particularly significant to patients who had no symptoms prior to medication. Recently, there has been some careful work on the utility of nonphysicians who can counsel on medication problems. Such counselors can markedly enhance adherence to therapeutic regimens.

Public Awareness

The National High Blood Pressure Education Program was initiated in 1972 as a result of survey research in the early 1970's that showed the low public awareness of high blood pressure and its consequences.[9] A far-reaching collaborative effort between the National Institutes of Health and various private organizations, it involves screening of the public for high blood pressure and careful follow-up. From start to present, the program has sharply raised the numbers of people using effective treatment for hypertension. A survey in 1977 of 100,000 residents of Chicago showed that 9 of 10 people with elevated blood pressure were aware of their condition and that about 60 percent of hypertensives were being effectively treated. This reflects a striking increase in recent years in both awareness and treatment.

Voluntary health organizations, health professional organizations, industry, consumer groups, and the federal government have worked together effectively in this educational program. Since hypertension is especially prevalent among blacks, a special effort in minority communities has been mounted in recent years. In view of the heavy burden of illness in these communities, the lessons learned from the National High Blood Pressure Education Program may well be useful.

CARDIOVASCULAR RISK FACTORS

Sudden death is the first symptom in at least one-fourth of those with coronary disease. This fact is a powerful stimulus to research on primary prevention and on noninvasive diagnosis before overt symptoms appear. Attention is turning increasingly to the detection of cardiovascular risk factors and means of reducing them before the disease becomes clinically apparent.

Cardiovascular risk factors include age, male sex, elevated blood pressure, cigarette smoking, elevated plasma cholesterol, elevated blood glucose, obesity, sedentary way of life, water hardness, family history of heart disease before age 65, personality type, and severe stress.[10,11] Not all of these factors have been investigated with equal thoroughness, but each has some predictive power for coronary disease.[12]

Epidemiologists have done large-scale prospective studies to delineate the biological and behavioral characteristics of individuals who are likely to develop cardiovascular disease; here considering hypertension and coronary heart disease separately and also their relationship with each other. These careful, systematic studies have been conducted with populations in different parts of the United States and in other countries.

Cholesterol

Particular attention has been given to cholesterol by the public and by clinical and basic researchers. Cholesterol in the plasma is part of several more complex structures, lipoproteins, which can be separated on the basis of their density. This density analysis has improved the predictive value of cholesterol as a risk factor: Either elevated total-plasma cholesterol or elevated low-density-lipoprotein cholesterol, which correlate closely with each other, tends to be harmful; elevated high-density-lipoprotein cholesterol, which does not correlate well with elevated total-plasma cholesterol, appears to protect against coronary heart disease.

A recent report of a World Health Organization clinical study showed that a drug (clofibrate) lowered lipids such as cholesterol in the blood, and this in turn was linked to a decrease in heart attacks. However, the status of the drug in clinical therapy is not yet clear.

Genetic Effects

A relationship between dietary saturated fats, plasma cholesterol levels, and arteriosclerosis has been demonstrated by many laboratory experiments in a variety of animal species. These relationships have been confirmed in humans: The nature and extent of fat in the human diet has a bearing on arteriosclerosis. But not everyone who consistently eats a diet high in saturated fats develops serious disease.

Genetic differences affecting individual responses to different kinds of dietary fats apparently have an important bearing on the extent and severity of the disease. Such genetic influences are illustrated by a recent discovery. High concentration of fat in human blood, or hyperlipidemia, is strongly influenced by certain genes. Indeed, three distinct single-gene disorders can predispose to hyperlipidemia.[13] One is called familiar hypercholesteremia, the second hereditary hypertriglyceridemia, and the third familiar "combined" hyperlipidemia. One or another of these conditions is found in a sizable minority of patients under age 60 who have heart attacks. Evidently, individuals with these particular genetic disorders are more vulnerable to a high-fat diet.

Modifying Risk Factors

Interest is high in modifying cardiovascular risk factors, especially among those who appear to be especially susceptible genetically. The task is not easy and will require significant educational and social changes. However, there are encouraging indications that better health is possible by altering firmly established patterns of behavior, such as smoking, exercise, diet, working, and coping with stress.

Significant changes in risk factors have occurred in the United States in the past 15 years, at least in part because of public concern with health. Consumption of tobacco products, milk and cream, butter, eggs, and animal fats have all declined among adults. The decline is greater in the more educated segments of the population and is especially striking among health professionals. These declines coincide with an accelerating decline in cardiovascular mortality. While the relationship may be a coincidence, it should be investigated. One point deserves emphasis: The link between diet and cardiovascular disease is suggestive,[14] but present evidence leaves little doubt that stopping cigarette smoking will reduce the risk of cardiovascular disease, and of lung cancer.

The Stanford Program The Stanford Heart Disease Prevention Program,[15,16] a multifaceted research effort to learn how to combat heart disease, has shown that it is possible to decrease risk factors for cardiovascular disease through health education using the media. The risk factors addressed in this two-year program in the mid-1970's were cigarette smoking, high plasma cholesterol concentrations, and high blood pressure. Three communities were involved: one with only a mass-media education program, one with a mass-media program supplemented by face-to-face counseling for high-risk individuals, and one control community with no special programs. All three communities were surveyed annually. Key aspects of the experiment were:

- The mass-media materials were devised to teach specific skills—for example, preparation of a palatable low-fat diet—and to provide information on health and motivate people to use it.
- Behavioral scientists advised on mass-media approaches and face-to-face instruction.
- The campaign was suited to the intended audience; for example, it took account of the fact that part of the audience was Spanish-speaking.

Among high-risk participants who received the annual survey, mass-media education, and intensive instruction, the overall risk of cardiovascular disease was reduced 30 percent. Almost all of the reduction was achieved in the first year of the program and sustained through the second year. High-risk individuals who received only the annual survey and mass-media education reduced their risk by 8 to 10 percent the first year and by 25 percent after two years. High-risk participants in the community exposed to the survey alone did not appreciably change their risk of heart disease.

The Finnish Program A larger study of the effectiveness of reducing risk factors has been in progress for six years in North Karelia, Finland.[17] The three major components of the program are: health education through community resources, including local newspapers and radio; hypertension screening with intensive group-health education of high-risk individuals; and early diagnosis, treatment, and rehabilitation involving existing health and social services. Education and medical measures are supplemented by a law forbidding smoking in public buildings and on public transport and by the cooperation of the local dairy and food industries in reducing fat in popular foods.

As a result of these efforts, the annual rates of incidence of acute myocardial infarction, which had been increasing in Finland for many years, have begun to decline in North Karelia. There also has been an apparent decrease in the severity of heart attacks. In addition, the annual incidence of strokes had fallen by the third year of the program.

Community-Based Prevention. The experiences in California and Finland will stimulate research on community-based prevention in the next five years and should be useful in planning new programs. Such investigations must explore ways of eliciting the interest and cooperation of both the public and private sectors in facilitating health-promoting behavior. If the findings thus far are duplicated in other communities, a sustained change in the methods of health education is likely.

CANCER AND RELATED PROBLEMS

Although cancer is not the most common serious disease, it is certainly the most feared. It is therefore not surprising that the National Cancer Institute (NCI) is the largest health-research organization in this country. During the 1970's NCI expanded very rapidly under the auspices of the "War on Cancer."

New leadership is now examining this enormous effort in order to achieve a balanced program sensitive to scientific opportunities, clinical needs, and social concerns. This constructive reassessment will be an important feature of science policy in the next few years.

Since cancer is also considered in other chapters, this section will be relatively brief and clinically oriented, with special attention to treatment and prevention.

NATURE OF CANCER

Cancer has many forms. All involve a defect that permits the unrestrained multiplication of cells to yield offspring cells whose growth is similarly unrestrained.[18] Some believe that the basic cell defect is the same for all cancers, but that it may be triggered in diverse ways to produce unrestrained growth.

Cancer is no longer viewed as an inevitable concomitant of aging. Environmental factors acting on genetic predispositions are now considered important in the origin of cancer. In this context, environmental factors are defined broadly and include, for example, diet and cigarette smoking. The growing recognition of environmental influences is based in part on large geographic variations in the incidence of specific cancers, linked with evidence that migrant populations tend to shift to the patterns of incidence of their new region.

Two lines of inquiry, which may have long-term clinical significance, have not yet been sufficiently resolved to assist in diagnoses, prevention, or therapy. These involve immune responses in carcinogenesis and the role of viruses in human cancer. One hypothesis is that carcinogenesis, or production of a cancer, occurs normally throughout life (perhaps in part owing to background radiation), but that it is kept under control by the immune system. Viruses are known to cause some cancers in animals—leukemia in cats is one example. But despite considerable research, it has not been proved that viruses cause cancer in humans. In the long run, research in both immunology and virology will clarify mechanisms of carcinogenic transformation and provide clues to how this transformation might be inhibited by pharmacological or immunological means.

Thresholds and Synergism

An important question is the level of exposure to carcinogenic substances that initiates carcinogenesis. Most toxic substances fed to test animals produce no observable effect or response below a characteristic threshold dose. However, a threshold dose for carcinogenesis has not been demonstrated for any substance; if threshold doses exist, they cannot be detected by the available toxicological methods. If carcinogens do not in fact have threshold doses, exposure to a very small amount of a carcinogen could initiate carcinogenesis, although the probability that it would do so would be slight; known carcinogens must be fed to 50 or 100 test animals in quite large amounts for months to induce cancer in only a few of them. Quantitative dose–response data are available for some agents that affect the incidence of cancer in humans. These agents include ionizing radiation and cigarettes. For each, the data are consistent with the absence of a threshold dose; if one exists, it is undetectable by current toxicological methods.

Aside from the probable absence of a threshold for carcinogens, synergism must also be considered; that is, the combined effect of two agents may be far greater than the sum of their individual effects. An example of this compounding of risk is the combined effect of cigarette smoking and exposure to asbestos. Smoking alone increases the risk of lung cancer about 10-fold—the actual increase depends on the number of cigarettes smoked. For asbestos workers who do not smoke, the risk of lung cancer is 7 times that for nonsmoking, nonasbestos workers. However, for an asbestos worker who also smokes, the risk of lung cancer is more than 12 times that for a nonsmoking, asbestos worker and 90 times that for a nonsmoking, nonasbestos worker.

Thus, current evidence provides no basis for assuming that there is a totally safe level of any carcinogen. In real life, however, it is often impossible to avoid completely a carcinogen. The establishment of socially acceptable exposure levels will require dependable, quantitative risk estimates based on better data than now available. It should also be kept in mind that cancer can develop in the absence of any known external carcinogens.

EPIDEMIOLOGY AND CANCER

Epidemiology has a long history of accomplishment in the understanding and control of infectious diseases through public-health measures. Today it is being applied increasingly to cancer, cardiovascular disease, and mental illness. The epidemiologist attempts to determine the cause of a disease by comparing possible causes statistically with the incidence and distribution of the disease in a population until a correlation emerges. With cancer the task is difficult for several reasons: physical, chemical, biological, and social, environments are very complex; many presumed carcinogens are present; exposure conditions are highly variable; and latent periods after exposure are very long. However, high exposure simplifies the problem. Epidemiological studies of groups of people exposed to unusually high doses, as in an industrial setting, have led to the identification of several dozen human carcinogens.

Large-scale epidemiological research will be required to correlate the changing patterns of environmental agents with parallel changes in the incidence of cancer. Special attention should be devoted to those cancers that occur most frequently and with large variations in geographic incidence (for example, gastrointestinal cancers).

Although some man-made chemicals have been identified as potential carcinogens, the firmly established incidence of cancer from these compounds accounts for only a tiny fraction of all cancers.[19] Cigarette smoking is the one environmental factor for which firm data demonstrate a strong association; occupational exposures are also an important source of environmental carcinogenesis. Other environmental factors, including those of natural origin, that are probably responsible for many and perhaps most cancers remain unknown. Research on these problems will intensify during the next five years.

DETECTION OF CANCER

Skin cancers are easily detected and therefore quickly treated. Deep-lying cancers, such as those in the lung, typically are detected late, and treatment is less effective. Special effort is being devoted to earlier detection of cancer in various deep-lying locations. One of the most successful screening techniques has been the "Pap smear," in which a few cells from the uterine cervix are examined. This test has contributed to a significant decrease in the death rate from cancer of the cervix among American women. The mortality rate dropped over 50 percent between 1948 and 1971, falling from 38 to 15 per 100,000 for white women and from 75 to 31 per 100,000 for nonwhite women.

The most efficient timing of screening tests is currently under study. Timing is important, for example, in mammography, a radiologic diagnosis for breast cancer. This disease affects nearly 1 woman in 13 and kills 30,000 annually. Regular screening of women by mammography during childbearing years probably exposes them to more radiation than the yield of discovered cancer is worth—the risks outweigh the benefits. But, the risks of mammography after age 50 may well be outweighed by the benefits of early detection of breast cancer. Such concern has led to technological improvement: dosage is about half what it was at first.

The new fiber-optic endoscope—an instrument for looking at the inside of a hollow organ—is much more flexible than instruments previously available, and therefore facilitates exploration for cancer of the lung and colon—two of the most common cancers. Similarly, the relatively noninvasive radiologic technique of computed tomographic scanning can detect tumors early, especially in the head. Other techniques of medical imaging for noninvasive diagnosis are under intensive development.

The immunology of tumors may become clinically useful in the next 5–10 years. Malignant tumors synthesize distinctive antigens, some of which are released into the circulating blood.[20] These antigens now are detectable in the blood only at an advanced stage of cancer. But the immensely sensitive and specific technique called radioimmunoassay is becoming increasingly useful for such analyses.[21] The method combines immunologic techniques with the measurement of substances labeled with radioisotopes. Its use to detect antigens in blood is especially promising for early diagnosis of gastrointestinal cancer.

TREATMENT OF CANCER

Surgical removal is still the best treatment for most cancers. If all the malignant cells can be excised, the cancer may not rec . Unfortunately, these abnormal cells commonly can invade surrounding tissues and, worse yet, spread through the blood or lymph systems to form new growths at distant locations. If this has occurred even microscopically by the time of initial diagnosis, removing the primary tumor will only buy time. A recurrence elsewhere is likely. Therefore, surgery is often supplemented by radiation therapy or chemotherapy, or both.

Surgery has a sharply focused target, radiation is usually directed at a specified region near the site of the tumor, and chemotherapeutic agents reach cells throughout the body. The well-established combination of surgery and radiation therapy is capable of curing (in terms of five-year survival) about one-third of all cancer patients. This fraction does not include patients with skin cancer, which is an easier problem.

Chemotherapy increasingly is being used after surgery when recurrence is especially likely. It works best when relatively few cancer cells have spread through the body. For the most part, chemotherapy is relatively new, and its long-term effectiveness is not yet clear. The technique does seem to lessen the risk of relapse in one form of bone cancer and may be helpful in some forms of breast cancer. Current research involves combined use of two or more chemotherapeutic agents, each of which has shown anticancer potency, at least in the short run. These chemotherapeutic combinations are also being tested in combination with established and newer forms of radiation therapy. Radiotherapists not only are trying new types of beams but also are combining them with radiosensitizers—chemical agents that selectively enhance the sensitivity of tumor cells to radiation.

A notable success of chemotherapy, particularly when combined with radiation therapy, is the treatment of Hodgkin's disease, which in the past was uniformly fatal; similar gains are being made in the treatment of acute lymphatic leukemia in children.[22] In both cases about half the patients can be cured. While these are certainly encouraging, indeed groundbreaking advances, these cancers account for less than 1.5 percent of all cancers.

PREVENTION OF CANCER

Given the gravity of cancer, the difficulty of early detection, and the limitations of treatment, it is crucial to ask what science can contribute to prevention. The past quarter century has seen unprecedented progress in the life sciences, and medical scientists should examine systematically the implications of these extraordinary advances for prevention of disease in general and of cancer in particular. A significant amount of cancer may be preventable by use of existing knowledge. Much can be accomplished by identifying risk factors and learning to modify exposure to them—as the record abundantly demonstrates. As we learn more and more of the mechanisms that underlie risk factors and their modes of action, systematic efforts can be made to use that knowledge for prevention.[23]

Identifying Carcinogens

A multifaceted approach will be necessary to reduce the cancers influenced strongly by environmental factors. This includes the use and further refinement of methods for identifying carcinogens, including *in vitro* screening tests, tests in mammals, and epidemiological studies. *In vitro* tests in particular warrant rapid development because they cost far less in time and money than do tests in laboratory animals. Chemicals that test positively for carcinogenicity should be seriously suspected of being carcinogenic in humans. However, it is still considered necessary to confirm the results of *in vitro* tests by tests in animals.

Chemicals shown to be carcinogenic in laboratory animals are likely to be carcinogenic for humans; in several well-documented cases, a carcinogen was identified in animal tests before its carcinogenicity for humans became clear. Conversely, with few exceptions, such as arsenic, substances that cause cancer in humans also do so in animals. The correlation is good but not perfect from animals to man and from man to animals. Animal tests, therefore, are very good but not infallible predictors of carcinogenic potential for humans. Also, animal species may vary widely in their sensitivity to a particular carcinogen, so that a quantitative estimate of degree of risk (or potency) can be extrapolated from animals to man only with considerable caution.

The several hundred compounds proven carcinogenic in animals should be viewed as potential carcinogens in man unless proven otherwise. Research on them is well justified; also the populations exposed to suspected carcinogens should be studied to provide better data for assessing risk to humans.

CIGARETTES AND HEALTH

Of all the opportunities for preventing cancer in the foreseeable future, the most important by far lies in cigarette smoking.[24] If research can help to markedly lower the number of cigarette smokers, gains will be made not only in prevention of lung and bladder cancer, but also of serious respiratory diseases and major cardiovascular diseases. Here then is a truly critical problem in health.

(Bill Gillette, EPA Documerica)

Cancer of the lung, the cause of 92,000 deaths annually, was uncommon in the early part of this century. It began to rise sharply around 1935 in men and 1965 in women. Each increase began about 20 years after cigarette smoking became widespread among the sex. The increase in deaths from lung and other respiratory cancers has been so large that, if these deaths were excluded, total cancer rates among persons aged 55–64 would drop approximately 40 percent for men and 15 percent for women.

A 20-year epidemiological study of mortality in relation to smoking among British physicians casts light on both the consequences of smoking and the benefits of stopping. The data indicate that between a third and a half of all cigarette smokers die because of their smoking. Deaths in the study were due chiefly to heart disease among middle-aged men, but lung cancer was also a prominent cause. As a whole, the population reduced its cigarette consumption substantially during the 20 years of thorough study. At the same time, the incidence of lung cancer became less common among the population. This and other studies indicate that pathology related to cigarettes is at least partially reversible, even after years of smoking. But here, as elsewhere, prevention is likely to be most beneficial to long-term health.

Children and nonsmoking adults may be exposed to the smoke of others. The unborn infant particularly needs protection, since it may suffer increased risk of death or retarded growth with long-term effects if the mother smokes. Increased risk of respiratory infections occurs among infants whose parents smoke, and these infections may predispose to adult chest disease.

INCIDENCE OF SMOKING

Knowledge of the risks of smoking has resulted in a decline in smoking among adult Americans. More than 30 million smokers have quit since the Surgeon General's report, *Smoking and Health*, was published in 1964. Sharper declines in smoking than in the general population have occurred among doctors, dentists, and pharmacists—professionals who daily observe the long-term effects of cigarettes on patients and customers.

Thirty-five percent of American adults smoke cigarettes. The proportion of teenagers who smoke has been increasing, especially in the 10–15 age-group and above all in females in this age-group; but there are indications of a very recent decline in this trend. Surveys show that Americans know of the relationships between smoking and lung cancer, but not that tobacco-related cardiovascular diseases take a heavier toll.

During the past 20 years, many clinics, techniques, and devices have been developed to help smokers quit. While several methods are effective for a short time, most smokers eventually resume smoking. The prevention of such relapses has been tackled recently as a problem that requires solutions different from those that help smokers first quit the habit, and will likely be given high research priority in the next five years.

Techniques based on learning principles show promise for stopping smoking permanently, but are still at an early stage of experimentation.

LOW TAR AND NICOTINE

There has been considerable effort in recent years to devise less hazardous cigarettes. Usually, lower hazard is attributed to "low tar and nicotine" in the inhaled smoke. While the low-tar-and-nicotine approach deserves systematic investigation for both risks and benefits, the facts are far from clear since cigarette smoke contains a variety of toxic substances. For example, carbon monoxide may be much more important in causing cardiovascular disease than in causing cancer.

SMOKING AND ADOLESCENCE

Adolescence is a critical period of biological and psychological change, and also of drastic change in social environment. And it is a time when lifelong behavioral patterns are formed involving cigarette smoking, use of alcohol and other drugs, automobile driving, habits of diet and exercise, and patterns of human relationships. Adolescence has been neglected until recently in biomedical and behavioral research.

The links between puberty and the endocrine hormones can now be studied with high precision as a result of recent advances. These include especially the discovery of the brain hormone that controls the reproductive system in both sexes and the development of radioimmunoassay methods for measuring hormones and their derivatives. Knowledge is needed relating endocrine and bodily changes of puberty to parallel emotional and behavioral changes.

Peer Counseling

The impact of peers, influential adults, and the mass media on adolescents needs careful, systematic study to clarify the determinants of behavior harmful to health. Peer counseling is one promising approach to guidance during adolescence. Such programs rely on the credibility of peers during adolescence in training students to help other students. The approach has been used in junior and senior high schools, and the findings suggest that such programs can be useful for both students and counselors.

Peer counseling is now being tested as a means of discouraging the onset of smoking. The test[25] is a field experiment, based on a model curriculum, in which 16-year-olds teach 12-year-olds how to resist peer pressure to smoke. Hundreds of students in matched experimental and control schools are compared for cigarette-smoking rates at the start of the test and during follow-up periods of up to two years. So far, the results indicate much less smoking among the adolescents in the peer-counseling program than among the controls.[26]

Since smoking rates have been rising more rapidly among early adolescents than in any other segment of the American population, this age group will be a critical focus in the next five years for research aimed at prevention of diseases related to smoking.

MENTAL ILLNESS AND BIOBEHAVIORAL SCIENCES

During the past quarter century, drug therapy for major psychiatric disorders has drastically decreased the numbers of patients in public mental hospitals. Research on drug therapy has led to more specific methods for diagnosis, assessment of severity, and criteria for improvement of mental illness. The new psychopharmacology also drew psychiatric research into modern biological science, stimulating investigation of possible biological mechanisms underlying mental disorders.[27] For the first time, a substantial cadre of scientifically well-trained people emerged to do research on mental illness.

MAJOR MENTAL DISORDERS

Schizophrenia is a profound disorder of behavior, emotional responses, thinking, and perception. About 200,000 schizophrenic patients are now hospitalized in the United States, occupying one-fourth of all hospital beds; and another 400,000 a[illegible] in outpatient clinics or out of treatment altogether.

Schizophrenia can be disabling. Its symptoms, which usually appear in late adolescence or early adulthood, include altered movements ranging from total immobilization to frenetic and purposeless activity with peculiar mannerisms. Perceptual disorders in schizophrenia often include hallucinations such as hearing voices when no one is present. Disturbances in thinking are common and often lead to distorted concepts, bizarre speech, and grossly illogical beliefs—most vividly expressed in paranoid delusions involving profound and pervasive distrust of others. Emotional expression is sometimes completely absent, sometimes highly inappropriate to the words and actions of the individual. During relapses, most schizophrenics seem incapable of experiencing genuine satisfaction of any kind.

Manic psychosis is another severe mental disorder. Its victims usually are intensely excited and pseudoelated and speak and move very rapidly. They often seem enormously energetic, relentlessly driven, and bold to the point of grandiosity. They tend to be easily distracted, easily irritated, impatient, and rarely able to complete a task. Words and fragmentary ideas may flow so rapidly as to be incoherent.

Psychotic depression is another disabling condition that frequently leads to hospitalization. During the 1970's, about 1,500,000 people have been treated annually in the United States for depression—some in hospitals but most outside. This condition goes far beyond the ordinary sadness of everyday life, even beyond the grief of bereavement. Psychotic depression involves deep and persistent feelings of hopelessness, helplessness, and worthlessness. These feelings are accompanied by severe loss of appetite and weight, disturbed sleep, and loss of interest in familiar activities. Some depressed patients also have delusions—for example, that one has committed hideous crimes or that one's internal organs are disintegrating. Patients think of death and suicide. Indeed, severe depression is the principal context in which suicide is undertaken, although it is associated with other disorders, including schizophrenia. Suicide is the tenth leading cause of death in the United States: about 24,000 suicides are reported each year, and many go unreported.

Treatment of Major Mental Disorders

Before drug therapies, there were few opportunities for treating outpatient mental cases, and most general medical hospitals did not admit severely ill psychiatric patients. Those who were admitted were quickly transferred as a rule to a public mental hospital, usually one under state auspices. Such hospitals were largely custodial institutions, not medical facilities oriented to active treatment. Pessimism prevailed. Each year the number of patients in mental hospitals increased; very few patients were discharged The patients' living areas typically were crowded and dilapidated. Though some such hospitals were humane and compassionate, others were of very poor quality.

Psychopharmacological Therapy

Drugs called phenothiazines were introduced in the United States in 1955. They quickly proved valuable for schizophrenic patients and moderately helpful for patients suffering from manic psychosis. Inhibitors of the brain enzyme monoamine oxidase (MAO) and the tricyclic antidepressants, two different classes of drugs, were first used in the United States in 1956 and 1957. They were reasonably effective in relieving severely depressed patients. Finally, the use of lithium carbonate to treat manic psychosis was discovered in Australia in the late 1940's and tested extensively in Europe in the 1950's. The compound was studied further in this country and found to be a remarkably effective treatment for manic-depressive illness. But dramatic as the treatment is in most cases, a significant minority of manic-depressive patients do not respond well to it. Research is under way to clarify the mechanism of action of lithium to enable still more effective treatment.

In the climate of optimism that the new drugs fostered, imaginative innovations in psychosocial therapies were undertaken. These gradually led to a set of useful therapies: crisis intervention, brief psychotherapy, group therapy, family therapy, the therapeutic community, and behavior-modification therapies that apply the principles of learning theory elucidated in psychology laboratories.

Effect on Patient Population

The patient population of public mental hospitals in the United States reached its peak of 559,000 in 1955. In 1956, for the first time since records have been kept, the number declined. This population has declined steadily since then despite a further increase in the national population and a steady increase in the admission rate. By 1967, the patient population had fallen to 426,000 and by 1973 to 249,000. If the pre-1956 trend had continued, the number of patients in public mental hospitals in 1979 would be almost a million.

Most of these patients are back in the communities from which they came. Many were prepared for reentry by teaching them new interpersonal and occupational skills and by the use of transitional facilities such as halfway houses or community lodges.

Need for Aftercare

A large proportion of the patients, especially those who have not been chronically hospitalized, are able to go to work, form attachments with family and friends, and take part in community life. Nevertheless, many discharged mental patients would benefit from moderate and continuing outpatient treatment and involvement in a community network of mutual assistance. Some do get such adequate aftercare; many do not. Providing such care is especially difficult in vast, disorganized, poor urban areas, but some encouraging prototypes exist.

In some cases, patients were moved too rapidly out of the large state and county mental hospitals and communities were not able to absorb them constructively. Many chronically ill patients discharged from mental hospitals are living in cheap hotels and deteriorated single-room occupancy dwellings. In these neighborhoods, former patients are often abused by people who take advantage of their helplessness. Sometimes, they come to public attention through violent crimes, either as victim or as perpetrator.

PSYCHOPHARMACOLOGY AS A STIMULUS TO RESEARCH

In addition to benefitting individual patients, psychopharmacological agents also have powerfully stimulated clinical and basic research. Clinical research was needed to check the effectiveness of medications against mental illness. This work led to efforts to define mental disorders and their subcategories more precisely and to the development of more rigorous methods for assessing improvement or worsening of psychiatric symptoms. The work brought into this field of medicine a crucial method of pharmacological research: the random-assignment, double-blind design of experiments. This research method eliminates much of the bias of the placebo effect—spontaneous recovery from symptoms—and the bias of the staff either for or against a new medication, which might influence staff ratings.

Research on brain function has also been stimulated by drug treatments for mental disorders, principally in initial attempts to explain the mechanism of action of psychopharmacological agents. The work brought a generation of young investigators into the emerging science of neurobiology and exposed them to the awesome, fascinating complexity of the human brain.

NEUROTRANSMITTERS AND PSYCHOPHARMACOLOGY

The human brain is composed of as many as 10 billion nerve cells or neurons, whose long processes or axons conduct electrical impulses. The axons connect to other neurons in elaborate circuitry. One axon does not actually touch the next neuron, but rather is connected to it by a synaptic cleft, a microscopic gap chemically bridged by a small molecule that transmits nerve impulses across the gap. When the neuron fires, it releases this molecule, a neurotransmitter, which then diffuses across the gap and attaches to a receptor on the second neuron. There the neurotransmitter, also called a neuroregulator, activates a mechanism that may stimulate, inhibit, or modify the firing of the second neuron. Several neurons—some stimulating, some inhibiting—may be connected to still another neuron by these neurotransmitters. Neurotransmitters include four compounds called biogenic amines: dopamine, norepinephrine, serotonin, and acetylcholine. Psychoactive drugs probably act by changing the functional activity of one or more neurotransmitters.

Two hypotheses of special interest relate neurotransmitter malfunction to severe mental illness. The biogenic-amine hypothesis states that vulnerability to depression is enhanced by a functional deficit in two neurotransmitters, norephinephrine or serotonin or both, in particular parts of the brain, and that a functional excess of these same neurotransmitters at crucial brain sites increases vulnerability to mania. The dopamine hypothesis states that excess activity of the brain neurons using the neurotransmitter dopamine predisposes to schizophrenia.

The Biogenic-Amine Hypothesis

The biogenic-amine hypothesis of depression and mania originated in observations on the effects of two psychoactive drugs, reserpine and iproniazid. Reserpine was used in the 1950's to treat manic psychosis. The drug was also useful for hypertension, but some patients treated chronically with it for hypertension developed a depressive disorder. Similarly, and also in the 1950's, some patients given iproniazid for tuberculosis developed feelings of intense well-being. In 1956, this observation led clinical investigators to treat depressed patients with iproniazid, with encouraging results.

Research soon revealed that reserpine depletes serotonin and norepinephrine in the brain, while iproniazid inhibits MAO, which is partly responsible for inactivating the biogenic amines. When MAO activity is blocked by iproniazid, the level of biogenic amines in the brain increases. So reserpine causes depression and lowers the level of brain biogenic amines, while iproniazid raises the level of biogenic amines and relieves depression. Subsequently, numerous other drugs were developed that also inhibited MAO and, like iproniazid, tended to decrease depressive symptoms.

The findings with reserpine and iproniazid suggested that chemical changes in the brain could alter emotional distress in predictable ways. Two parallel and important research strategies followed. One was to look for new agents that might alter the mood of severely depressed patients; the other was to study the chemical mechanisms by which such substances might alter mood.

The tricyclic antidepressants proved to be even more effective and considerably safer than the inhibitors of MAO in treating depression.[27] The effects of both lithium and the tricyclic antidepressants on brain chemistry are consistent with the biogenic-amine hypothesis. Lithium slows the release of biogenic amines from neurons and speeds the removal of biogenic amines from the synaptic cleft, thus decreasing their functional activity. Tricyclic antidepressants, on the other hand, slow the removal of biogenic amines, presumably increasing their functional activity. The biogenic-amine hypothesis is, therefore, supported by the action of four pharmacological agents: reserpine, iproniazid, lithium, and the tricyclics.

But the tests of the biogenic-amine hypothesis of depression and mania have not been truly decisive. Drugs with effects similar to those of the tricyclic antidepressants on biogenic amines are not effective antidepressants. Moreover, the biogenic-amine hypothesis has not been confirmed by the precursor-loading strategy. The precursors are dietary substances converted to biogenic amines in the brain and the strategy is straightforward: If depression is caused by decreased functional activity of biogenic amines, such activity might be restored by feeding patients large amounts of precursors of biogenic amines. Several of these precursors have produced disappointing results in depressed patients. They have some effect on mood, but are not effective against depression. Also, direct attempts to find altered levels of biogenic amines in patients with mania and depression have produced suggestive but inconclusive results. Thus, the biogenic-amine hypothesis of depression and mania is stimulating and useful, but may well be an oversimplification. Nevertheless, the research has yielded significant new insights into brain function and has improved the quality of treatment.

The Dopamine Hypothesis

The dopamine hypothesis—that excess functional activity of the neurotransmitter dopamine predisposes to schizophrenic symptoms—also originated in both biochemical and clinical studies of psychoactive compounds. Drugs effective against schizophrenia seem to work by blocking the receptor for dopamine in brain neurons since their antipsychotic potency correlates fairly closely with their potency in blocking the dopamine receptor. If dopamine blocking and hence a decrease in dopamine neurotransmission helps to reduce schizophrenic symptoms, increased dopamine activity might cause or worsen schizophrenia. There is some evidence that it does. Some chronic users of the psychoactive drug amphetamine, which increases the activity of circuits in the brain that respond to dopamine, develop a syndrome remarkably like paranoid schizophrenia. Many of these paranoid amphetamine users have no history of mental illness.

Still, the dopamine hypothesis remains unproved. Direct attempts to find increased dopamine activity in schizophrenic patients have been inconclusive. Nonetheless, like the biogenic-amine hypothesis of depression and mania, the dopamine hypothesis of schizophrenia is an important stimulus to clinical and neurochemical research. It has led to the discovery of a number of pharmacological agents now in use that have improved the treatment of schizophrenia.

DRUG SIDE EFFECTS

Drugs that relieve severe mental illness can have important side effects. The antipsychotic medications—those used to treat schizophrenia, for example—are remarkably safe, considering their long-term use. However, a few patients who take them for many years develop involuntary twitches of muscles of the jaws, cheeks, and tongue. The disorder sometimes subsides gradually when the drug is stopped, but it is disturbing and sometimes continues long after the drug is stopped. Recently, it has been discovered that this unfortunate side effect can be alleviated by administering choline, one of the B vitamins.

The antidepressant drugs so far have not shown adverse long-term side effects, but they do have an unfortunate property. Unlike the phenothiazines, the tricyclic antidepressants can be taken in fatal overdoses by patients attempting suicide. The problem is particularly poignant, because, while the tricyclic antidepressants are quite useful for many depressed patients who may kill themselves, the drugs usually require about two weeks to work. The search for safer, more rapidly acting and effective antidepressants, with fewer side effects, will be a dynamic area in the next five years, one relying on expanding knowledge of neurobiology.[28–30] As in the past, the psychopharmacological efforts will be paralleled by work on the psychotherapy of depression.

Abuse of Psychopharmacological Drugs and Alcohol

Psychopharmacological agents have also been abused by some for their effects on mood, thinking, and behavior. On balance, however, these drugs are much less serious in this respect than is alcohol. Recent epidemiological, toxicological, and clinical research has shown that alcohol-related disorders constitute a major portion of the burden of illness. Only now—and quite belatedly—is alcohol abuse becoming a truly major focus for research in the biomedical and behavioral sciences.[31] This trend is likely to accelerate in the next five years.

ANIMAL MODELS

However new drugs are developed, they must be tested in animals before they are given to humans. Tests in animals, besides revealing toxic effects, help investigators to predict the pharmacological actions of drugs in humans. Thus, animal pharmacologists have developed a number of tests in rodents that predict antidepressant and antipsychotic activity of new medications in man. These tests were developed empirically by examining the actions of drugs known to be effective in humans on numerous animals.

A promising but expensive recent approach to animal testing is the development of animal models that more closely approximate human mental illness. Models of behavioral disorders using nonhuman primates are being created and evaluated. Several of them are useful not only in tests of individual animals, but also in studies of nonhuman primates in social groups. For example, separating a mother monkey from her infant produces a monkey model of human depression. Chronic administration of amphetamine to nonhuman primates produces a disorder resembling a simplified form of paranoid behavior.

The principle of a relatively close biological relationship of nonhuman primates to humans that is helpful in this work can also be applied in other contexts. For example, further research on immunization against a form of viral hepatitis and on reproductive biology pertinent to contraception relies heavily on the use of higher primates, especially chimpanzees. However, the availability of primates has been sharply curtailed by countries that traditionally have supplied them, and the United States does not have adequate breeding colonies. This situation could seriously inhibit progress on problems of biobehavior, reproduction, and infectious disease.[32]

AGING AND HEALTH

The number of Americans aged 65 and over has risen from 4 percent of the population in 1900 to 11 percent today. Their personal health-care expenditures in fiscal 1976, when they comprised about 10 percent of the population, were nearly 30 percent of the total for all Americans.

The National Center for Health Statistics (NCHS) has recently projected health trends for the next 25 years. Especially significant for health services is the NCHS projection for the postretirement portion of the population. The population aged 65 years and older grew 76 percent from 1953–78, compared with 38 percent for the entire population. If mortality rates decline as expected, the population aged 65 and older will continue to grow about twice as fast as the entire population during the next quarter century and will be the fastest growing segment of our population.

Chronic disorders tend to increase with age. Their effect on limitation of activity is a measure of health status reported regularly in NCHS Health Interview Survey. The current rates of activity-limitation, applied to the populations projected for 25 years hence, indicate that the number of persons with activity-limitation will rise from 31 million in 1978 to 42–46 million in 2003, an increase of 35–48 percent among all ages. We can reasonably expect a corresponding increase in demand for health services.

The main causes of limited activity today in persons over age 65 are cardiovascular diseases, arthritis, and mental and neurological disorders.[33] Impaired vision and hearing are also especially important among the aged. In recent years, the number of elderly people in mental hospitals has decreased substantially. However, nursing homes—varying greatly in nature and quality of care—are becoming increasingly important in the care of the elderly. A very large increase in the number of nursing-home residents is likely in the next 25 years. The special problems of health care for the elderly, such as transportation to the source of care, will also have to be dealt with.

RESEARCH ON AGING

The maximum life span for humans is about 100 years, but few people live that long because of the interaction of aging and disease. The distinction between aging and the diseases associated with aging is important. Atherosclerosis, cancer, and senile dementia, for example, are not inevitable consequences of aging and can be considered separately from normal aging and each of these conditions investigated in its own right.

Cellular Biology

Research in cellular biology is clarifying the basic processes of aging. Cultivation of human cells *in vitro* provides a model for studying cellular changes in aging, since such cells have a finite life. Using cell fusion, investigators are beginning to determine the extent to which longevity is programmed in the nucleus or in the cytoplasm. This work is an attempt to find the basis for the difference in longevity between cell types. Cells from patients with conditions associated with premature aging may provide important clues to the genetic basis of aging.

Immunological Studies

The immune system, which produces antibodies in response to invading antigens, deteriorates with age; the decline can lead to increased susceptibility to infection and possibly to some tumors. In one study, 80 percent of individuals over age 80 who had no or minimal immune response to specific antigens were dead within two years, as contrasted with 35 percent of those who had a vigorous immune response.

Autoimmunity, or immune responses to one's own tissues, generally increases with age in experimental animals and in humans. Animal studies have shown that increases in autoantibodies and abnormal immunoglobulins, or antibody proteins, can be slowed by manipulating diet and other environmental factors. The potential reversibility of these immunologic changes suggests that research may be able eventually to ameliorate the effects of aging or at least the susceptibility with age to specific diseases.

Neurological Studies

Senile dementia is pathologically distinct from normal aging. This condition, severe in 5 percent and moderate in 10 percent of those 65 and over, is characterized by loss of initiative, decrease in judgment, difficulty in selecting appropriate words, severe loss of recent memory, difficulty in performing calculations, disorientation, and personality deterioration.

Recent research on senile dementia has focused on changes in choline acetyltransferase, an enzyme responsible for production of the neurotransmitter acetylcholine in certain nerve cells. Although the level of the enzyme declines in normal aging, it declines much more in patients with senile dementia (of the important Alzheimer type) than in age-matched controls. Acetylcholine is one of the major neurotransmitters in a part of the brain having a vital role in memory. Receptors for acetylcholine are present in normal quantity even when the enzyme responsible for producing the neurotransmitter is not. It may thus be possible to find a substance that activates the receptors and so alleviates Alzheimer senile dementia. Very recent evidence indicates that feeding extra choline to patients with senile dementia can improve memory.

Other neurologic research has demonstrated age-related changes in functional activity of neurotransmitters. The finding may explain the higher incidence of Parkinsonism in older patients. The ailment is associated with inadequate functioning of brain cells that normally function in response to the neurotransmitter dopamine. The life of rats has been prolonged by feeding them the amino acid L-dopa, the precursor of dopamine. The compound presumably activates pathways in the brain that respond to dopamine and has proved useful in relieving the symptoms of Parkinsonism in humans.

Other studies have suggested that pituitary glands in older people may secrete a compound that diminishes the response of peripheral tissues, such as toes and fingers, to thyroid hormone. Individual differences in secretion of this compound could affect activity and vulnerability to cold. Also, some investigators are seeking a pacemaker for the aging process in the brain's pituitary–endocrine system.

Various hormones and hormone fragments are being studied with respect to brain function, including learning and memory. A new approach to the treatment of senile brain disease involves the use of a hormone fragment that seems to improve memory in certain behavioral tests in animals. This medication, ACTH (4-10), is a fragment of a stress-related human pituitary hormone. It is being studied in patients with defective memory.

Pharmacology and the Aged

Pharmacological treatments currently show some promise for certain ailments of the aged. Drugs that dilate blood vessels may be useful in treating senile patients whose symptoms are due to decreased blood flow to the brain. Such medication is already in clinical use. However, further clinical trials are needed to determine how effective the drugs are and how to identify patients most likely to benefit from them. The same is true of anticoagulant therapy, such as aspirin on a long-term basis, for prevention of stroke.

The clinical pharmacology of elderly patients presents distinctive problems. We are only beginning to understand the changes in enzyme and organ function, the effects of drug interactions, and the difficulties in adhering to a therapeutic regimen because of poor memory and altered eating habits. For practical reasons, most studies are performed on young people, and their applicability to the older population is limited. Research in clinical pharmacology will benefit from a perspective that considers the entire life span.

DEPRESSION IN THE ELDERLY

Depression is common among the elderly, who account for almost one-quarter of all reported suicides. Some patients with the personality changes, thinking, and memory defects of senility seem to improve markedly if treated with the same medications given younger patients suffering from depression. There is no reliable way so far to predict which elderly patients are most likely to benefit from antidepressants. In this respect, as in others, the aged have been neglected both in current services and research.

Of the mental illnesses of the aged, depression is more common than is serious loss of awareness and judgment resulting from pathological changes in the brain. The aged are subject to multiple stresses, including loss of family members, friends, jobs, economic assets, all of which contribute to their depressions and other psychiatric disorders. Social isolation is significant in these difficulties.

SOCIAL EFFECTS

Clinical research has shown that senile, incompetent behaviors of elderly people are often related to social factors. Many persons can function effectively at advanced ages when placed in a network of mutual aid and social support and given meaningful tasks and a basis for self-respect.

Retirement is one of the critical changes of later life. We know little about biological and physiological changes in response to retirement. Preparation for retirement can be studied in field experiments, by systematically comparing different approaches in similar populations and examining their outcomes.

The family remains the major social, economic, and emotional resource for dependent older persons. However, many problems related to dependency require study, especially because they are so pertinent to the individual's ability to function in society. Special attention should be directed to the socially isolated, who make up most of the 20 percent of those people over 65 who will spend some time in a long-term care facility and who constitute a vulnerable subset of the elderly population.

CHANGES IN SLEEP/WAKE MECHANISMS

While the changes in sleep/wake mechanisms with aging are poorly understood, they are associated with many complaints of insomnia among older people.

There is a resultant medical problem because the elderly are particularly vulnerable to the hazards of sleeping medications, since they are more likely to already have the problems associated with those same hazards, such as respiratory disorders and impaired kidney function. Also, they are more likely to suffer from disorders that require the use of other prescription drugs and so are exposed to the risk of toxic interactions among several different medications.

There is little data on the efficacy of sleeping medication for older people. The few available studies show no drug more effective than another. Given the paucity of data on a clinical problem that is truly difficult for so many elderly people, a serious effort to learn more of the physiology and pharmacology of sleep and its disorders in the elderly is indicated in the years ahead.

HEALTH SERVICES FOR OLDER PEOPLE

There is a growing consensus in clinical medicine and public health on two principles regarding health, disease, and adaptation in older people. First, the care of the elderly should be designed to maintain maximum possible functional and social independence. And, second, health care and social services should be provided in a manner that preserves the dignity of the elderly individual and provides opportunities for personal choice.

Although aging is likely to bring some decline, many elderly people function well until shortly before death. For those who are functionally disabled, the rate of decline may be decreased substantially or minimized through early detection and appropriate assistance. To promote independence among the elderly, we must improve our ability to identify environmental, genetic, and social risk factors in the development of functional dependency.

Current practice emphasizes skilled institutional care for the elderly. Many individuals require and benefit from intensive institutional care, but the lack of alternative types of care can result in inappropriate use of high-intensity services. Indeed, studies of older people have demonstrated the enhancement of dependency by unnecessary services. If individuals are blocked from taking responsibility for certain daily tasks, such as preparing meals and personal care, their abilities tend to diminish rapidly.

Public concern for the functionally dependent elderly is warranted for economic as well as humanitarian reasons. It is in the interest of both society and the elderly individual to forestall dependency or to minimize its impact once functional capacity begins to decline. It is important to consider those who are at high risk of becoming dependent, such as the recently widowed, in addition to individuals who already depend on others for care.

We need better ways to compare systematically the different arrangements for minimizing functional dependency in the elderly. We must also seek more effective means of linking medical and social services. And we must examine ways of improving both long-term institutional care (as Sweden has done) and home-care services (as Great Britain has done).

A sustained effort will be required to meet the health needs of the older people of this country. Their rising numbers and special problems offer real challenges in basic science, in clinical investigation, and in health-services research.

GENETIC FACTORS IN DISEASE

An emerging line of inquiry that will become more visible in the next five years is the influence of genetic factors on the response to environmental agents. The combination of biomedical research with family and population studies is likely to bring new understanding to the prevention and treatment of diseases that are so burdensome in industrialized countries—for example, cancer, arteriosclerosis, depression. This work will require identification of genes involved in susceptibility and resistance to these diseases. Greater understanding of the interaction of such genes with specific environmental factors may point the way to sharply focused preventive techniques.[33–37]

A special target of such research is likely to be substances to which a relatively large fraction of the population is genetically susceptible and to which it is widely exposed—substances found in diet, medication, and occupation.

INNOVATION IN HEALTH CARE DELIVERY

The importance that American society attaches to health has provoked in recent decades a continuing debate about how health care should be provided. Likely to intensify in the coming five years, this debate has been accompanied by new arrangements for providing health care. We seem to be moving gradually away from the traditional arrangement, whose dominant characteristics include the delivery of medical services by physicians working independently, greater emphasis on curing than on preventing illness, and the payment of fees after the delivery of services. This mode remains strong, but others are emerging.[38]

Innovation in providing medical services is hardly surprising, given the increase in the nation's population, the intensive development of medical technology, the rise of the medical specialist, the growth of complexity and cost in health care, and the social and technical inventiveness that characterizes American society.

The earliest questions about the effectiveness of the traditional method of delivering health care in the United States were essentially technical ones of a kind familiar to those in many other occupations. For example, exponential growth of knowledge in this century, combined with unprecedented technological advances, have made specialization imperative, particularly in science and the science-based professions. In medicine, specialization has meant a declining role for the general practitioner, the individual physician able to deal with every health problem from measles and heart disease to broken bones and depression.

GROUP PRACTICE

As diagnosis and treatment became the joint effort of two or more physicians, a part of the medical profession concluded that technical efficiency in curing disease could be improved if physicians became partners or members of larger groups. By the early 1960's, physicians were tending increasingly to enter private group practice, with continued emphasis on curative medicine and no change in the standard fee-for-service arrangement. This trend toward multispecialty group practice is stronger and more varied in this country than elsewhere.

Service innovations must be as much concerned with the quality of care as with its cost.[39,40] The various kinds of group arrangements bring to clinical practice the kind of peer review that is effective in judging scientific quality. Peer review is quite formal in the clinical context, but it also includes informal peer contact, particularly among people of different clinical disciplines. Such interactions tend to maximize the diffusion of information, to keep practioners abreast of new developments, and to ensure that procedures are executed in accordance with well-recognized standards based on the best-available evidence.

The basic asset of group practice in organized settings is the pooling of resources, including ideas, techniques, instruments, facilities, and clinical judgments. Within this framework, it is quite possible to have an enduring doctor–patient relationship with a primary-care physician, specialists being brought in when needed. Group ventures can be organized in many ways, and we are now seeing traditional American ingenuity and pluralism being applied to the challenges of the new science base and changing social context of health care.

ACCESS TO HEALTH CARE

In the United States in the middle and late 1960's, public attention shifted to another problem: how to assure that all Americans—regardless of race, age, income, or where they live—have access to at least a decent minimum of health care.[41] This goal has been achieved in almost all advanced countries. In the United States, it became apparent that millions of citizens could not obtain such care—or were severely restricted in their efforts to do so—because of inadequate income, racial discrimination, or residence in areas where doctors were reluctant to live, particularly inner-city neighborhoods and rural areas. The national response to these problems included the creation of the multibillion-dollar Medicare and Medicaid programs, efforts to increase the numbers of minority students in medical schools, publicly funded clinics in poor areas of the large cities, and continuing nationwide effort by certain groups (particularly those representing labor unions, poor people, and minorities) to establish through federal legislation a national health insurance program that would pay most health care bills of all Americans.

More recently, concern over costs, combined with growing appreciation of the value of preventing illness, has led to increasing interest in alternatives to both the traditional private method for delivering health care and to a federally controlled and financed system. Several alternative delivery systems have become reasonably well established in the 1970's. Indeed, there is now a ferment of innovation in the organization and financing of health care delivery systems. These pluralistic developments are likely to ramify in the next five years, and they deserve careful evaluation.

ALTERNATIVE DELIVERY SYSTEMS

Several kinds of alternative delivery systems now exist; the most prevalent is the Health Maintenance Organization, or HMO. As of August 1978, 199 HMO's were providing health care to more than 7 million Americans, or over 1 million more than a year earlier.

Prepaid Group Practice

The predominant type of HMO is the Prepaid Group Practice (PGP) plan. More than 90 percent of those enrolled in HMO's in August 1978 were members of either a PGP plan or of its principal variation. PGP plans have two essential characteristics. The first is that families or individuals enrolled in them agree to pay a set monthly premium to the HMO, whether or not they need medical care. That fee covers all care, including hospitalization. In many cases the monthly premium is paid by an employer or by government; the federal government, for example, pays monthly premiums for Medicare recipients who belong to HMO's. The second characteristic of a PGP plan is that those enrolled are not entirely free to choose their own physicians or other health personnel. In most cases, the HMO's staff physicians provide primary care and also specialized care to the extent that they can. Where additional specialized care is needed, some HMO's require, and others recommend, that the patients consult specific specialists.

Independent Practice Associations

The other principal type of HMO is the Independent Practice Association (IPA); by August 1978, IPA's were providing service to 9.5 percent of the people enrolled in HMO's. In essence, the goal of IPA's is to give their members the advantage of fixed monthly payments and at the same time let them choose their own doctors. Participating physicians practice as they always have, either alone or as members of a private partnership or other group. The IPA customarily reimburses them out of the monthly premiums on a fee-for-service basis. The appeal of the IPA depends in part on how many physicians in a community agree to participate. Giving patients the opportunity to choose their own physicians has little meaning unless most of the community's physicians participate in the association.

Assessments of HMO's

HMO's must be managed carefully and must deliver services efficiently. Since HMO's agree to provide their stated benefits for a fixed payment in advance, they must do so or risk a loss. It is this fixed payment in advance which is largely responsible for a HMO's ability to hold costs down, since it forces the staff to provide care within an annual budgetary limit. In contrast, the traditional method requires payment for costs already incurred, regardless of amount.

Hospitalization costs are the highest of all types of medical care. They were also most prone to inflation between 1972 and 1977, when hospital service charges rose 60.1 percent, or an average of 12 percent a year. HMO's have been able to reduce the cost of health care chiefly by reducing the time that their members spend in hospitals. A comparison of hospital use in seven large states between HMO members and persons with Blue Cross/Blue Shield coverage, for example, showed that the HMO members spent 782,000 fewer days in the hospital in 1976 and had less surgery.

Such examples are encouraging but not compelling. It is possible that the savings are achieved by reducing medical services to the patient below the level necessary for good health. It is also known that some patients go outside the system for surgery; if such patients were to become a large proportion of HMO members, the cost savings would be more apparent than real.

One substantial systematic study compared the quality of care under an HMO, the Health Insurance Plan of Greater New York, with other forms of delivery. The results indicated that HMO members experienced fewer premature births and fewer newborn deaths than did a New York City sample under the care of private physicians. The observations on premature births are of special interest because these babies are at high risk of mental retardation and long-term disability. Other studies of maternal and child health care in organized settings, not necessarily HMO's, indicate that such approaches can be remarkably effective in reducing death and illness among mothers and children. However, such experiments are complicated by uncertainties about the populations studied. For example, members of HMO's may be habitually more careful of their health than nonmembers. We need a much better understanding of the true effectiveness of HMO's and similar organizations. In fact, few comparisons have been made of the health outcomes of different service-delivery systems. Such comparisons require greater attention in the next 5–10 years.

Scholars and clinicians concerned with innovation in health care are increasingly interested in creating efficient and high-quality systems that can compete effectively in the health care marketplace on a nationwide basis.[42] A variety of useful models are already functioning in several regions. Whether they spread widely will depend on the professional and social context. The following characteristics are desired in such systems:

- A group or association of physicians accepts responsibility for providing comprehensive health services to a defined population.
- People join through free choice among this and other health plans, thus providing a competitive incentive for the health plan to see to it that its personnel give patients judicious, caring, willing service.
- The system provides health maintenance services to the extent that they are effective; this includes such preventive measures as immunizations, prenatal and perinatal care and counseling on health-protective behavior.
- The health plan has built-in cost constraints. The physicians accept responsibility for the total per capita cost of the care of their patients, and they use their best judgment of how to give appropriate care. They systematically consider such trade-offs as substituting ambulatory care or home care for more costly inpatient care—for example, to make services more accessible to patients who are poor or who live far away.
- The mix of specialists, facilities, and other resources are matched to the needs of the enrolled population.
- There is built-in quality control through peer review, ease of consultation, and follow-up on patient satisfaction and health outcome. Specialized procedures are performed by specialists whose annual volume of such cases is sufficient to maintain their proficiency.
- There is continuity of health plan membership.
- The health plan keeps a unit medical record for each patient, so that a new doctor can quickly and reliably ascertain what has been and is being done for the patient and so that unnecessary duplicate tests and conflicting prescriptions can be avoided.
- Primary care physicians and specialists work closely together in the same system.
- General and mental health professionals interact freely within the system, so that mentally ill patients are not excluded or stigmatized and so that the behavioral aspects of general health, such as cigarette smoking, can be handled effectively.

How and to what extent these characteristics can be achieved in actual practice remains to be seen.

Other Alternatives

Efforts are being made to extend primary, or "front-line," health care to rural areas and low-income urban areas. One such effort is the National Health Service Corps—young physicians whose medical education is subsidized in return for postgraduate service in underserved areas. Also emerging is a cadre of specially prepared nurses—nurse practitioners—who provide primary care in underserved areas.

Innovations in health insurance, both public and private, include a trend toward providing a reasonable "floor" of health care for those who can afford very little or no health insurance at present, and also increasing coverage among the general population for preventive and ambulatory services.

Efforts also are being made to build close functional links between mental-health and general-health services. Also under way are analyses of organizational changes aimed at more effective health services for children and for the elderly.

HEALTH-SERVICES RESEARCH

The evolution and assessment of useful innovations in health care will depend on continuing advances in biomedical research, including clinical investigation and to an increasing extent on the young field of health-services research.

The need for accurate, dependable information about health services in the United States is becoming increasingly apparent. Management of the personnel, facilities, and technologies that comprise modern health care institutions requires information similar to that needed to manage other complex enterprises. The growing involvement of the public sector in the financing and provision of health services, combined with the need to address increasingly complex issues of resource allocation, requires more knowledge than ever before. The systematic comparison of alternative delivery systems will be a great challenge to the health sciences in the years ahead.

PERSPECTIVES ON HEALTH

At the end of World War II, medical care in the United States was provided largely by physicians working as independent general practitioners and by a modest network of community hospitals with limited diagnostic and therapeutic capabilities. Penicillin—the first wonder drug—had recently become available; the chief diagnostic tools of the practicing physician were the stethoscope and the X-ray. Few could have foreseen broad-spectrum antibiotics, polio immunization, open-heart surgery, organ transplants, and computed tomographic scanning.[43]

The American system of health care today bears only a family resemblance to that cottage industry of 35 years ago. Health professionals use a wide array of drugs and diagnostic and treatment equipment requiring a high level of technical competence. The nation has hundreds of thousands of physicians and other health professionals in many specialties and subspecialties—surgery, internal medicine, gynecology, urology, psychiatry, pediatrics, dentistry, nursing, and many more—as well as thousands of laboratory and equipment technicians, physical therapists, physicians' assistants, and other support personnel. Furthermore, hospital operation is now so complex that a cadre of specifically trained hospital administrators is required. The American system of health care, in short, has become a large and complex enterprise, accounting in 1978 for almost 9 percent of the Gross National Product.

Recent years have brought profound insight into the causal processes of many diseases, including nephritis, endocrine disorders, arthritis, gout, ulcer, and hepatitis; and of such genetic disorders as cystic fibrosis, muscular dystrophy, and lipidoses. Numerous surgical, pharmaceutical, and palliative measures prolong life in some cases and make it tolerable in many others. There have in fact been clear advances on many fronts in the vast field of health.

NEW PROBLEMS IN HEALTH CARE

The recent growth of American medicine has been accompanied by significant new problems: doubts about the wisdom of some of the newer medical techniques, uncertainties about the proper relationship between patient and physician, and continuing national concern about those—such as the poor and the elderly—still unable to obtain the full benefits of current medical knowledge. There is also reason for concern that the needs of children and adolescents, regardless of family income or social status, are not being met adequately. Within the past five years, these problems have been augmented by serious concern with the rising costs of medical care—recently increasing at about twice the general inflation rate. Furthermore, there is a realization that a continued rise in spending on personal health services may not yield commensurate benefits; greater attention to hazards to health in the environment and in individual behavior might be a better investment.

Means of containing health care costs that serve both the public health and our personal and national financial imperatives will not be found easily. In the long run, we may reduce costs by acquiring deeper understanding of human biology and behavior and applying this understanding to prevention of disease.

Research and Health

In principle, all basic research in the life sciences has implications for the prevention of disease.[44] The more we know of the human organism, the more we should be able to affect basic life processes—such as the reproduction of cells or the secretions of glands—so as to prevent disease. Research in the life sciences already has contributed much, but scientists still do not understand most of the bodily mechanisms to a degree that would allow prevention of disease at its primary site. "The Living State" chapter illustrates progress being made in this direction.

IMPROVEMENT OF HEALTH

The health of the public may be improved by preventing disease in various ways: personal health services, environmental measures, and changes in individual behavior.

Personal health services include the procedures and counseling by physicians and their collaborators in private offices, clinics, hospitals, and other organized settings—what is usually called medical care. The most striking example of prevention in medical care is immunization against communicable disease. The effectiveness of immunization against such diseases as poliomyelitis, diphtheria, and measles has been demonstrated beyond all reasonable doubt. A vaccine against pneumococcal pneumonia is now coming into use, and work is proceeding on vaccines against gonorrhea and malaria, although both are some years away from use. Yet millions of Americans remain unimmunized against diseases for which effective vaccines are available.[45] Increasing the immunization rates for polio and measles to the levels of Western Europe is a challenge that can be met in the next five years. Efforts in this direction in the U.S. Public Health Service are gaining momentum.

Environmental measures for improving health are well illustrated by the effectiveness of fluoridated drinking water in reducing dental caries. For example, 12–15 years of communal water fluoridation in five cities reduced the incidence of caries in children 41–70 percent. The benefits for adults are more difficult to assess. However, a comparative study of two communities—one fluoridated, the other not—indicated that people over 45 had at least 25 percent fewer caries with fluoridation.

In recent years, individual behavior has been viewed increasingly as a leading key to good health. The health sciences now are extending their scope to individual behavior patterns and their relationships to the illnesses most prevalent in American society. For example, deaths from lung cancer, coronary heart disease, chronic respiratory disease, and other conditions attributable to cigarette smoking constitute a major part of the burden of illness in this country. Yet surveys indicate that more than one-third of American adults still smoke cigarettes and that, until very recently, an increasing proportion of adolescents aged 10–15, especially girls, were taking up the habit. Smoking in the 10–15 age-group is at a very high level. Should these patterns of behavior persist, the consequences for the health of the American people could be serious indeed.

Building public understanding to the point where Americans take these behavioral risks to health seriously will be a long-term and difficult process. We badly need to find more effective methods of health education in the next 5–10 years. The smoking problem in particular is likely to receive much more public and scientific attention than in the past.

Because health depends on medical care, on environmental conditions, and on individual behavior, we must take a variety of approaches to maintaining and improving the health of the American people. The circumstances call for a broad spectrum of sciences working together over the long term.[46,47]

GOALS FOR THE HEALTH SCIENCES

The burden of illness in this country shows clearly that science still has much to do in reaching a fundamental understanding of how the body functions and why it malfunctions. Achievement of deeper understanding of the body's vital processes will require sustained effort in every part of the health-sciences spectrum, from basic science through clinical investigation to health-services research. One can observe today a revival of interest in older disciplines, such as epidemiology and biostatistics, the strengthening of such relatively new disciplines as biomedical engineering and behavioral sciences, and novel combinations of disciplines for doing research on prevention of disease and delivery of medical services.

As new scientific opportunities arise, their relevance to disease must be determined. Not so long ago the then-emerging discipline of biochemistry was viewed with suspicion by chemists as weak chemistry and by biologists as weak biology. Today, this hybrid discipline plays a central role in biomedical research. Similarly, not so long ago most medical scientists doubted that genetics would have any practical significance for health in the twentieth century. Today, genetics is one of the most dynamic fields in medicine.

At one end of the health-sciences spectrum is basic research, usually initiated by a laboratory investigator with no particular prevention or treatment goal in mind. Next in the spectrum are small-scale clinical investigations designed to determine the significance of new knowledge to humans. Such clinical research often stimulates basic science. Small-scale investigations sometimes are followed by medium-scale experiments, usually involving several hundred people, and then by large-scale, controlled field trials involving thousands of persons. Such large-scale experiments typically seek to delineate the effects of particular risk factors and the results of specific kinds of interventions on human health.

At the other end of the research spectrum is health-services research, which is concerned with organization of medical services and their quality, availability, and cost. Careful and systematic investigation of these matters is a recent development. For that reason, such research still lacks a secure institutional base, either in government or among the nation's leading universities.

The prevention of rheumatic heart disease illustrates the value of linkages across the health-research spectrum. Basic research on infectious diseases led to clinical investigation of the link between streptococcal infection and rheumatic fever. This resulted in clinical work that showed the efficacy of penicillin for preventing rheumatic fever and thus rheumatic heart disease. These interrelated investigations took years to complete, and work is still progressing on some aspects of the problem—for example, health-services research on effective implementation of methods of preventing rheumatic fever in ordinary and, especially, poor communities. But the example is clear: The sequential linkage of basic research, clinical investigation, and health-services research drastically reduced the toll of a serious disease in a few decades.

Assessing New Techniques

An important aspect of health-services research is assessment of the benefits, costs, and risks of new interventions and treatments. These investigations are difficult and sometimes expensive. In some cases, we are still hard pressed to measure benefits; the difficulties are compounded when a long time passes between treatment and a measurable effect on health.

Despite the difficulty and added expense, these assessments of new technology and treatments are crucial because of escalating costs, intensified public scrutiny, and the potential of great harm that accompanies the potential of great benefit. More and more in recent years we have seen new kinds of medical techniques widely used and accepted without a thorough, objective assessment of potential clinical utility or probable impact on the existing health care system. Computed tomographic scanning is one recent example.[6] In the next five years, serious efforts will be made to strengthen clinical systems for determining the proper applications of promising technology.

BROAD APPROACH TO RESEARCH

Each part of the spectrum of research in the health sciences must be fostered if the task of improving health is to be pursued effectively. Overriding emphasis in any one part of the spectrum may act against vigorous cross-stimulation among the sciences and continuing improvement in the health of our population. Communication and cooperation among various parts of the spectrum must be strong, so that knowledge developed through basic research can be translated into practical applications at the clinical level and so that the clinical problems uncovered by medical practitioners can be relayed back to those conducting research. Mechanisms are being developed—for example, by the National Institutes of Health, by the Congressional Office of Technology Assessment, and by professional societies—for arriving at a technical consensus on the potentials of new information for providing specific diagnostic, therapeutic, and preventive advances.

As the national prevalence of various illnesses changes, the scientific efforts needed to reduce illness also change. During the first half of this century, the chief goal of American medicine and public health was to develop ways to control the most serious contagious diseases, which were then the principal causes of death in the United States. As that goal has been approached, with remarkable though not total success, the health sciences have gradually shifted their attention to the types of illness and injury most characteristic of American life today: cardiovascular diseases, cancer, respiratory diseases, mental illness, major vehicle accidents, abuse of alcohol and other drugs, and chronic disorders associated with aging. Most of these problems have biomedical, environmental, and behavioral components, so that multifaceted approaches are needed to make progress against them.

The nation is beginning to test the extent to which the methods of science can be brought to bear on all factors that determine the health of the American people—building on the firm base of biomedical research to include behavioral and environmental influences and the effects of health care per se. Research now extends from the laboratory bench to the patient's bedside to daily community activities. To further improve health, we must have deeper insight into molecules, tissues, organisms, populations, and health care systems.

OUTLOOK

The following outlook section on health of the American people is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

The nation is learning that health depends only partly on the traditional methods of health care. Environment and behavior are also crucial, especially in preventing disease. The scope of research in the health sciences has been expanding steadily and today extends from traditional biomedical investigations to the patient bedside to environmental and behavioral factors in everyday life to the organization of health services in the community.

In recent decades, new kinds of prevention and treatment have been applied successfully to a wide range of disorders, including infectious diseases, vitamin deficiencies, damaged organs, mental illness, cancer, and cardiovascular diseases. But a heavy burden of illness remains.

BURDEN OF ILLNESS

The burden of illness can be assessed both clinically and economically; clinically—how many people suffer and die—cardiovascular disease and cancer predominate; economically, respiratory diseases have particularly heavy impact in terms of work days lost and other indices.

Mental illness, although not a leading cause of death, takes a heavy toll in suffering, disability, and economic costs. Diabetes and arthritis also are large components of the national burden of illness. Recent research has shown clearly that alcohol-related disorders comprise an important part of the burden of illness. Only now—and quite belatedly—is alcohol abuse becoming a truly major focus for research in the biomedical and behavioral sciences. This trend in research is likely to accelerate in the next five years.

The next decade is likely to see considerable improvement in the reliability of clinical and economic measures of burden of illness. These measures can enable a more rational allocation of resources for research, education, and services. For instance, the population 65 years and over is increasing steadily. Since chronic disorders tend to increase with age, it is reasonable to expect a corresponding increase in demand for health services. We can also anticipate an increase in the special problems in delivery of health care to the elderly, such as tranportation to the source of care.

HEALTH AND BASIC SCIENCE

As the prevalence of various kinds of illness changes, the scientific efforts needed to reduce illness will also change. One can observe, for example, a revival of interest in older disciplines, such as epidemiology and biostatistics, the strengthening of such relatively new disciplines as biomedical engineering and the behavioral sciences, and an upsurge of interdisciplinary research on prevention of disease and delivery of health services. Advances in these fields—and in such fields as biochemistry, genetics, neurobiology, immunology, and pharmacology—potentially could yield marked gains in health over the next few decades.

Major strides in basic science can be expected in the next five years. Still, we know too little to be able to relieve much of the burden of illness in the near future. The interplay of basic research with clinical investigation can provide deeper insight into molecules, tissues, organisms, populations, and health care systems.

CARDIOVASCULAR DISEASE

Large-scale experiments on the prevention of cardiovascular diseases have highlighted the need to strengthen the linkages of basic research, clinical investigation, and health-services research. Such linkages can greatly reduce the toll of these serious diseases. Efforts will be focused on prevention and treatment of arteriosclerosis and hypertension.

More must be known about the basic processes of hypertension and arteriosclerosis to enhance our ability to prevent the diseases and their complications by both pharmacological and behavioral means. Also, we must learn to identify cardiovascular disease patients more precisely to better fit appropriate therapy.

Improved drug therapy for hypertension should become available in the next five years. It will take longer, however, to determine which therapeutic regimens will work best for which patients in preventing such complications of the disease as stroke or heart failure.

Hypertension is more frequent among blacks than among whites in this country. We do not know why. Identification of specific genetic and environmental factors in this public health problem remains a major scientific challenge.

Basic research on arteriosclerosis is likely to advance rapidly in the next five years. Progress should be made in clarifying the nature and formation of the basic lesion causing the disease, in developing technical advances that might lead to noninvasive diagnostic techniques, and in providing a firmer foundation for preventing the disorder.

CANCER

Environmental factors acting upon genetic predispositions are now considered important in the orig of cancer. Where such causal factors can be identified, they can often be minimized in the environment so as to prevent cancer. A goal for epidemiology is improved understanding of the relationship of changing patterns of environmental agents to the incidence of cancer. Special attention should be given to those cancers that occur most frequently and with large geographic variations in their incidence.

Even after a carcinogen has been identified, it may not be possible to remove it from the environment or to avoid it totally. Thus, we will need to establish socially acceptable exposure levels. Such levels must be based on dependable, quantitative risk estimates derived from data much better than most of that now available.

Cigarette smoking is responsible for 80 percent of the incidence of lung cancer, so that the advantages and disadvantages of low-tar-and-nicotine cigarettes deserve systematic investigation. Careful attention should be paid to the effects of cigarette smoking, not only on lung cancer, but also on cardiovascular and other cigarette-related diseases.

The search for more effective combinations of surgery, radiotherapy, and chemotherapy for treatment of various cancers will proceed intensively during the next five years.

BEHAVIOR AND DISEASE

Modifying risk factors for various diseases through changes in behavior is important but not easy. Significant educational and social changes will be required. However, there are encouraging indications that firmly established patterns of behavior—such as smoking, exercise, diet, working, and coping with stress—can be changed. Work in the area will intensify during the next five years, and will involve combined efforts in biomedical and behavioral research.

To the extent that behavior harmful to health can be reduced, the unborn infant also will benefit. The unborn infant is at increased risk of death or abnormality if the mother smokes or drinks alcohol during pregnancy. The vulnerability of the fetus to maternal intake will be investigated intensively in the next several years.

Because cigarette smoking contributes so heavily to cancer and cardiovascular disease, research on stopping smoking is highly pertinent. The focus will be not so much on stopping initially as on preventing relapse. This is likely to be an important area of study in the next five years.

SMOKING AND ADOLESCENCE

Smoking rates in the 1970's have been rising more rapidly among early adolescents, but have stabilized or even decreased in other parts of the American population. Early adolescence, therefore, will be a critical focus in the next five years for research on prevention of diseases related to smoking. Adolescence is a critical period of biological and psychological change, and of drastic social change. It is a time when lifelong behavior patterns crucial to health are formed. These patterns involve not only cigarette smoking, but also use of alcohol and other drugs, automobile driving, diet and exercise, and human relationships. The impact of peers, influential adults, and the mass media on adolescents needs careful, systematic study to clarify the bases of health-related behavior. Research on prevention of diseases due to smoking will focus on the adolescent age group in the next five years.

RESEARCH ON ANIMALS

Some areas of health research, including behavioral research, rely heavily on the use of higher primates, because of their biological similarity to humans. It will be essential in the next five years to develop adequate breeding colonies and research facilities for such primates. They are needed to make critical progress on problems of behavioral biology, of reproduction (e.g., contraception), and infectious disease (e.g., viral hepatitis vaccine).

MENTAL ILLNESS

Drugs for treating mental disorders have more than halved the population of public mental hospitals in this country since about 1955. The drugs also have spurred rapid growth in research on the chemistry and biology of the brain and particularly on the biochemical bases of mental disorders. This research in part has led to more effective drugs for mental illness and is likely to continue to do so. Better drugs also can be expected to result from systematic modification of existing compounds.

AGING

The rising proportion of elderly people in the American population is leading to a new era in basic, clinical, and health-services investigations of the problems of aging and health. These efforts include work on immune responses as protective mechanisms; changes in patterns of hormonal secretion with age; neurobiology in relation to memory, sleep, and other brain functions; the distinctive problems of drug therapy in the elderly; maintenance of functional independence in coping with chronic diseases; and linkage of health and social services in maintaining health into later life. Such efforts will intensify in the next several years

GENES AND ENVIRONMENT

Studies of the influence of genetic factors on responses to environmental agents will expand in the next five years. The work will require the identification of genes involved in susceptibility and resistance to particular diseases. The interaction of such genes with specific environmental factors may point the way to sharply focused preventive techniques. These investigations are likely to concentrate on substances that are widespread—as in diet, medication, and occupation—and to which a relatively large fraction of the population is genetically susceptible.

HEALTH CARE DELIVERY

Innovations in health care delivery are likely to be prominent in the next five years. Various organized settings, such as group practice and health maintenance organizations, are being used increasingly. Efforts to extend primary health care to rural areas and low-income urban areas include the National Health Service Corps—young physicians whose medical education is subsidized in return for postgraduate service in underserved areas. Also emerging are specially trained nurse practitioners to provide primary care in underserved areas.

Efforts are being made to build close functional links between mental-health and general-health services. Also under way are analyses of organizational changes aimed at more effective health services for children and for the elderly.

HEALTH-SERVICES RESEARCH

The search for improvement in health services will require not only carefully designed innovations, but also systematic assessments—the function of health-services research.

The need for accurate, dependable information on the organization, quality, availability, and cost of health services in the United States is becoming increasingly apparent to health care professionals, government officials, and the public. Management of the personnel, facilities, and technologies that comprise modern health care institutions requires information similar to that needed to manage other complex enterprises. The systematic comparison of alternative health care delivery systems will be a great challenge to the health sciences in the years ahead. However, such research still lacks a secure institutional base, either in government or among the nation's leading universities.

REFERENCES

1. Shapiro, S., E.K. Schlesinger, and R.E.L. Nesbitt. *Vital and Health Statistics* (National Center for Health Statistics) 3(4):1, 1965.

2. *Research Needs in Nephrology and Urology, Vol. 1.* Report of the Coordinating Committee, National Institutes of Arthritis, Metabolism and Digestive Diseases (DHEW Pub. No. [NIH] 78-1481). Washington, D.C., 1978.

3. MacMahon, B., and J.E. Berlin. Health of the United States Population. In: *Horizons of Health.* H. Wechsler *et al.* (eds.). Cambridge, Mass.: Harvard University Press, 1977, pp. 13–21.

4. Rice, D.P., J.J. Feldman, and K.L. White. *The Current Burden of Illness in the United States* (Occasional Paper of the Institute of Medicine) Washington, D.C.: National Academy of Sciences, 1976, p. 26.

5. Smith, T.W. The Heart and the Vascular System. In: Horizons of Health, H. Wechsler *et al.* (eds.). Cambridge, Mass.:Harvard University Press, 1977, pp. 179–197.

6. Moore, F.D. Surgical Care. In: Horizons of Health, H. Wechsler *et al.* (eds.). Cambridge, Mass.: Harvard University Press, 1977, pp. 343–353.

7. Melmon, K., and H. Morelli (eds.). *Clinical Pharmacology,* 2nd ed. New York: Macmillan, 1978.

8. Sackett, D.L., and R.B. Haynes. *Compliance with Therapeutic Regimens.* Baltimore: The Johns Hopkins University Press, 1976.

9. Shapiro, A.P. Behavioral and Environmental Aspects of Hypertension. *Journal of Human Stress* 4(4):9–17, 1978.

10. Levi, L. (ed.). *Society, Stress and Disease, Vol 1., The Psychosocial Environment and Psychosomatic Diseases.* London: Oxford University Press, 1974.

11. Rose, R.M., C.D. Jenkins, and M.W. Hurst. *Air Traffic Controller Health Change Study: A Prospective Investigation of Physical, Psychological and Work-Related Changes.* Boston: Boston University School of Medicine, 1978.

12. Levy, R.I., and M. Feinleib. Coronary Artery Disease: Risk Factors and Their Management. In *Heart Disease* (in press).

13. Havel, R.J. Classification of Hyperlipidemias. *Annual Review of Medicine* 28:195–209, 1977.

14. Truswell, A.S. Diet and Plasma Lipids—A Reappraisal. *The American Journal of Clinical Nutrition* 31:977–989, 1978.

15. Farquhar, J.W., *et al.* Community Education for Cardiovascular Health. *The Lancet,* 2:1192–1195, 1977.

16. Maccoby, N., *et al.* Reducing the Risk of Cardiovascular Disease: Effects of a Community-Based Campaign on Knowledge and Behavior. *Journal of Community Health* 3(2):100–114, 1977.

17. Puska, P. High Risk Hearts. *World Health.* Geneva: The World Health Organization, 1976, pp. 12–15.

18. Frei, E. Cancer. In: *Horizons of Health,* H. Wechsler *et al.* (eds.). Cambridge, Mass.: Harvard University Press, 1977, pp. 25–39.

19. Tomatis, L., *et al.* Evaluation of the Carcinogenicity of Chemicals: A Review of the Monograph Program of the International Agency for Research on Cancer (1971–1977). *Cancer Research* 38:877–885, 1978.

20. Isselbacher, K.J. The Gastrointestinal System. In: Horizons of Health, H. Wechsler *et al.* (eds). Cambridge, Mass.: Harvard University Press, 1977, pp. 243–251.

21. Yalow, R.S. Radioimmunoassay: A Probe for the Fine Structure of Biologic Systems. *Science* 200:1236–1250, 1978.

22. Kaplin, H.S. *Hodgkin's Disease.* Cambridge, Mass.: Harvard University Press, 1972.

23. Upton, A.C. "Prevention—The Ultimate Goal." Presented at Conference on Cancer Prevention—Quantitative Aspects sponsored by National Cancer Institute, NIH, held in Reston, Va., September 1978.

24. *Smoking and Health: A Report of the Surgeon General.* Prepublication copy, Office of the Assistant Secretary for Health, DHEW, January 1979.

25. Kety, S.S. The Biological Bases of Mental Illness. In: *Horizons of Health.* H. Wechsler *et al.* (eds.). Cambridge, Mass.: Harvard University Press, 1977, pp. 111–123.

26. McAlister, A.L. Tobacco, Alcohol and Drug Abuse: Onset and Prevention. In: *Disease Prevention: Report to the Surgeon General on Health Promotion and Disease Prevention, Part II.* Washington, D.C.: Institute of Medicine, 1979.

27. Hollister, L.E. Tricyclic Antidepressants. *The New England Journal of Medicine* 299:1106–1109, 1978.

28. Bloom, F., *et al.* Endorphins: Profound Behavioral Effects in Rats Suggest New Etiological Factors in Mental Illness. *Science* 194(4265):630–632, 1976.

29. Guillemin, R. Peptides in the Brain: The New Endocrinology of the Neuron. *Science* 202(4366):390–402, 1978.

30. Schally, A.V. Aspects of Hypothalamic Regulation of the Pituitary Gland. *Science* 202(4363):18–28, 1978.

31. Mendelson, J.H., and N.K. Mello. Alchohol and Drug Abuse. In: Horizons of Health, H. Wechsler *et al.* (eds.). Cambridge, Mass.: Harvard University Press, 1973, pp. 124–140.

32. *Report of the Task Force on the Use and Need for Chimpanzees of the Interagency Primate Steering Committee.* Bethesda, Md.: National Institutes of Health, 1978.

33. Geschwind, N. Neurological Disorders. In: *Horizons of Health,* H. Wechsler *et al.* (eds.). Cambridge, Mass.: Harvard University Press, 1977, pp. 161–175.

34. Motulsky, A.G. Family Detection of Genetic Diseases. In: *Early Diagnosis and Prevention of Genetic Diseases,* L.N. Went *et al.* (eds.). Leiden: Leiden University Press, 1975, pp. 101–109.

35. Motulsky, A.G. "Genetic Approaches to Chronic Common Diseases." *Centenary Lecture Series. I. Medical Genetics.* University of Western Ontario, London, Canada, March 1978, in press.

36. Omenn, G.S., and A.G. Motulsky. "Eco-Genetics: Genetic Variation in Susceptibility to Environmental Agents." *Genetic Issues in Public Health and Medicine.* Charles C Thomas: Springfield, Ill., 1978, pp. 83–111.

37. Rotter, J.I., and D.L. Rimoin. Heterogeneity in Diabetes Mellitus—Update 1978. Evidence for Further Genetic Heterogeneity Within Juvenile-Onset Insulin-Dependent. *The Journal of the American Diabetic Association* 27:599–608, 1978.

38. Rutstein, D.D. *Blueprint for Medical Care.* Cambridge, Mass.: The Massachusetts Institute of Technology, 1974.

39. Fine, J. Proposal for a National System of Peer Review. *Bulletin of Atomic Scientists* 33(7):38–43, 1977.

40. Rutstein, D.D., *et al.* Measuring the Quality of Medical Care. *The New England Journal of Medicine* 294:582–588, 1976.

41. Lewis, C.E., R. Fein, and D. Mechanic. *A Right to Health: The Problem of Access to Primary Medical Care.* New York: John Wiley and Sons, 1976.

42. Enthoven, A. "Incentives and Innovation in Health Services Organization." Paper delivered at the Annual Meeting of the Institute of Medicine, October 26, 1978.

43. *Computed Tomographic Scanning* (IOM Pub. No. 77-02, Washington, D.C.), April 1977.

44. Frederickson, D.S. Health and the Search for New Knowledge. *Daedalus* 106(1):159–170, 1977.

45. *Evaluation of Poliomyelitis Vaccines* (IOM Pub. No. 77-02, Washington, D.C.), April 1977, p. 15.

46. Comroe, J.H. The Evolution of Biomedical Science: Past, Present, Future Perspectives. *Journal of Medical Education* 52:3–10, 1977.

47. Hamburg, D.A., and S.S. Brown. The Science Base and Social Context of Health Maintenance: An Overview. *Science* 200(4344):847–849, 1978.

BIBLIOGRAPHY

Aging and Medical Education (IOM Pub. No. 78-04, Washington, D.C.), September 1978.

The Application of Advances in Neurosciences for the Control of Neurological Disorders. Geneva: World Health Organization, 1978.

Arteriosclerosis: The Report of the 1977 Working Group to Review the 1971 Report of the National Heart and Lung Institute Task Force on Arteriosclerosis, December 1977 (DHEW Pub. [NIH]No. 78-1526), 1978.

Assessing Quality in Health Care: An Evaluation (IOM Pub. No. 76-04, Washington, D.C.), November 1976.

Barchas, J. *et al.* (eds.). *Psycho-Pharmacology: From Theory to Practice.* New York: Oxford University Press, 1977.

Baselines for Setting Health Goals and Standards. Papers on the National Health Guidelines (DHEW Pub. No. [HRA] 76-640), September 1976.

Berger, P., B. Hamburg, and D. Hamburg. Mental Health: Progress and Problems. *Daedalus* 106:261–276, 1977.

Besser, M. (ed.). *Medicine. 1977.* New York: John Wiley and Sons, 1977.

Beyond Tomorrow: Trends and Prospects in Medical Science, A Seventy-Fifth Anniversary Conference. New York: The Rockefeller University, 1977.

Biomedical Research in the Veterans Administration (NRC Committee on Biomedical Research in the Veterans Administration). Washington, D.C.: National Academy of Sciences, 1977.

Breslow, L., and A.R. Somers. The Lifetime Health Monitoring Program: A Practical Approach to Preventive Medicine. *The New England Journal of Medicine* 296(11):601–608, 1977.

Brown, S.S. *Policy Issues in the Health Sciences: A Staff Paper* (IOM Pub. No. 77-002, Washington, D.C.), October 1977.

Bunker, J.P., B.A. Barnes, F. Mosteller (eds.). *Costs, Risks, and Benefits of Surgery.* New York: Oxford University Press, 1977.

Cairns, J. *Cancer: Science and Society.* San Francisco: W.H. Freeman, 1978.

Carlson, R.J., and R. Cunningham (eds.). *Future Direction in Health Care: A New Public Policy.* Cambridge, Mass.: Ballinger Publishing Co., 1978.

Comroe, J.H., and R.D. Dripps. Scientific Basis for the Support of Biomedical Science. *Science* 192(4235):105–111, 1976.

Conference on Health Promotion and Disease Prevention (Vol. I, Themes) (IOM Pub. No. 78-002, Washington, D.C.), June 1978.

Conference on Health Promotion and Disease Prevention (Vol. II Summaries). (IOM Pub. No. 78-003, Washington, D.C.), June 1978.

Cooper, B.S., and D.P. Rice. The Economic Cost of Illness Revisited. *Social Security Bulletin* (DHEW) February: Vol. 39, pp. 21–36, 1976.

Davis, K., and C. Schoen. *Health and the War on Poverty: A Ten-Year Appraisal.* Washington, D.C.: The Brookings Institution, 1978.

The Elderly and Functional Dependency (IOM Pub. No. 77-04, Washington, D.C.), June 1977.

Fisher, K.D., and A.U. Nixon (eds.). *The Science of Life: Contributions of Biology to Human Welfare.* New York: Plenum Press, 1972.

Forward Plan for Health, FY 1978–82. U.S. Department of Health, Education and Welfare, Public Health Service (DHEW Pub. No. [OS] 76-50046), 1976.

Freinkel, N. (ed.). *The Year in Metabolism: 1977.* New York: Plenum Publishing Corporation, 1978.

Hamburg, D.A., and H.K.H. Brodie (eds.). *American Handbook of Psychiatry.* New York: Basic Books, Inc., 1975.

Health in America: 1776–1976. U.S. Department of Health, Education, and Welfare, Public Health Service, Health Resources Administration (DHEW Pub. No. [HRA] 76-616), 1976.

Health in the United States, 1975 (DHEW Pub. No. [HRA] 76-1232), 1976.

Health in the United States 1976–1977 (DHEW Pub. No. [HRA] 77-1232), 1977.

Lalonde, M., *A New Perspective on the Health of Canadians.* (Information Canada, Ottowa), 1975.

Luria, S.E. *Life: The Unfinished Experiment.* New York: Charles Scribner's Sons, 1973.

McLachlan, G. (ed.) *A Question of Quality?: Roads to Assurance in Medical Care.* New York: Oxford University Press, 1976.

Papers on the National Health Guidelines: The Priorities of Section 1502 (DHEW Pub. No. [HRA] 77-641, Washington, D.C.), 1977.

Perspectives on Health Promotion and Disease Prevention in the United States (IOM Pub. No. 78-001, Washington, D.C.), January 1978.

Preventive Medicine USA. Task Force reports sponsored by The John E. Fogarty International Center for Advanced Study in the Health Sciences, National Institutes of Health and the American College of Preventive Medicine. New York: Prodist, 1976.

Priorities for the Use of Resources in Medicine (DHEW Pub. No. [NIH] 77-1288, Bethesda, Md.), 1976.

Recommendations for a National Strategy for Disease Prevention. Atlanta, Ga.: U.S. Department of Health, Education, and Welfare, Center for Disease Control, June 30, 1978.

Report of the President's Biomedical Research Panel (DHEW Pub. No. [OS] 76-500, Washington, D.C.), 1976.

Report of the President's Biomedical Research Panel, Appendix A: The Place of Biomedical Science in Medicine, and the State of the Science (DHEW Pub. No. [OS] 76-501), 1976.

Respiratory Diseases: Task Force Report on Prevention, Control, Education (DHEW Pub. No. [NIH] 77-1248, Washington, D.C.), March 1977.

Rogers, D.E. *American Medicine: Challenge for the 1980s.* Cambridge, Mass.: Ballinger Publishing Company, 1978.

Segal, J. (ed.). *Research in the Service of Mental Health: Report of the Research Task Force of the National Institute of Mental Health* (DHEW Pub. No. [ADM] 75-236, Rockville, Md.), 1975.

Shannon, J. Federal Support of Biomedical Sciences: Development and Academic Impact. *Journal of Medical Education.* 51(7):1–98, 1976.

Smith-Exton, A.N., and J.G. Evans (eds.). *Care of the Elderly: Meeting the Challenge of Dependency.* New York: Grune & Stratton, 1977.

Somers, A.R., and H.M. Somers. *Health and Health Care.* Germantown, Md.: Aspen Systems, 1977.

Surgery in the United States: A Summary Report of the Study on Surgical Services for the United States. The American College of Surgeons and The American Surgical Association, 1975.

Tanner, J.M. (ed.). *Developments in Psychiatric Research.* London: Hodder and Stoughton, 1977.

Usdin, E., D.A. Hamburg, and J.D. Barchas (eds.). *Neuroregulators and Psychiatric Disorders.* New York: Oxford University Press, 1977.

Walsh, J. *The Biomedical Sciences—1975: Report of a Macy Conference.* New York: Josiah Macy Jr. Foundation, 1975.

Wechsler, H., J. Gurin, and G.F. Cahill (eds.). *The Horizons of Health.* Cambridge, Mass.: Harvard University Press, 1977.

White, A. *et al. Principles of Biochemistry,* 6th ed. New York: McGraw-Hill, 1978.

White, K., and M.M. Henderson. (eds.). *Epidemiology as a Fundamental Science: Its Uses in Health Services Planning, Administration, and Evaluation.* New York: Oxford University Press, Inc., 1976.

9 Toxic Substances in the Environment

INTRODUCTION

A principal effect of technological development during recent decades has been the accelerating rate of human alterations of land, vegetation, water, and air. Most visible is the transformation of the American landscape. Cities spread outward, carrying residences, factories, and roads, while inner sections are rebuilt. Streams are dammed and channeled; wetlands drained or restored; forests cleared, reserved, thinned, and replanted. Farmlands are affected by changes in patterns of crops, cultivation practices, and uses of fertilizer, pesticides, herbicides, and water. Mineral lands are stripped and sometimes reclaimed.

Each transformation affects in some fashion the basic processes of the interlocking systems of land, water, and air that support life. Each is intended to enhance the capacity of parts of these systems to serve human needs—for food, fiber, energy, transport, recreation, and solitude. Each, however, may also entail hazard to other parts of the system; each may reduce the productive capacity of the earth and perhaps impair the health and welfare of the people whom the changes are expected to benefit. The hazard may result from inadvertently altering a system, as when land cultivation disturbs stream flow, or from exposing people to new risks, as when a hospital is constructed in an active seismic zone.

Some of these linkages are direct and relatively measurable—such as the obliteration of farmland by the asphalt parking lot of an industrial plant, thus impeding groundwater infiltration in exchange for new manufacture. Others are far more complex and difficult to quantify: For example, the cutting of a forest reduces the standing stock of carbon and possibly renders global climate more vulnerable to carbon dioxide (CO_2) buildup, as noted in Chapters 1 and 5.

In this chapter we review only one facet of the massive changes under way: the issues raised by the presence of toxic substances in the environment—in particular those related to the effects of man-made chemicals on human health. An immense array of substances in the natural environment may also be toxic to plants, animals, and humans. Indeed, some of the chemicals that are necessary for life are nevertheless toxic or even lethal in high doses. Ions of zinc and copper, for example, are essential nutrients, but fatal at high concentrations. Even salt and oxygen are toxic in large enough amounts. Natural selenium in certain soils causes disease in grazing animals and tooth deformation in humans.[1] Marine plants and sponges generate a great variety of halogenated organics and have the ability to tolerate and detoxify such materials.[2] Many chemical interactions in undisturbed natural systems are imperfectly understood or barely suspected. For example, natural sources appear to account for the methyl mercury found in the livers of some marine fish and mammals. However, the adverse effects of methyl mercury on these animals may be ameliorated when selenium is also present.[3]

Interactions of this sort among humans are extraordinarily difficult to study. Epidemiologists find that their subjects move from place to place, drawing upon different soils and groundwater. In any given area, new substances are being introduced, thus changing the background. Nevertheless, there is great pressure on science to understand the consequences of the increasing introduction of man-made substances into the environment and to consider possible ways of dealing with them. Of most public concern has been the diffusion of man-made chemicals for household, agricultural, and industrial purposes. They attract attention because of dramatic instances of local contamination and because they are suspected of contributing to broader injury both to the health of humans and of ecosystems.

There is a linkage between the effects of toxic chemicals on human health and on the environment. For example, air pollutants such as ozone and sulfur dioxide, which can damage crops, are also associated with human respiratory ailments.

Much remains to be learned. Some toxic chemical hazards are well known. Other chemicals are suspect. They arouse concern but their dangers have not been clearly identified. Furthermore, an earlier viewpoint, that there are low levels of exposures to chemicals that are totally without effect (totally safe levels), is currently being challenged in some areas, especially in regard to cancers. At the same time, rapid advances in analytical techniques make it possible to detect smaller and smaller quantities of hazardous chemicals. Thus, for a number of reasons we are moving towards a position where quantitative assessment of both risks and benefits must be increasingly used. Since it is impossible to avoid all risk, society will have to decide in each case if the benefit from a hazardous material justifies the risk.

Environment is emerging as a factor in today's major diseases. The preceding chapter has shown that since 1900 cardiovascular disease and cancer have replaced microbial infection as the major health threats to Americans. Among the cardiovascular diseases "the elucidation of specific genetic and environmental factors" implicated in hypertension remains a scientific challenge. Water hardness is one of the dozen or so factors deserving investigation as a basis for preventing coronary disease.

Environmental factors, defined broadly to include diet and cigarette smoking, are now considered important in the origin of cancer. This recognition is based in part on the large geographic variations in incidence of specific cancers, and it encourages speculation as to what kind of background or anthropogenic factors may account for the differences in incidence among people in different regions of the United States. The large number of presumed carcinogens, the highly variable conditions of exposure to them, the long latent periods after exposure, and the high mobility of population make it extremely difficult to work out the epidemiology of cancer. The five years ahead will see major efforts by epidemiologists to refine their understanding of the large variations in geographic incidence.

The situation at present is summed up in the preceding chapter:

> Although some man-made chemicals have been identified as potential carcinogens, the firmly established incidence of cancer from these compounds accounts for only a tiny fraction of all cancers. Cigarette smoking is the one environmental factor for which firm data demonstrate a strong association; occupational exposures are also an important source of environmental carcinogenesis. Other environmental factors, including those of natural origin, that are probably responsible for many and perhaps most cancers remain unknown. Research on these problems will intensify during the next five years.

Interest in the outcome and application of that research will expand even more rapidly as more refined measurements reveal the presence of substances whose presence was unsuspected, as additional carcinogenic factors are identified, and as more toxic chemicals are found. Even if there were no expansion in the production and number of chemical compounds, the question of the hazard of those already circulating in the environment would still be important. To the extent that production enlarges, that question will command even more public attention.

GROWTH OF CHEMICAL TECHNOLOGY

Over the past three decades, the production and use of industrial chemicals expanded a hundredfold. The welfare of our nation has come to depend in no small measure upon these and other chemical products. Synthetic fertilizer is essential to food production at current levels in many parts of the world. Reliability and output of crop yields would be substantially diminished without agricultural chemicals such as pesticides and weed killers. Almost all pharmaceuticals are synthetic chemicals. The smelting and purification of metals are chemical processes. Most of the gasoline, rubber, plastic, adhesives, detergents, textiles, antifreezes, disinfectants, cosmetics, solid-state devices, films, paints, and much of the building material we use are synthetic.

For many years the American people looked upon expansion in the manufacture and use of synthetic chemicals as a major area of national progress. Many of the newly generated substances apparently have no harmful effects at current levels. Some of them, however, are toxic, and over the past decade we have come increasingly to realize that this advance in productivity and convenience has been accompanied by both real and potential impacts on human and environmental health now and in the future. This focuses much public attention on the dangers inherent in the uncontrolled manufacture and use of synthetic chemicals and on the ways in which we can in the future improve the safety of the products and processes we use.

Some 80 generally distributed chemicals (not immediately used in the synthesis of other chemicals) are each produced in amounts greater than 100 million pounds annually. Over 25 million pounds each of 25 other such chemicals are manufactured yearly.[4]

Figure 29 shows the growth of the synthetic organic chemical industry in the United States since 1917. Reported annual production of manufactured chemicals includes some 10,000 individual chemicals. (However, about 70 percent of the total output is made up of some 500 large-volume chemicals.[5]) Many more, produced in small volume, are not reported. It has been estimated that approximately 70,000 chemicals are in some degree of current use, with perhaps 1,000 new chemicals being introduced each year.[6] Growth rates for six of the major basic compounds used in the manufacture of large volume chemicals during the past two decades are shown in Figure 30. Some of these starting chemicals are utilized in a wide variety of ways. As an example, the many uses of ethylene, the largest volume starting chemical, are shown in Figure 31. Whether or not these production rates continue, the number and volume of substances released into the environment and their uses are large and promise to remain so.

In the face of increasing demand and production over the past several years, considerable advance has been made in curbing pollution and controlling the spread of toxic chemicals. Federal regulations have

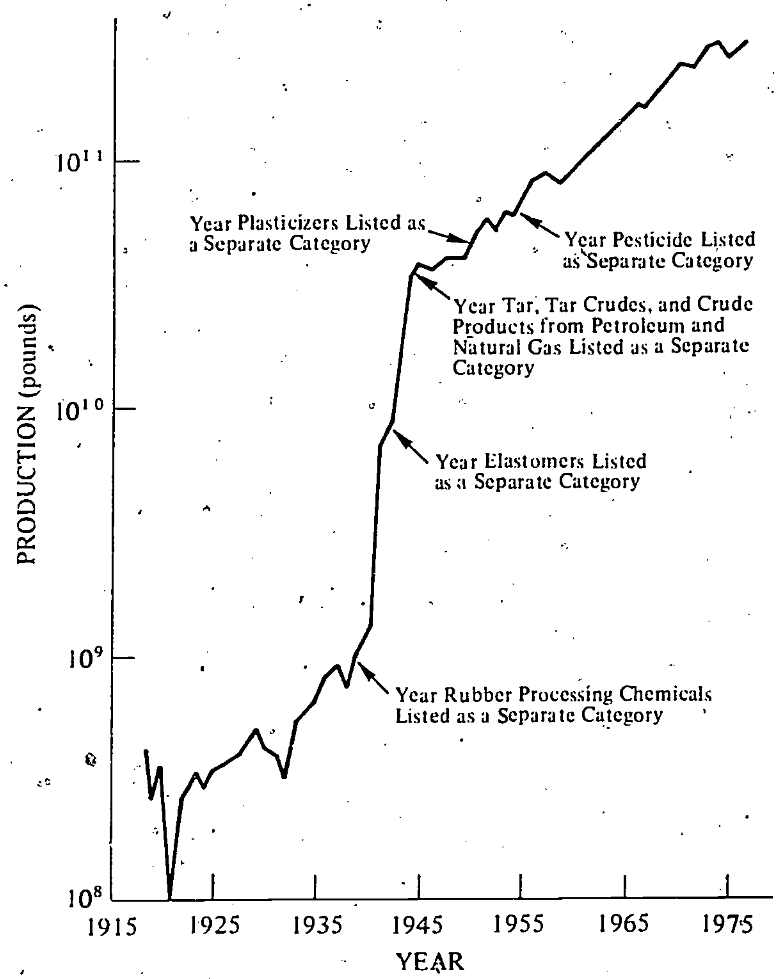


FIGURE 29 Synthetic organic chemical production. (Synthetic Organic Chemicals: U.S. Production and Sales, 1918-1976, Washington, D.C., U.S. International Trade Commission)

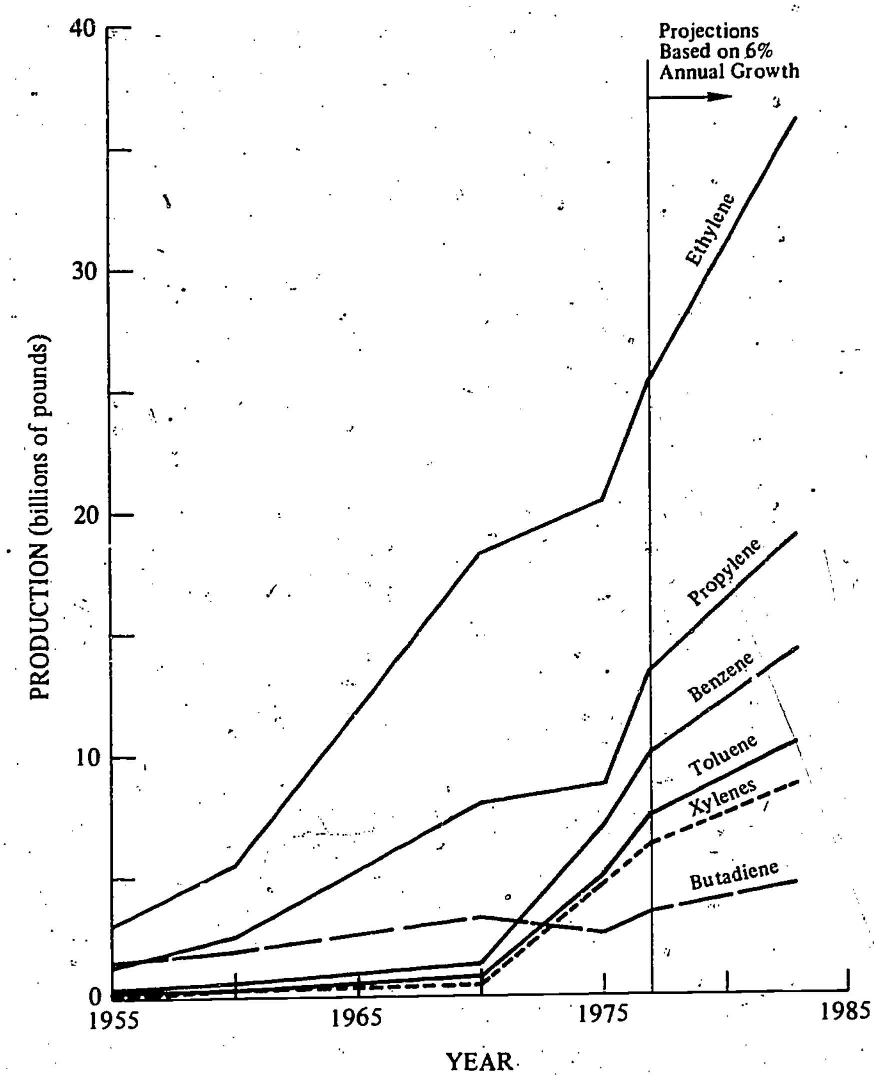


FIGURE 30 Production trends for major petrochemical starting materials. (Synthetic Organic Chemicals, U.S. Production and Sales, various years, Washington, D.C., U.S. International Trade Commission [Benzene production estimates for 1955, 1960, and 1970 courtesy of *Chemical Economics Handbook*, SRI International, Menlo Park, Calif.])

brought substantial reductions in air pollutants such as sulfur dioxide, workplace contaminants such as vinyl chloride, and industrial-waste discharges into rivers and lakes. Some of this progress has resulted from the voluntary cooperation of industry with federal agencies in restricting uses and production of PCB's, and in early reduction in the use of halocarbons as an aerosol propellant.

Much remains to be done. The pressing need now is for better knowledge—both to determine the dimensions and importance of suspected chemical hazards and to improve regulatory efforts to control recognized problems. Only thus can we separate large from small problems in order to set our priorities. In many cases the necessary knowledge will be slow in coming or never completely available.

The following discussion of a few known and suspected chemical hazards (both man-made and natural) is intended to illustrate the pervasiveness of the problems, gaps in our knowledge of causes and effects, and ways that scientific information can be brought to bear in dealing with issues that threaten to become yet more troublesome.

PERSISTENCE, TRANSFORMATION, AND MOVEMENT OF CHEMICALS

PERSISTENCE

Some chemicals are extremely persistent in a natural environment—highly resistant to physicochemical breakdown and to biological degradation. Such persistent chemicals collect and remain in the sediments of lakes and rivers. Some move up the food chain from microorganisms to other creatures, to fish, and finally to humans, increasing in tissue concentration at each

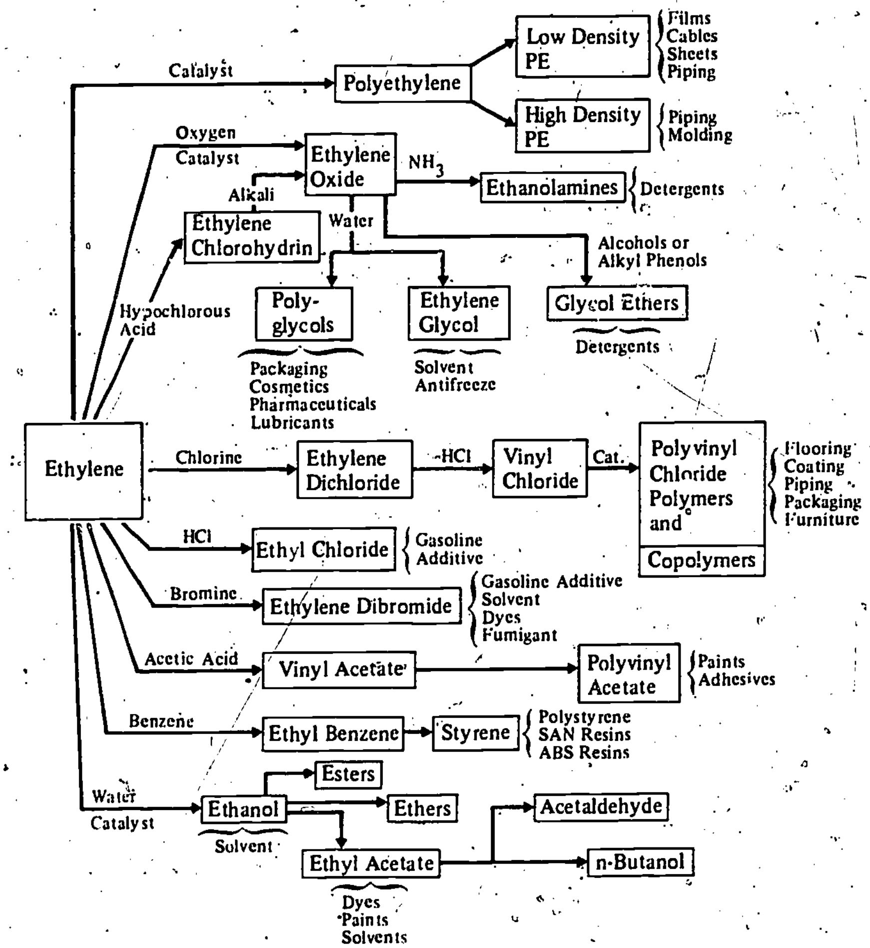


FIGURE 31 Uses of ethylene.

stage. The accumulations of these chemicals in the tissue of fish are thus higher than the average levels in surrounding waters.[7] The stability of these persistent chemicals tends to make possible their diffusion over substantial areas by both air and water (especially air over long distances).[8] The resulting contamination can be very widespread. For example, the highly persistent PCB's are found in both humans and nonhuman species throughout the world.[9]

CHEMICALLY TRANSFORMED, SECONDARY, AND TRACE CHEMICALS

Chemicals can be transformed by various means—for example, by reaction with other chemicals or by the action of bacteria. Bacteria in the sediments of streams and lakes can act on inorganic mercury and produce the more toxic methyl mercury. Because methyl mercury derivatives are volatile, they can contribute mercury to the atmosphere, which can then spread over long distances and add to the atmosphere's natural burden of mercury.[10]

Secondary chemical pollutants derive from complex processes in which a totally new toxic agent is generated. For example, ozone and peroxyacetyl nitrate (PAN) are produced in the lower atmosphere through the photochemical action of sunlight on nitrogen oxides and hydrocarbons in the air (from auto exhausts, power plants, and so on).[11] Ozone and PAN can damage agricultural crops. (It is of interest here to note the relationship between site and effect. In the upper atmosphere, ozone filters some of the potentially carcinogenic ultraviolet rays from sunlight and is beneficial rather than harmful.)

Sometimes a chemical product presents a relatively minor hazard compared to that posed by trace amounts of contaminants that it contains. An example is dioxin (tetrachlorodibenzedioxin or TCDD), a substance found in the defoliant 2,4,5-T. Although dioxin is present in only minute quantities, it has been a major cause of concern about the defoliant's possible toxicity.[12–14] We are now aware of the importance of some trace contaminants even at concentrations of a few parts per billion.

Deposits of Chemical Wastes

One of the most perplexing aspects of chemical pollution is the accumulation of chemical wastes in landfills and sediments. Hudson River sediments, for instance, contain the accumulation of many years of PCB discharges. Dredging is being proposed to remove the PCB's from the river. But it has not been shown that this approach would be successful. Indeed, it could make the situation worse by simply spreading the contamination. Moreover, there would be the problem of where and how such a vast volume of contaminated material (an estimated 1.7 million cubic yards)[15] could be safely disposed of with no environmental or health impact.

Other approaches have been suggested: overlaying the sediment with impervious material, sequestration, and bacterial or chemical degradation. A key question in each case is whether it is possible to improve on natural degradation and sequestration, slow and inadequate though they may be.

Widespread chemical dumping has been a pattern of the past, and some dump sites have been later used for landfill. The Love Canal near Niagara Falls, New York, is one such site. Thus far, the human health impacts of these abandoned and refilled chemical dumps have not been fully evaluated.[16]

MAJOR PATHWAYS OF CHEMICAL POLLUTION

Air Pollutants

The nation has been concerned about air pollution for some years. The general dimensions of the health impacts of pollutants such as the sulfur oxides, ozone, nitrogen oxides, carbon monoxide, and particulates are only now being defined.[17] Current estimates vary widely. Major gaps in our knowledge of these health effects persist—especially detailed information on the health effects of actual concentrations of specific pollutants on specific segments of the population.

It is generally believed, for example, that acid particulates alone or in combination with irritant gases contribute to the respiratory effects of air pollution. If the responsible chemical species and various combinations among a constellation of chemical agents could be identified, it should be possible to design more efficient and selective controls.[18]

More refined knowledge of the quantitative impacts of air pollutants, in terms of specific agents and disease patterns, is also needed. Irritant air pollutants such as sulfur oxides, nitrogen oxides, acid particulates, and ozone are thought to contribute to morbidity and mortality. But not all parts of the population are equally endangered by these pollutants. The young and the old seem to be more susceptible, especially those elderly persons who already suffer from cardiorespiratory disease.[11,19,20] Knowledge is needed of the role of host factors, such as age, sex, prior or concurrent disease, and genetic makeup. Also, a better understanding of the interaction and role of the various components of the acid particulate complex could aid in refining and improving control procedures.

The role of air pollution in contributing to lung cancer remains controversial. It is generally agreed that there is an excess of lung cancer in urban as compared with rural areas and that carcinogenic and cancer-enhancing air pollutants are in higher (but not massive) concentrations in urban than in rural areas. However, in view of other differences in the two settings, there is no general consensus that the urban-rural differences in lung cancer rates can be confidently attributed to air pollution. One still-controversial estimate suggests that general air pollutants can increase the incidence of lung cancer in males who smoke by perhaps 10 percent beyond the average effect of cigarette smoking alone.[21] This estimate does not apply to nonsmokers.

Water Pollutants

There are two well-known pathways for the chemical contamination of lakes, rivers, and groundwaters: current discharges or runoffs, and gradual releases from accumulated deposits of chemical wastes. During recent years, persistent pesticides (such as Mirex, Kepone, and DDT), mercury, PCB's, and other chemicals have been found in many of our groundwaters and freshwater lakes and streams.

A third, perhaps significant pathway derives from the internationally employed chlorination of water supplies. Because of chemical treatment, microbial and viral contamination of water bodies, the classical potential hazard to human health, is no longer an active threat of massive dimensions. The control of intestinal infections by the chlorination of water supplies has been one of the most dramatic public-health advances of the past century. However, there is now concern that this may produce chemicals such as chloroform by interaction with otherwise harmless organic substances in the water. Laboratory data show that chloroform and several similar compounds cause liver cancer in rodents.[22] There is also debated epidemiological evidence of increases of cancer among populations using chlorine-treated water.[23]

Any change in this extremely successful chlorination process should be approached carefully. To remove those organic chemicals that can react with chlorine to produce chloroform or similar compounds, the Environmental Protection Agency has proposed charcoal treatment of water prior to chlorination in cities with populations above 75,000.[24] This additional treatment, however, may have undesirable side effects, and would be costly and an additional burden on public health authorities. Hence, alternative methods for the control of waterborne pathogens, such as ozone treatment, are being investigated. The magnitude of the cancer risks involved in chlorination should, in addition, be more thoroughly assessed.

Trace quantities of many other organic compounds are present in water. Some of these compounds have been identified as weak carcinogens or mutagens in experimental animals or other biological systems.[25] There are currently no fully acceptable means for evaluating the health risks (if any) from drinking water containing trace amounts of such compounds; appropriate techniques should be sought.

Exposure in the Workplace

The workplace has been a tragic but revealing source of information about chemical carcinogens. Some 20 chemicals or processes have been linked to the incidence of cancer in exposed workers. Some, such as 2-naphthylamine, chromium metal production, and asbestos, have been well known for years. Others, such as vinyl chloride and the chloromethyl ethers, have been recognized more recently.[26] The contribution of occupational exposure to chemicals to overall incidence of cancer in the United States is uncertain; published estimates have ranged from 1 to 20 percent.[27]

Cancer-causing chemicals in the workplace are presently regulated through a number of procedures, such as allowable exposure concentrations for asbestos and vinyl chloride and stipulated work practices for certain chemical carcinogens. Several new approaches have been suggested, among them a permit system requiring the stipulation of control procedures and a total-enclosure approach, under which carcinogenic compounds would be isolated from workers and monitoring would involve the testing of tightness rather than allowable concentrations. An evaluation of new approaches might yield a reduction in risk for workers and better, more cost-effective controls.

Some occupational diseases, such as silicosis, coal miners' pneumoconiosis (black lung disease), and byssinosis (brown lung disease, associated with the vegetable fiber industries) have been with us for many years. Recently, attention has been given to the effects of chemicals on the reproductive and nervous systems. The soil fumigant, 1,2-dibromo-3-chloropropane, has been linked to sterility in production workers[28] and lowered sperm counts in field workers.[29] Some female anesthetists have suffered abortions and congenital defects in offspring.[30] The solvent methyl-*n*-butyl ketone has been associated with neurological disorders.[31]

What is urgently needed is an assessment of the prevalence and seriousness of the various occupational diseases to aid in setting control priorities. Also needed are evaluations in the laboratory of many of the workplace chemicals now in use and or to be introduced.

The Marketplace

Among consumer products, both the degree of control and appreciation of hazards remain very limited. A case in point was the rapid introduction of the flame retardant chemical Tris (2,3-dibromopropylphosphate) for use in treating children's sleepwear. The chemical is now restricted after being found in laboratory tests to be mutagenic and carcinogenic.[32] The Tris incident clearly suggests that more rigorous pretesting is needed before product introduction.

Many household products (so-called under-the-sink chemicals and agents) have not been adequately tested, especially for chronic effects. There is a clear need for better understanding of the effects of chemicals used in consumer products, with particular emphasis on chronic effects and chemical interactions.

Toxins in food may be introduced by man or occur naturally. Aflatoxin, for example, is a product of mold growth that can contaminate peanuts, cotton seed, and other agricultural products. There is strong suspicion of an association of increases in liver cancer in Asia and Africa with the consumption of food contaminated with aflatoxins.[33] Even more important is the epidemiological evidence that general dietary choices may substantially influence the incidence of human cancer.

Man-made chemicals in food have aroused widespread but sometimes inappropriate concern. Such chemicals include food additives that have been used for decades under the so-called GRAS (generally regarded as safe) concept. The GRAS list is being reexamined, a process that may raise many problems of hazard evaluation.

An important legal element is the Delaney Clause, which forbids the deliberate use, in any amount, of any food additive found to cause cancer in man or animal. The wisdom of the clause has been much debated, and the issues will be clarified as valid test cases are examined.

Two recent cases have attracted much public interest. Recent laboratory tests showed that saccharin is a definite but weak carcinogen.[34] These data would seem to require action under the Delaney Clause, even though the risks to humans from saccharin seem very modest.

Nitrite, a preservative used in meats such as salami, wieners, bacon, ham, and so on, is the focus of another controversy. It can combine with secondary amines in the stomach to form nitrosamines, a class of chemicals known to be carcinogenic.[35] Although nitrite does provide protection against botulism, the same result might be achieved with smaller amounts of the chemical than are now being used. However, higher levels are presumably favored by consumers (and meat processors) because of the red color they bring to meats. A further complication is that nitrates—found in most diets—are reduced naturally in the body to nitrites in significant quantities.

The principle underlying the Delaney Clause is clearly acceptable for compounds with minimal benefits or for which noncarcinogenic substitutes are available. But quantitative risk assessment becomes necessary when it is asserted that there are benefits (including benefits to health) and no ready substitutes, or when an agent is of natural origin and its complete removal impractical. Even though assessments of both risks and benefits have substantial uncertainties at this time, they should under these circumstances be taken into account in regulatory decisions.

Cigarette smoking and alcohol and drug abuse are also major health problems. It is useful to note, however, that smoking can substantially increase the adverse health effects of toxic substances in the environment. The most vivid example is the interaction of smoking with certain inhaled asbestos fibers—raising the risk of lung cancer some eight times.[36] As discussed earlier in this chapter, there also seems to be an interaction between smoking and air pollution.

COMPOUNDS OF SPECIAL CONCERN

SOME PROBLEM METALS

Compounds of some of the old, familiar metals such as arsenic, mercury, cadmium, and lead are still potentially serious hazards to human health. There appears to be clear evidence that some arsenic compounds are carcinogenic in humans, especially in association with pollutants such as sulfur dioxide. There is also evidence that high natural levels of arsenic compounds in water supplies are associated with cancer in certain population groups such as in Taiwan. Strangely, tests of arsenic carcinogenicity in animals have all proved negative.[37]

Much public attention was given to methyl-mercury poisoning after the discovery that many Japanese families eating contaminated fish caught in Minimata Bay were paying a heavy price. Clinical studies revealed damage to the central nervous system and congenital defects in the offspring of exposed mothers.[10] Sale of fish containing mercury is now regulated in the United States, although the actual health effects of exposure to low levels of methyl mercury are unknown. Since methyl mercury is found in fish from many U.S. waters, including the sport fishing waters of the Northeast, health and wildlife-management officials face difficult decisions.[38]

Exposure to cadmium in the workplace is associated with respiratory and renal disease. More significant to the general population is the fact that cadmium largely remains in the body once it is absorbed. Thus, a newborn infant starts life with only a few micrograms of cadmium in his body. But by the age of 60 or 70, the cadmium burden, mainly from food, in his kidneys and liver is many milligrams—a 1,000-fold increase. The level in the kidneys by this age is approximately one-fourth of the average required to produce renal disease.[39] It is not known whether and to what extent kidney deterioration normally associated with the aging process is due to the natural accumulation of cadmium in normal kidneys.

Cigarette smoke is another source of cadmium. Absorption of cadmium through the lungs from cigarette smoke is, in contrast to absorption from food, very efficient—up to 40 percent.[39,40] As a result, smokers generally have higher cadmium burdens than nonsmokers.

The ancient problem of lead poisoning still persists today, sometimes in acute forms. There are poorly maintained buildings in ghetto neighborhoods of older cities whose interior walls are covered by many layers of paint with high lead content. Over the years, significant levels of lead have been found in children from such areas, with noticeable adverse effects on mental development and neurological health.

There is also concern over so-called normal levels of lead in the body. Lead levels in the blood of city dwellers (heightened by the lead from auto exhausts) are higher than those of rural residents by some 25–30 percent. Our understanding of biochemical disturbances stemming from increased lead burdens in the body is beginning to suggest that moderately elevated levels may be associated with biochemical changes related to synthesis of hemoglobin in the oxygen-carrying pigment of red blood cells.[41] The practical implications of these findings are unknown.

Again, there is a lack of knowledge as to whether, how, and to what extent chronic low-level exposures to toxic metals contribute to adverse health effects.

SOME HALOGENATED HYDROCARBONS

Three halogenated hydrocarbons—DDT, PCB's, and PBB's (polybrominated biphenyls)—illustrate different patterns of distribution, exposure, and sources of exposure. All are fat-soluble chemicals that are slowly metabolized. They tend to accumulate in the fat where they remain as a source of continuing exposure to the rest of the body.

DDT saved millions of lives through its contribution to the control of malaria, but its wide use as an insecticide spread the chemical into the atmosphere, into lakes and rivers, and, through concentration in the food chain, into fish and humans.

PCB's were originally developed for use in electrical equipment, especially condensers and transformers. The special properties of these compounds, such as low flammability, low freezing point, and low volatility, also led to their extensive use as heat-exchange liquids, plasticizers, and so on. PCB's have now spread throughout the environment, reaching many parts of the biosphere, including humans.[9]

The carcinogenic potential of PCB's has caused concern. But the more serious problem may be the effects of PCB's on nursing infants. They are the most intensely exposed to the chemicals, since they are dependent on mothers' milk, which, in many parts of the country, contains PCB's.[42] Laboratory studies on nonhuman primates have shown that a dietary intake of PCB's has produced abortions and underweight offspring.[43] The application of these findings to humans is uncertain.

The problem of PBB's appears to be mainly restricted to Michigan where, through a packaging error, the chemicals were added to cattle feed supplement. This resulted in contamination of some cattle and the exposure of some people in the area through the consumption of milk and beef.[44] The extent of the human impact is still being evaluated.

Chlorinated pesticides, such as Chlordane, Mirex, Kepone, and so on, present dissemination problems similar to those of DDT, PCB's, and other such chemicals. In addition, laboratory experiments have shown that the chlorinated pesticides DDE, Mirex Lindane and Chlordane produce liver tumors in rodents.[45]

TCDD (dioxin) first came into prominence as an impurity in the herbicide 2,4,5-T. It has since been identified as being involved in several episodes of an accidental nature, the most recent of which was an explosion in Seveso, Italy.[46] Health effects of TCDD on humans are not well established. However, TCDD has been extensively studied in animals and has been found to produce skin disorders, to be teratogenic (producing congenital defects), and injurious to the immune system.[13]

ENVIRONMENTAL FACTORS IN DISEASE

The widespread publicity accorded carcinogenic chemicals has created apprehension in many people. The identification of a significant number of new chemical carcinogens has led to the mistaken impression that nearly all chemicals can cause cancer. In fact, only a fraction of the chemicals tested are carcinogenic.

It is also important to recognize that the modest increase in the total cancer rate over the past few years is due to the prevalence of lung cancer associated with smoking; otherwise, cancer incidence has actually somewhat declined.[47] However, in terms of the impact of chemicals in the environment, these total incidence rates mean less than statistics on cancer rates by organ site, since different chemicals affect different organs. Lung cancer, mainly attributable to cigarette smoking, has gone up dramatically and is still increasing, especially among women. Stomach cancer has dropped sharply. Pancreatic cancer is rising.[48] Thus, the pattern is mixed and requires detailed analysis. Except for lung cancer, the other trends remain largely unexplained.

An important concept involving the role of external, or environmental, factors in carcinogenesis has developed over the past 20 years. Different ethnic groups have cancer patterns that seem to be related to their cultural patterns—such as diet, occupation, housing, etc.—rather than to their genetic makeup. Thus, native Japanese have a high stomach cancer rate. Their descendants in Hawaii have a significantly lower rate, and their descendants who live in the continental United States, a still lower incidence.[49] Ethnic groups tend to assume the cancer rate (by organ site) of the region to which they have migrated.

Thus cultural patterns, especially diet, may be the basis for these different rates. This does not mean that genetic factors are unimportant, since individual susceptibility, within specific populations, almost certainly must vary because of individual genetic variation. The range of such genetic variation may be more or less the same in different population groups.

The strong implication of these epidemiological findings is that cultural factors may be the source of major differences in cancer rates in various ethnic groups. If this inference is correct, then the excesses over the minimal background rate of cancer suggest that the additional cancers attributable to cultural factors are substantial.

This observation, together with data on such known cancer sources as smoking and occupational exposures, has led to the estimate that a large proportion (possibly high as 80–90 percent) of human cancer may stem from cultural factors,[49] including foods that have been part of the human diet for thousands of years.

There is a need to refine our knowledge of the contributory causes of cancer, whether from diet, occupation, or other sources. A number of approaches are required to develop this knowledge, including epidemiology, the continued testing of agents (alone, and together with other carcinogens and enhancing agents), and studies of mechanisms of action. Risk-assessment procedures can be improved by refining methods of translating laboratory data into human risk evaluations, especially in extrapolating from high doses to low doses over the long term.

In the long run, critical to all such evaluations will be knowledge of the minimal rate of carcinogenesis in man, inherent in our own biology, and the mechanism of such carcinogenesis. And it seems equally important that understanding be sought of the fact that so seemingly bizarre a diversity of chemical structures all engender cancerous transformation of animal cells.

Cardiovascular Disease

Evidence continues to accumulate showing lower rates of cardiovascular disease in hard-water areas than in areas where the water is soft.[21] The issue is still unresolved. We need more understanding of the role of water hardness in cardiovascular disease to determine whether hard water is beneficial and, if so, what compounds are responsible.

Germ-Cell Mutations and Injury to the Reproductive Process

There is a strong possibility that mutations of germ cells could result in damage that would become apparent in future generations. Mutations arise naturally from imperfections in the biochemical process of replication. In fact, without such mutations organisms could not evolve; but additional mutations are caused by chemicals and radiation, and the majority of these are deleterious. Indeed, the Environmental Protection Agency has already moved ahead on a proposal that the potential of pesticides to produce heritable mutations should be systematically tested as part of the normal registration procedure.[50]

Intense attention is currently being given to the potential of chemicals to produce birth defects. Well-developed methods of testing in this important area are now available, but there is a further need to focus more generally on the impact of chemicals on reproductive physiology and on the reproductive process in general. There are instances of sterility and other alterations of reproductive physiology that were caused by certain chemicals. This field has been relatively neglected.

ADVERSE EFFECTS ON NONHUMAN SPECIES

There is a large variety of materials and processes in natural systems that need to be understood for purposes of resource management and as background for assessing the effects of man-made additions to the natural background. Birds and fish are widely contaminated by such persistent chemicals as PCB's[9] and methyl mercury.[51] Acid rain from sulfur oxides and nitrogen oxides in the air appears to be affecting aquatic life, especially in lakes in the Northeast.[18] The movement of sewage effluent, agricultural wastes, and fertilizer residues is also a major problem. Such pollution changes aquatic ecosystems by increasing the volume of nitrates in lakes and streams.[52] As noted earlier, interactions among a pollutant's impacts on the environment, nonhuman species, and human beings are not infrequent. When a lake is contaminated by acid rain, for instance, acidification can alter the leaching of metallic compounds from the lake bed,[53] thus increasing the exposure of aquatic biota (and people, if the lake is a source of drinking water) to toxic substances that are a part of the natural sediments. We need better understanding of the extent, nature, and importance of the effects of acid rain. The damaging effects of individual air pollutants on some agricultural crops have been fairly well described. However, knowledge of the impacts of some chemicals remains inadequate, as does understanding of the role of synergism between air pollutants.

The impacts of chemicals on nonhuman species involve a rather different set of considerations than is the case with human beings. When assessing the effects of chemical contamination on a natural ecosystem, there is less concern over injury to a single individual (or small numbers of individuals) than with the survival of populations, communities, or even entire species. Weakening of the hold of a particular species in an ecosystem could lead to the proliferation of other species; for example, the emergence of the red spider mite after an area has been sprayed with DDT.[54] The consequences for a natural ecosystem could vary, depending on the circumstances. We need more and better information on the effects of pollutants on natural systems. Efficient pretesting methods are needed to predict these effects as well as the degree of natural resilience that such systems might have.

SCIENTIFIC, TECHNICAL, AND POLICY RESPONSES

REDUCING OUR RISKS

Safety evaluation of chemicals was relatively simple in the past. It was usually assumed that there is a safe level for every chemical and that this level can be determined through modest laboratory tests (to which a safety factor was applied). This concept was based in part on the assumptiom that a threshold exposure level existed below which no adverse effects would occur. Paracelsus (1493–1541) wrote: "All things are poisons, for there is nothing without poisonous qualities. It is only the dose that makes a thing a poison." For a wide range of chemicals that statement holds true today.

The situation may, however, prove different for mutagens and carcinogens as compared to other poisons.* Quite possibly for some carcinogens there is no "threshold," no concentration below which these compounds are safe; it may turn out that, with respect to these materials, health problems can be expressed only in quantitative, statistical terms. A trace of some carcinogens may inevitably occasion a minute but nonzero increase in the probability of cancer, whereas a larger dose will lead to a much greater probability. The situation is like that of crossing a street; some crossings are more dangerous than others, but none is absolutely safe. Devising proper procedures under these circumstances involves difficult social and political as well as medical questions.

The concept of a threshold for carcinogens has been intensely debated over the past few years. At this time, it is fair to say that, although there may be cases in which such thresholds appear to exist, they have not yet been reliably demonstrated.

Several instances are now known—for example, cigarette smoking and aflatoxin—in which the human dose–response curves for cancer appear to pass through or near zero and can be regarded as linear over the lower dose regions. In other words, even a very small dose would have an effect, and the effect is directly proportional to the size of the dose.

Although there are as yet no reliable examples, a true threshold may exist for some chemicals as cancer-causing agents. The probability that such thresholds exist is suggested by the presence of mechanisms that are known to inactivate harmful chemicals in varying degrees.†

The Congress has set up a number of federal agencies designed to protect the public against poisonous chemicals and advance the cause of safety. The Environmental Protection Agency has primary authority over much of this area. The Toxic Substances Control Act (TSCA), the Clean Air Act, the Safe Drinking Water Act, and the Occupational Health and Safety Act have been added to earlier legislation on pure food and drugs. Since many of these laws are new and since experience with their administration is limited, Congress will likely find ways to provide even more effective protection. We must learn to deal more rapidly with the large numbers of toxic and potentially toxic substances in commerce and in our environment and at the same time to encourage development of safer products, such as less harmful pesticides and new pharmaceuticals. Although much has been and is being accomplished, experience with the new laws has quite naturally illuminated some areas of difficulty.

A major uncertainty today relates to the concept embodied in the legislation established in TSCA of "unreasonable risk." Each of us voluntarily accepts certain risks. The dangers of smoking are well advertised. A cigarette smoker is essentially making a personal risk–benefit assessment. But there is no practical alternative to using the community water supply or breathing air. In many instances, as in choosing modes of travel or buying paint, the individual may not be fully aware of the possible risks involved. The administrator of TSCA is asked to evaluate whether the benefits from the manufacture of a particular chemical justify the imposition of involuntary risks. The situation is still more complex when the benefits accrue to a different group from those who suffer the risks.

The problems are not easy ones, and the assignments given by law to federal regulators are difficult in the extreme. The most difficult decisions, perhaps, are those of the Commissioner of Food and Drugs, who puts some patients at risk if he grants approval for a new and possibly useful drug and others at risk if he refuses it.

It is becoming recognized that total safety is unattainable and that benefits from the use of chemicals range from trivial (and not worth even a minor risk) to vitally significant (and worth a substantial risk). It is clear that quantitative assessments are desirable to arrive at balanced judgments as to the full consequences—both social benefits and social costs—of using or not using a particular suspect chemical. What is much less clear is how to quantify benefits when they involve health, environmental quality, and other value-laden areas.

In part because of the paucity of basic data and of experience in evaluation methods, society has tended to respond erratically to assertions of chemical threats. In the case of the 1973 restriction on the use of spray adhesives, federal response proved hasty and overreac-

*There is a qualitative difference between carcinogens and other toxins. A low dose of a kidney poison, for example, may irreversibly damage a small fraction of kidney cells, but not constitute a threat to life. A very low dose of a carcinogen, in theory, could occasion the neoplastic transformation in only one or a few cells; yet, should they escape the immunologic surveillance mechanisms and reproduce sufficiently, lethal cancer might result.

†Generalizations concerning carcinogens are fraught with risk, however. The specific change(s) in the cell genetic apparatus are unknown as is the rate of spontaneous carcinogenesis. To the extent that the alterations in DNA occasioned by mutagens are repairable by the known efficient repair enzymes of normal cells, such effects may find no biological expression. Some potential carcinogens are active only after they have been metabolically altered by the liver or other cells; other carcinogens are excreted or inactivated by the body. Hence, it seems unreasonable to consider that there should be a single form of the dose response curve, differing only in slope (potency).

tive. On the other hand, the action to curb the exposure of workers to asbestos after health effects had been demonstrated came at a rate that many observers considered unduly slow. Indications are that we are now entering an era in which there will be increasing demand for quantitative risk–benefit assessments. This balancing process could lead to a new formulation of acceptable risk. Acceptability is a social, not a scientific, criterion.

What is the best procedure for those most affected (or their representatives) to balance risks and benefits? In most current situations, a regulatory agency makes the judgments. In others, the Congress has laid down specific guidelines for acceptable risks—air-pollution laws, for example.

In some countries, tenured boards have been set up, with representations from the various constituencies affected by benefits and risks. Although this approach may be more cumbersome, it provides for open public debate on an issue and its consequences prior to the final decision.

Sociopolitical balances are mainly shaped by current value judgments, but they may also involve technical considerations. One is the development of a quantitative risk-assessment procedure. Others involve the procedures for estimates of benefits and costs of control. The development of quantitative assessments and their accurate interpretation to the people at risk will be difficult. The applicable science is still very much in its infancy, and precise estimates will rarely be possible; yet, such estimates, even crude ones, are better than none.

The public response to new scientific findings concerning chemical risks may be influenced by a variety of political and social factors not directly related to the evidence. For example, it is known that municipal referendums on adoption of water-supply fluoridation may be affected by such considerations as administrative methods and attitudes toward government.

RESEARCH RESOURCES

There are several resources available to develop the needed knowledge on environmental and health effects of chemicals: the academic community; the National Institutes of Health, especially the National Cancer Institute (NCI) and the National Institute of Environmental Health Science (NIEHS); the National Institute of Occupational Safety and Health (NIOSH); the National Center for Toxicological Research (NCTR); the Environmental Protection Agency (EPA); the Department of Energy; and, to a lesser extent, other federal agencies.

Nevertheless, a number of shortcomings remain. The greatest need lies in the area of basic research. We are only now beginning to understand the fundamental nature of cancer. Just as medieval communities were helpless in the face of bubonic plague because they did not realize that it was carried by rats and fleas, so we may be unable to check the inroads of cancer until we understand its etiology. Meanwhile, we must do the best we can. There is clearly a need for better coordination and especially better linkage between the more research-oriented units and the regulatory agencies. The new National Toxicology Program, bringing together the NCI Bioassay activity and parts of NCTR and NIEHS under the direction of the director of NIEHS, and the oversight of an executive committee that includes the heads of the regulatory agencies, could improve this linkage. A present limitation in this field is the shortage of trained toxicologists and epidemiologists.

The impacts of chemicals on nonhuman species are not receiving sufficient attention at the national level. As a result, there are difficulties in making judgments on the risks many chemicals pose for them.

A field that has been seriously neglected, but which now has been recognized by both EPA and the Congress, is research to anticipate future problems. It is desirable to anticipate potential problems from chemicals as early as possible, perhaps at the early industrial development stage. An early-warning research program could establish a basis for developing low-risk alternatives or needed control procedures; such measures would be superior to the present tendency to wait for problems to emerge as full-blown crises. Such anticipatory research must, of course, be based on carefully reasoned and highly effective technological forecasting.

STANDARDIZATION OF TESTS

Both administrators and manufacturers understandably seek as much simplification and standardization of safety testing as possible. While some standardization is desirable (such as minimum standards of good laboratory practices), there is a danger in carrying it too far too soon. Premature freezing of techniques can stifle the development of what is still an infant science. It would be unfortunate to lock in procedures that may not be adequately informative or efficient. In a rapidly evolving field such as this, a best-available-procedures approach to safety evaluation may be best.

INTERNATIONAL COLLABORATION

Ongoing research in the United States needs to be closely linked with parallel activities in other nations, not only because of what can be learned from their experience, but also because of the influence the United States has on standards elsewhere. To these ends there is also a need for mechanisms to insure that international resources for such evaluations are used effectively, without unnecessary duplication or overlap, and that they take into account differences in physical and cultural environments.

The World Health Organization has authorized a major expansion of its program for evaluating chemical safety for humans through a greater (and to some degree decentralized) reliance on cooperative centers throughout the world. Similar resources for evaluating environmental effects of chemicals are needed but are not now available.

OUTLOOK

The following outlook section on toxic substances in the environment is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

Some toxic chemical hazards are well known; others remain poorly defined. Gaps in our knowledge of causes and effects are serious, and the need to fill them is urgent. More knowledge will help to reduce or eliminate the hazards from toxic agents while permitting us to enjoy the benefits of the great majority of chemicals without fear of harm to human health or the environment.

Manufactured chemicals have become ubiquitous. Their production is steadily rising and their uses multiplying. Over the next five years, a great number of new chemicals will be introduced—perhaps as many as 1,000 annually.

But we are still struggling with problems from the past. Chemical wastes were often dumped haphazardly in disposal sites, and some of these sites (e.g., the Love Canal near Niagara Falls, New York) were later used for landfill. Estimates are that hundreds of these abandoned chemical dumps exist, most still undiscovered. Many will be found in the coming years, and the government will have to establish procedures for dealing with them. Thus far, the human health hazards posed by these discarded and refilled dumps have not been fully evaluated.

Highly persistent chemicals, such as PCB's, are continuing to build up in the sediments of our freshwater lakes and streams. These chemicals move up the food chain from microorganisms and other creatures to fish and finally to humans. Through this biomagnification, such chemicals are further accumulating in humans and nonhuman species. No effective means have yet been developed to sequester, degrade, contain, or remove chemically laden sediments. Moreover, there are fears that some of the proposed solutions to the dilemma may actually worsen the situation.

A rising proportion of the nation's electrical power will probably be generated by coal. Although power-plant pollution controls will be maintained, the total environmental burden of such air contaminants as acid particulates may well increase. While the general dimensions of the health impacts of these and other pollutants are being studied, detailed knowledge on the impacts of actual concentrations of specific pollutants on specific segments of the population is still lacking.

Rapid advances in analytical techniques are making it possible to detect smaller and smaller amounts of toxic chemicals—as small as parts per billion. In consequence, trace chemicals will be discovered in previously unsuspected places. Already trace amounts of organic compounds, some identified as weak carcinogens or mutagens in experimental animals, have been found in drinking water. Currently there is no fully acceptable way of evaluating health risks (if any) from drinking water containing traces of toxic chemicals. A related concern is that the chlorination of drinking water may, through interaction with otherwise harmless organic substances, produce potential carcinogens.

A great deal more attention is being focused on workplace exposures to chemicals following the passage of the Occupational Safety and Health Act. Some 20 chemicals or processes have already been linked to cancer in exposed workers. More such associations, involving mutagens as well as carcinogens, may be found in the next few years as new evidence of long-term exposure effects is discovered.

IMPLICATIONS

Known and potential problems stemming from toxic chemicals raise a range of implications, some general and others specific. Underlying a number of them is the challenge to the threshold concept, which holds that certain low levels of exposures to chemicals are totally without effect. The threshold theory has been intensely debated over the past few years, particularly with regard to cancer. At this time, it is fair to say that although such thresholds may exist for some toxic chemicals, none has yet been reliably demonstrated for any carcinogen.

As a result of the growing conviction that there is no completely safe level of exposure to any carcinogen, we are entering an era of quantitative risk-benefit assessment. If complete safety is unattainable, we must be able to evaluate the benefits of individual chemicals: Are they vitally significant (and worth a significant risk) or trivial (and not worth even a minor risk)?

The techniques of quantitative risk assessment are still in their infancy. Over the next five years, there will be increasing efforts to evolve better methods of extrapolating the results of tests on laboratory animals to humans—in particular extrapolating from high doses given to test animals to long-term low doses to which humans are generally exposed—and to gain better understanding of the biological processes involved. The goal is to move, as rapidly as possible, to quantitative descriptions of dose-response curves at low-dose levels.

Even more elusive than quantitative risk assessment is the quantitative evaluation of benefits. There will be efforts in the near future to develop more sophisticated means of both presenting and interpreting quantitative risk-benefit evidence so it can be understood by affected groups and responsible government officials.

This approach to safety evaluation will increasingly be applied to new chemicals, chemicals in the workplace, chemicals in the environment, and the reassessment of existing household chemicals. There will be no single definition of acceptable risk. Each decision will depend on the uses of a particular chemical, its potential substitutes, the risk-benefit ratio, and yet other factors—all within a context of prevailing social values. Such decisions will be as important in determining control strategies as in arriving at acceptable risks.

Continuing efforts will be required to develop new technology for effectively sequestering, degrading, containing, or removing chemically laden wastes in freshwater lakes and rivers as well as in abandoned dumps. The challenge is to isolate or neutralize such wastes.

More knowledge will be gained of the specific effects and components of air pollutants, such as acid particulates, on specific segments of the population. The results could make possible the design of more efficient and selective controls.

Epidemiological studies will further refine knowledge of the effects of chemical exposures in the workplace. There will be more laboratory evaluations of those chemicals now in use and new chemicals being introduced. There will also be increased efforts to uncover presently unrecognized chemical hazards in the workplace. New and better control methods to protect workers from dangerous chemicals will be stressed, including education in their handling. What is further needed is an evaluation of the prevalence and seriousness of various occupational diseases as a basis for setting control priorities.

The cancer risk involved in chlorination of drinking water will be more thoroughly assessed and, if necessary, alternative methods of controlling waterborne pathogens investigated. Also, better methods will be sought to evaluate the hazards of trace chemicals in drinking water.

The growing awareness of potential chemical hazards is a positive development. During the next five years, we will come to know more about the nature, extent, and seriousness of the problems posed by toxic chemicals. Indeed, new problems may be uncovered. But as more knowledge is gained, we will learn how to better control and reduce toxic chemical hazards.

Increased training now getting underway will add significantly to the ranks of qualified toxicologists and epidemiologists within the next five years but will only begin to meet major needs toward the end of that period.

REFERENCES

1. *Selenium* (NRC Committee on Medical and Biologic Effects of Environmental Pollutants). Washington, D.C.: National Academy of Sciences, 1976.

2. Fenical, W. Halogenation in the Rhodophyta. *Journal of Phycology* 11(3):245–259, 1975.

3. Deijer, J., and A. Jernelov. *Environmental Health Perspectives* 25:43–45, 1978.

4. Report of Workshop Panel to Select Organic Compounds Hazardous to the Environment (National Science Foundation). Springfield, Va.: Natic Technical Information Service, No. PB-287-996, 1975. Resear.. Program on Hazard Priority Ranking of Manufactured Chemicals: Phase II Chemical Report (prepared for the National Science Foundation by the Stanford Research Institute). Springfield, Va.: National Technical Information Service, Nos. PB-263-161 (Chemical Nos. 1-20); PB-263-162 (Chemical Nos. 21-40); PB-263-163 (Chemical Nos. 41-60); PB-263-164 (Chemical Nos. 61-79); PB-263-165 (Appendix and References), 1975.

5. *Annual Reports on U.S. Production and Sales of Synthetic Organic Chemicals, 1917–1976*. Washington, D.C.: U.S. International Trade Commission.

6. *Implementing the Toxic Substances Control Act: Where We Stand*. Washington, D.C.: EPA Office of Toxic Substances, 1978.

7. Metcalf, R.L. Biological Fate and Transformation of Pollutants in Water. *Chemical and Biological Fate of Pollutants in the Environment*, Part 2, I.H. Suffet, ed. New York: John Wiley & Sons, 1975.

8. *Principles for Evaluating Chemicals in the Environment*. (NRC Committee for the Working Conference on Principles of Protocols for Evaluating Chemicals in the Environment). Washington, D.C.: National Academy of Sciences, 1975.

9. *Polychlorinated Biphenyls and Terphenyls*(U.N. Environmental Program and WHO). Geneva: World Health Organization, 1976, p. 65.

10. *Environmental Resources* 4:1–69, 1971.

11. *Ozone and Other Photochemical Oxidants* (NRC Committee on Medical and Biologic Effects of Environmental Pollutants). Washington, D.C.: National Academy of Sciences, 1977.

12. MacLeod, C.M. A Report of the Panel on Herbicides of the President's Science Advisory Committee. Washington, D.C.: Office of Science and Technology, 1971.

13. Firestone, D. *Ecological Bulletin* (Stockholm) 27 (in press).

14. Huff, J.E., and J.S. Wassom. Chlorinated Dibenzodioxins and Dibenzofurans, A Bibliography. *Environmental Health Perspectives*, Experimental Issue 5, September 1973.

15. Axelrod, David, personal communication.

16. *Love Canal: Public Health Time Bomb*. Albany: New York State Department of Health, 1978.

17. *Air Quality and Automobile Emission Control: Summary Rept.* (NRC Coordinating Committee on Air Quality Standards), Vol. 1. 1974, p. 11.

18. *Federal Register*, 43:2229–2240, 1978.

19. *Airborne Particles* (NRC Committee on Medical and Biologic Effects of Environmental Pollutants), 1977.

20. *Sulfur Oxides* (NRC Committee on Sulphur Oxides). National Academy of Sciences, 1978.

21. Cederlof, R., *et al.* Air Pollution and Cancer: Risk Assessment Methodology and Epidemiological Evidence. *Environmental Health Perspectives* 22:1–13, February 1978.

22. *Drinking Water and Health* (NRC Committee on Safe Drinking Water). Washington, D.C.: National Academy of Sciences, 1977, pp. 439 and 715.

23. *Epidemiological Studies of Cancer Frequency and Certain Organic Constituents of Drinking Water—A Review of Recent Literature Published and Unpublished* (NRC Safe Drinking Water Committee). Washington, D.C.: National Academy of Sciences, 1978.

24. *Federal Register* 43:5756–5780, 1978.

25. *Annals of the New York Academy of Sciences*, Vol. 298, 1977.

26. Cole, P., and M.B. Goldman. Environmental Factors, Occupation. *Persons at High Risk of Cancer: An Approach to Cancer Etiology and Control*, Joseph F. Fraumeni, Jr., ed. New York: Academic Press, 1975, p. 167.

27. *Estimates of the Fraction of Cancer in the U.S. Related to Occupational Factors*. Washington, D.C.: DHEW, September 15, 1978.

28. Whorton, D., *et al.* Infertility in Male Pesticide Workers. *Lancet* 2:1259–1261, 1977.

29. Glass, R.I., *et al.* Sperm Count Depression in Pesticide Applicators Exposed to Dibromochloropropane. *American Journal of Epidemiology*, Vol. 9, 1979 (in press).

30. Corbett, T.H., *et al.* Birth Defects Among Children of Nurse Anesthetists. *Anesthesiology* 41:341–344, 1974.

31. Allen, N. Chemical Neurotoxins in Industry and the Environment. In: *The Nervous System*. New York: Raven Press, 1975, pp. 235–248.

32. Blum, A., *et al.* Children Absorb Tris-BP Flame Retardant from Sleepwear: Urine Contains the Mutagenic Metabolite, 2,3-Dibromopropanol. *Science* 201(4360):1020–1023, 1978.

33. Linsell, C.A., and F.G. Peers. Aflatoxin and Liver Cell Cancer. Symposium on Liver Carcinoma. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 71(6), 471–473, 1977.

34. *Saccharin: Technical Assessment of Risks and Benefits*, Part 1 (IOM and NRC Committee for a Study on Saccharin and Food Safety Policy). Washington, D.C.: National Academy of Sciences, 1978, p. 16. Also, *Food Safety Policy: Scientific and Societal Considerations* (IOM and NRC Committee for a Study on Saccharin and Food Safety Policy). Washington, D.C.: National Academy of Sciences, 1979.

35. *Human Health and the Environment—Some Research Needs—2* (Second Task Force for Research Planning in Environmental Health Science) (DHEW Pub. No. [NIH] 77-1277). Washington, D.C.: U.S. Government Printing Office, 1977.

36. *Asbestos* (IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man, Vol. 14). Lyon, France: International Agency for Research on Cancer, 1977, p. 102

37. *Arsenic* (NRC Committee on Medical and Biological Effects of Environmental Pollutants). Washington, D.C.: National Academy of Sciences, 1976.

38. Eisenbud, M. *The Health Implications of Methyl Mercury in Adirondacks Lakes*. Albany: New York State Department of Health, 1978.

39. Friberg, L., *et al. Cadmium in the Environment*. Cleveland, Ohio: CRC Press, 1974.

40. Elinder, C., *et al.* Cadmium in Kidney Cortex, Livers, and Pancreas from Swedish Autopsies. *Archives of Environmental Health* 31(6):292–302, 1976.

41. *Air Quality Criteria for Lead* (EPA Pub. No. 600 8-77-017). Washington, D.C.: Office of Research and Development, 1977, pp. 1-9.

42. Savage, E.P. *National Study to Determine Levels of Chlorinated Hydrocarbon Insecticides in Human Milk*, 1975–76, and supplementary report, 1977. Springfield, Va.: National Technical Information Service, No. PB-284-393.

43. Allen, J.R., and D.A. Barsotti. The Effects of Transplacental and Mammary Movement of PCBs in Infant Rhesus Monkeys. *Toxicology* 6:331–340, 1976.

44. Proceedings of a Workshop on Scientific Aspects of Polybrominated Biphenyls, Oct. 24–25, 1977 at Michigan State Univ., *Environmental Health Perspectives*, Vol. 23, April 1978.

45. International Agency for Research on Cancer. *The Evaluation of the Carcinogenic Risk of Chemicals to Humans: Some Organochlorine Pesticides, 5, 1974. Some Halogenated Hydrocarbons* (in press).

46. Resoconto, D. *Riunione Di Esperti Sui Problemi Determinati Dall-Inquinamento da Diossina*, Proceedings of the Expert Meeting on the Problems Raised by TCDD Pollution. Milan: Commission on the European Communities, 30 September–1 October, 1976.

47. Chiazze, L., D. Levin, and T. Silverman. Recent Changes in Estimated Cancer Mortality. *Incidence of Cancer in Humans.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratories, 1977, pp. 33–44.

48. Levin, D.L., *et al.* In: *Cancer Rates and Risks;* 2nd ed. Bethesda, Md.: National Institutes of Health, 1974.

49. Higginson, J., and C.S. Muir. The Role of Epidemiology in Elucidating the Importance of Environmental Factors in Human Cancer. *Cancer Detection and Prevention* 1:79–105, 1976.

50. EPA Proposed Rules for Pesticide Programs: "Mutagenicity Testing." *Federal Register* 43:37388–37394, 1978.

51. *Mercury* (U.N. Environment Program and WHO). Geneva: World Health Organization, 1976, p. 60.

52. Hutchinson, G. E. *A Treatise on Limnology*, 1. Chemistry of Lakes. New York: John Wiley & Sons, 1957, p. 386.

53. Likens, Eugene, personal communication.

54. Nisbet, Ian, personal communication.

IV INSTITUTIONS

10 Academic Science and Graduate Education

INTRODUCTION

Preceding chapters of this report have presented some of the more recent accomplishments of American basic and applied science. Much of this work has been carried out in U.S. universities by their faculties and their graduate and postdoctoral students. With few exceptions, the individuals responsible for these advances have been educated in the public and private universities and colleges distributed throughout the nation.

Thus the universities are a major force in American science. Today, their faculties, students, and support staff are responsible for more than half the basic scientific research performed in this country. Their graduates in science and engineering populate private and government laboratories and engineering centers throughout the United States.

The conduct of basic research in U.S. universities historically has been closely linked to the graduate education of scientists and engineers. One feeds the other, in contrast to the pattern in Europe and the Soviet Union, where centralized research institutes are essentially separated from education systems. Basic science in the United States is dependent upon the coupling of research and education at the postgraduate level, so that the vitality of American science is a function of the vitality of academic institutions, particularly of the major U.S. universities that carry out most academic science and award most of the advanced degrees in the sciences and engineering.

Although federal support lessened in the early 1970's, the strong national commitment to fundamental scientific research continues—perhaps best symbolized by the provision of $3 billion federal in 1978 to fund research and graduate education in U.S. universities. As a result of this strong support, particularly during the growth years of the 1950's and 1960's, the United States has developed a broad, diversified, and highly competent academic research and related postgraduate capability. The universities are now and will continue to be remarkably productive both in increasing our understanding of nature and man and in educating new generations of scientists and engineers.

Nevertheless, the institutions of academic R&D have been facing serious problems that threaten their productivity and may well slow their present momentum. The earlier rate of growth in financial support for academic R&D has slowed to relatively level funding in constant dollars over the past decade. Averaged annual growth (in constant 1972 dollars) from 1953 to 1960 was 12 percent; from 1960 to 1964, 14 percent; from 1968 to 1974, essentially zero percent; and from

1974 to 1978, about 4 percent.[10] Yet over this period the academic research population has continued to grow, increasing the demand for research support. As research becomes more sophisticated, the required instrumentation, facilities, and supporting services become more expensive, adding further costs beyond those due to inflation. Federal agencies can now fund a smaller fraction of worthy research proposals; and other demands on university resources have sharply increased, causing some private institutions to draw upon their basic endowment capital.

The increased competition for research funding has led to longer, more detailed proposals, prepared and reviewed by scientists on time formerly spent on research. Funding agencies, in an effort to meet increased demand, have tended to reduce both the amounts and the time period of support. Funding available for direct support of research has also been eroded by rising indirect cost rates resulting in part from the need to satisfy regulations of federal and local agencies. These requirements often make serious demands on the time of the research investigator as well as on university administrators.

Given these problems, it is appropriate to ask whether academic science in the United States can maintain its high rank. There is concern that academic science is on the verge of a decline, and that it may be a decade before its decreased vitality and momentum are fully apparent, by which time the direction may be very difficult and expensive to reverse.

The period immediately ahead will require adjustments in the policies of universities and in their relationships with the federal government as the principal external sponsor of academic research. Some of these problems are now being addressed by *ad hoc* groups, notably the Sloan Commission on Government and Higher Education and the National Commission on Research. The general resolution of the problem, however, will require the thoughtful attention and best effort of all partners—universities, government, and industry—if the U.S. research enterprise is not to suffer.

SCIENTIFIC RESEARCH AND GRADUATE EDUCATION

American scientists and engineers begin their professional education as college undergraduates. Through introductory and intermediate courses in mathematics and the various natural sciences, they learn scientific principles and gain an overall view of the history and subject matter of the various disciplines. Inadequate education at this level is difficult for students to overcome later.

Graduate education in the sciences and engineering consists of advanced instruction in the current body of knowledge, methods of thought, and research techniques, all of which must be assimilated by those gaining expertise in these professions. However, in addition to receiving highly specialized instruction, graduate students participate in and contribute to research, under the guidance of senior faculty members, often as paid assistants. This firsthand experience in the laboratory starts them on the path to maturity as independent investigators. Formal graduate education concludes with the writing of a dissertation based on the student's research and intended to be an original contribution to knowledge.

Many of those who receive doctoral degrees in the sciences subsequently accept postdoctoral research appointments at academic institutions. Those who hold these positions are given the opportunity to engage in one or two years of advanced research, often of their own choosing and design, at leading scientific centers having specialized facilities and faculties. Postdoctoral appointments, which usually place few other responsibilities on their holders, give younger scientists the variety and intensity of training they need to become fully qualified investigators.

Active research, then, permeates academic science and is interwoven with the education of both undergraduates and graduate students. This interrelated arrangement not only gives students a better understanding of the current status of their science but also provides members of the faculty and their graduate assistants with the intellectual stimulus that strengthens and enlivens teaching.

ENROLLMENT, DEGREES, AND JOBS

THE 1960'S

The 1960's were a period of expansion for higher education, including academic science, in the United States. Undergraduate enrollment increased by an average 7 percent a year, which meant that colleges and universities everywhere needed more teachers, administrators, buildings, and facilities. During this period, the federal government substantially increased its funding for scientific investigations sponsored by federal agencies and for graduate fellowships and traineeships, particularly in the physical and biomedical sciences.

The natural consequence of these trends was a substantial increase in the number of graduate students and in the number of persons gaining Ph.D. or Sc.D. degrees. Between 1965 and 1971, the awarding of doctorates in all fields increased at an annual rate of 11.7 percent; the number of those with doctorates in science or engineering rose at a rate of 10.3 percent.[1]

By 1971, the nation's universities were awarding more than 30,000 doctorates annually, both scientific and nonscientific, and many of those earning doctorates were able to secure academic jobs.

THE 1970'S

However, the early 1970's marked the beginning of a difficult period of readjustment for higher education, particularly in the natural sciences. Large annual increases in earlier years of funds for scientific research by academic investigators were replaced by more modest increments or by actual declines measured in dollars of constant purchasing power. The number of doctorates awarded annually in the physical sciences, engineering, and the life sciences began to fall, as did graduate enrollments in the physical sciences, mathematics, and engineering. After reaching a peak of 33,755 in 1973, awards of doctoral degrees in all fields, both scientific and nonscientific, started to decline. As shown in Table 8, by 1977, doctoral production in the sciences and engineering had dropped from the 1971 figure of 14,311 to 11,777. Undergraduate enrollments, however, continued to increase during the decade, but at an average rate of only 4 percent.

General financial pressures within the universities to maintain their expanded establishments, combined with a slower rate of growth in undergraduate enrollment, also reduced opportunities for those seeking academic employment. Between 1969 and 1977 the proportion of new doctorate recipients in engineering, mathematics, and the physical sciences who secured academic positions dropped from 31 to 25 percent (see Table 9).

TABLE 8 Doctorates Awarded in the United States

	Totals		Sciences and Engineering[a]	
Year	Number	Index	Number[b]	Index
1960	9,732	100	4,674[b]	100
1965	16,340	168	8,307[b]	178
1970	29,500	303	13,637	292
1971	31,872	327	14,311	307
1972	33,044	340	13,984	299
1973	33,755	347	13,674	293
1974	33,046	340	12,950	277
1975	32,948	339	12,763	273
1976	32,936	338	12,222	261
1977	31,672	325	11,777	252

[a] Engineering, mathematics, physical sciences, and life sciences.
[b] From unpublished NRC tables.
SOURCES: National Research Council. *Summary Reports, 1977, Doctorate Recipients from U.S. Universities*, Washington.

TABLE 9 Employment Plans of New Doctorate Recipients (in percentages)

	Engineering Mathematics, and Physical Sciences		Nonsciences	
Fields of Employment	1969[a]	1977	1969[a]	1977
Postdoctoral study	20	28	2	4
Academic	31	25	80	70
Industry, government, and nonprofit	42	38	7	15
Other and unknown	6	9	11	11
TOTALS	100	100	100	100

[a] From unpublished NRC tables.
SOURCE: National Research Council. *Summary Report, 1977, Doctorate Recipients from U.S. Universities*. Washington.

THE PRESENT AND THE FUTURE

Overall undergraduate enrollment in the nation's colleges and universities is expected to begin a decline within the next two or three years. The reason is that the number of births in the United States has been declining since the end of the baby boom in 1964. As a result, we can expect a decline in the number of high school graduates seeking to enroll in universities during the early 1980's, returning to a situation that existed in the late 1950's.

This decline, however, will be more gradual than might be predicted on demographic grounds. There has been a significant increase in the number of adults—particularly women and minority-group members—attending college. More than half of all college students in 1977 were older than 21,[2] which was once the age at which the great majority of students received a first degree. Many of these older students, however, are found at community colleges and vocational institutions rather than at liberal arts colleges or universities.

A decline in the number of undergraduates, whatever its size, will probably not have a corresponding effect on the number of graduate students. Graduate enrollment in both the sciences and other fields is expected to continue to rise as more and more college graduates seek advanced degrees to bolster their chances on the job market. Many of these will be students who enter doctoral programs, only to drop out before they obtain their degree. Furthermore, foreign students comprise a substantial percentage of those enrolled in U.S. graduate schools, and that percentage is expected to continue growing for the next several years.[3]

Data for 1977–78 on first-year enrollments of graduate students suggest that earlier trends in individual disciplines may now be changing. First-year enrollments of graduate students in the physical sciences and engineering are up slightly, while first-year enrollments in the life sciences are slightly down.

Despite increased numbers of graduate students, total doctoral production—down by 25 percent in the physical sciences and engineering between 1971 and 1977[4]—seems likely to continue to decrease.

A smaller percentage of the new doctorate holders will obtain academic jobs than at any time in the recent past, partly because the number of undergraduates will decline and partly because the retirement rate of professors will be low. As of 1976, over half of those on college and university faculties had been hired during the preceding 15 years. Unless there is a marked movement among current faculty members to early retirement—an event that seems unlikely in a period of inflation and the recently legislated minimum mandatory retirement age of 70—the result will be relatively few openings for new faculty members until the 1990's.

Furthermore, as indicated earlier, a substantially larger proportion of individuals with recent doctorates in the sciences are seeking and obtaining jobs outside the universities than was true 10 years ago. This indicates that younger scientists are adjusting to the realities of the job market and the situation of the universities. Their ability to make this adjustment, however, is complicated by differences in demand. The demand for solid-state physicists, for example, has remained much higher than the demand for nuclear physicists. The demand for analytical chemists has stayed at a high level, even though the market for chemists generally has been relatively unfavorable. Predicting differences in future job markets is difficult, and it would seem to be increasingly desirable to search for modes of graduate and postdoctoral education that enable students in the sciences to adapt more easily to changes in demand.

MAINTAINING FACULTY BALANCE

The prospect of a period during which few younger scientists receive academic appointments brings with it the possibility of some impairment of the capacity of the universities to carry out basic research and sustain the vitality of their faculties. Younger faculty, as a group, bring to their work a high degree of enthusiasm, singleness of purpose, and freedom from administrative duties. As recent graduates, they also benefit from the currency of their knowledge and perspectives.

The existence of organized research units of the kind described in a subsequent section will enable universities to absorb some scientists and engineers without teaching responsibilities. The possibility that universities also will be engaged more extensively in applied research may provide additional employment opportunities for younger scientists and engineers. However, these positions are generally divorced from the central educational role of the universities.

Although the general outlook for faculty positions over the next five years is predictable, it will be important during this period to monitor and analyze such factors as the number of academic jobs available for younger faculty, trends in academic and nonacademic salaries, shifts of faculty out of academic life into other positions, retirement rates (including early retirement), and changes in tenure ratios, rules, and customs. Such studies should be helpful in assessing the need for remedial measures necessary to assure competent and balanced science faculties.

SUPPORT OF GRADUATE STUDENTS

During the past few years there has been a substantial shift in the sources of financial support for graduate students (see Table 10). For many years, most of the federal funds used to provide fellowships and traineeships for graduate students came from the G.I. bill. In the peak year of 1975, $370 million was provided for graduate students under this legislation,[5] but this funding will fall to an estimated $207 million in 1979. Although substantially lesser in amount, the National Institutes of Health (NIH) training program in the biomedical sciences ($130 million in 1977)[15] and the National Science Foundation (NSF) graduate fellowship program ($11 million in 1978) have been especially helpful in providing assistance to graduate students in the sciences. However, these programs have declined in some areas or remained at the same level. The proportion of graduate students whose major source of support was the federal government has declined, from 40 percent in 1968 to 23 percent in 1977. Most of the difference has been made up by students themselves (by working, by the earnings of

TABLE 10 Sources of Support for Full-Time Graduate Students in Sciences and Engineering in Doctorate-Granting Institutions (in percentages)

Sources of Major Support	1968	1971	1977
Federal	40	31	23
Institutional	32	37	37
Self	20	22	32
Other	8	10	8

SOURCES: National Science Foundation *Graduate Student Support and Manpower Resources in Graduate Education, Fall 1970*, p. 73 and *Fall 1971*, p. 46; and *Graduate Science Education Student Support and Postdoctorals, Fall 1977*, p. 77. Washington.

spouses, and by borrowing from private lenders) and by the universities, which have increased their funding for fellowships. One effect of this changed pattern has been to divert larger amounts of university funds to the support of graduate students, with a resulting reduction in the university funds available for other university functions.

WOMEN AND MINORITIES

The proportion of female graduate students in the universities increased from 38 percent in 1969 to 47 percent in 1978.[6] This proportion will probably increase very slowly over the decade ahead. The majority of women at the graduate level are still found in education, the humanities, and the social sciences.

Both the number of women receiving a doctorate in science or engineering and the proportion of all doctoral recipients in science or engineering who are women increased significantly between 1968 and 1977. The number of women receiving doctorates in science and engineering rose from 1,306 to 3,292; the proportion rose from 9.6 percent to 18.0 percent (see Table 11). These numbers and proportions, however, are both heavily weighted by the relatively large numbers of women in the life sciences, biology and medicine, and in the social sciences.

The number and proportion of women on university faculties will probably remain relatively small over the next decade, even with expected increases. As noted in the previous section of this chapter, there will be relatively few academic positions available during the 1980's, except in collegiate institutions not offering graduate degrees in the physical and life sciences and engineering.

Progress in enlarging the representation of nonoriental minorities in science has been slow and irregular. In 1977, for example, the proportion of doctorates granted to minority groups was quite small compared with that of whites, as shown in Table 12.

The low representation of blacks and other [illegible]ties among recipients of doctorates can be sign[illegible]antly increased only as efforts to erase the ef[illegible] generations of discrimination become effective. [illegible] will take time. Meanwhile, efforts must continue to strengthen the educational background of minority students who need help in order to deal successfully with graduate studies.

TABLE 11 Women Securing Doctorates in Sciences and Engineering

Field	1968 Number	1968 Percent of Doctorates Awarded	1977 Number	1977 Percent of Doctorates Awarded
Physical sciences	232	5.0	431	9.9
Life sciences	510	13.8	957	20.1
Social sciences	552	15.8	1,830	28.1
Engineering	12	0.4	74	2.8
TOTALS	1,306	8.9	3,292	18.0

SOURCE: National Research Council. *Summary Report, 1977, Doctorate Recipients from U.S. Universities*, p. 8. Washington.

FUNDING OF ACADEMIC SCIENCE

Funds separately budgeted for the support of academic R&D (that is, funds designated specifically for research) come from a variety of sources: the general funds of universities, the federal government, state government, private foundations, industry, and individual donors. Virtually all the federal government's support comes to the universities as grant and contract funds and, in 1977, constituted two-thirds of the separately budgeted funding. The universities themselves, the next most important source of funds for separately budgeted research, provided 22 percent. Private foundations and industry provided 8 percent and 3 percent, respectively.[7]

University contributions from general funds that are not separately budgeted are less clearly identifiable but they are large. Universities provide faculty and support staff salaries and benefits, as well as the general facilities and environment that make universities a particularly produ[illegible] for basic science.

In the case of [illegible] universities, these general funds are derived primarily from funds that the universities receive from state governments. In the case of private universities, the funds are taken primarily from tuition revenues, endowment income, and private gifts. The data presented in this section reflect only separately budgeted funds and therefore understate the actual level of university financial support for academic science.

LEVELS OF SUPPORT

Total support for academic R&D (that is, basic research, applied research, and development) has risen steadily since World War II in real terms except for the period 1968–74—when there was no increase in constant dollars, and the federal contribution actually declined by 8 percent, causing universities to increase their contribution in compensation. During the period 1974–79 there has been a 17 percent increase in support, which is helping to offset the adverse effects resulting from the years of level funding (see Table 13).

TABLE 12 Doctorate Recipient by Racial/Ethnic Group and Fields of Science (fiscal year 1977)

Racial-Ethnic Group	All Fields No.	All Fields %	Physical Sciences No.	Physical Sciences %	Engineering No.	Engineering %	Life Sciences No.	Life Sciences %
American Indian	215	1	16	[b]	12	1	37	1
Asian	907	3	223	6	248	14	182	4
Black	1,186	4	44	1	15	1	68	2
Hispanic	471	2	56	2	22	1	35	1
White	23,411	86	3,051	85	1,412	79	3,506	88
Other and unknown	1,181	4	210	6	84	5	158	4
TOTAL[a]	27,371		3,600		1,793		3,986	

[a] Include U.S. citizens and non-U.S. with permanent visas.
[b] Less than 0.5 percent.
SOURCE: National Research Council. *Summary Report, 1977. Doctorate Recipients from U.S. Universities*, p. 16. Washington

The funds over the last decade have had to be shared among a growing group of claimants. The number of scientists and engineers with doctorates employed by colleges and universities engaged primarily in research and development increased by 50 percent (from 33,600 to 50,400) from 1961 to 1969 as undergraduate enrollment and federal funds for research and development rose.[8] This group has now matured, and, as principal investigators, many now seek support. The combined effect of a slow increase in available funds and a rapid increase in claimants has been a general decline in the fraction of proposals funded by federal agencies. For example, in 1967, 82 percent of the proposals found worthy of support by the NIH were funded. By 1977, this percentage had fallen to 48.[9]

Funding for Basic Research

Basic research comprises about 68 percent of all research conducted in universities. In constant dollars, total support for basic academic research declined by 8 percent between 1968 and 1976, and federal support for basic academic science fell by 10 percent in constant dollars over that period.[10] To offset the cut in federal funds, universities increased their support of separately budgeted research from $334 million to $360 million in constant 1972 dollars, or 8 percent, over the 1968–76 period.[11] This increase in expenditures strained the resources of many of the universities.

Virtually level expenditures for all academic science, combined with the increased number of eligible investigators, had significant effects over this 1968–76 period, in addition to a reduction in the fraction of proposals accepted. As federal agencies sought to stretch their funds, budgets proposed by academic scientists were often cut, and a smaller proportion of proposals, as noted above, were funded for shorter time periods. Increasingly stringent economies became necessary in such things as shop and technical services, equipment purchases, graphics, and travel.

TABLE 13 Recent Trends in Expenditures for Academic Research and Development (constant dollars in 1972 dollars)

	All Sources				Federal Government			
Year	Current Dollars	Percent Increase Over 1974	Constant Dollars	Percent Increase Over 1974	Current Dollars	Percent Increase Over 1974	Constant Dollars	Percent Increase Over 1974
1974	3,023		2,606		2,032		1,751	
1975	3,409	13	2,681	3	2,288	12	1,799	3
1976	3,730	23	2,789	7	2,501	23	1,870	7
1977	4,064	34	2,870	10	2,717	34	1,919	10
1978	4,585	52	3,018	16	3,075	51	2,024	16
1979	4,965	64	3,055	17	3,315	63	2,040	17

SOURCE: National Science Foundation. *National Patterns of R&D Resources: Funds and Personnel in the United States, 1953–1978/79*, pp. 29, 37. Washington.

Meanwhile, the rising competition for funds forced many researchers to devote much of their time to preparing lengthier and more detailed applications for research grants and contracts.

In recognition of the need, the federal government has recently raised its level of support. Between 1974 and 1977, federal funds for basic scientific research were increased 10 percent, in constant dollars. Acceptance by Congress of the President's proposed funding for basic scientific research in the universities during fiscal year 1980 would result in a 15 percent total increase, in constant dollars, between 1976 and 1980.

EFFECT OF CHANGES IN NATIONAL PRIORITIES

Changes in national priorities lead to changes in the research budgets of federal agencies. Between 1968 and 1978, for example, the portion of federal research funds accounted for by the Department of Health, Education, and Welfare (primarily NIH), NSF, and other agencies rose from 50 percent to 64 percent, while the portion accounted for by the Department of Defense, the Department of Energy, and the National Aeronautics and Space Administration (NASA) dropped from 50 percent to 33 percent.[16]

Such shifts affect both the universities and the agencies. Some individual researchers must change the direction of their research or seek new sponsors; opportunities for graduate students decline in some fields and rise in others; new research groups and research structures (such as centers and institutes) are formed, while older ones may be disbanded. Mission-oriented agencies seeking solutions for real and pressing problems tend to emphasize applied research and development at the expense of their support of basic research. NSF, which has sole responsibility among federal agencies for assuring adequate balance and general support of basic science, is faced with many demands as a result of decisions of other agencies. If these shifts continue, as is likely, it is possible that the coming years will see an increase in the proportion of university-based applied science and specialized institutes and nonteaching research centers.

CAPITAL EXPENDITURES

When funds are relatively plentiful (as they were in the 19[illegible]'s), capital investments tend to rise. When funds are relatively scarce (as they have been in the 1970's), the tendency is to pay operating costs first and use what may be left over for capital investment. This is what accounts for the reduced federal investment in capital facilities for basic scientific research. Academic research in the natural sciences requires capital investment in buildings and large research installations. Laboratory equipment and special instrumentation are also capital costs, but their useful life is typically shorter than that of buildings.

Federal obligations for academic research facilities dropped from $212 million in 1966 to $28 million in 1975, in constant 1972 dollars.[12] Federal funds available to the universities for purchasing laboratory equipment also declined, although not as greatly. The proportion of funds allocated to NIH for laboratory equipment, for example, declined from 11.7 percent in 1966 to 5.7 percent in 1974.[13]

The federal government's decision to reduce allocations for capital investment and equipment was a rational response to budgetary problems. However, the period of low investment was so protracted that many research installations have become, or are becoming, obsolescent. Some experiments simply cannot be performed in existing facilities with an earlier generation of equipment. In short, the academic research system is consuming its capital, and the grace period during which the system could operate effectively on earlier capital investments is running out.

The need to lift the level of investment in research facilities and equipment has been recognized by the Administration and by Congress. Special funds—$93 million in 1978, $88 million in 1979, were provided, and $101 million is proposed by NSF for new capital investment and equipment in its 1980 appropriation request. Several years of funding at this general level will be required to refurbish the facilities and instruments necessary in modern academic science.

SUPPORT STRUCTURES

Effective research in scientific laboratories also requires an array of support services—shops, libraries, equipment, technicians. Some of the most urgent problems of the universities over the next five years will arise in funding these services.

PROJECT FUNDING

Although federal agencies support some broad areas of investigation through large grants or contracts that finance the work of many individuals and groups, the dominant mode is support of individual projects—that is, support for a defined research task spelled out in a proposal made by an investigator to a federal agency. This method of operation has many fundamental strengths:

- Recognition and encouragement of initiative on the part of individual investigators.
- Provision of an efficient mechanism for matching the legitimate needs of mission-oriented agencies for

research in given areas with the legitimate desire of scientists to work on problems that interest them.

• Large-scale participation by good scientists in the decision process and assurance of quality through peer review of individual projects, while retaining for responsible government officials large-scale program decisions.

• Establishment of a means for shifting agency program emphasis without substantial disruption of specific research projects.

• Avoidance of the necessity, in distributing federal research funds, to judge the relative merits of universities.

• Disclosure to the Congress and the public of the content of research financed from public funds.

WEAKNESSES OF THE PROJECT SYSTEM

Yet the project system as currently administered is not without flaws. One problem, noted earlier, is a decreasing length of average periods of support for projects. The system is one that has been characterized as financing 10-year ideas with 3-year grants and 1-year appropriations. There is no short-term solution to this problem. If average periods of support are lengthened while total funds remain constant, fewer new starts in each year are possible or the average size of grants has to be cut. Worthwhile lines of investigation may be unfunded; cutting back the size of grants often impairs progress. Thus, the decisions of program managers are invariably compromises.

Federal agencies must obtain enough information about proposed research projects to enable them, or peer-review groups, to judge whether the work merits support. As funds have become tighter, applicants have taken to writing more detailed and documented proposals in the hope of gaining a competitive edge; and some agencies have encouraged, or even required, needlessly detailed applications. As periods of support and the average grant become smaller, the burden of more-frequent application writing rises. Many principal investigators now spend substantial portions of their time writing proposals—a fraction that could and should be substantially reduced without decreasing the rigor of the review process. NSF recently placed a 15-page limit on grant applications in order to reduce this burden on both reviewers and investigators.

Under the project system, expenditures under a given grant must be limited to activities specified under that grant, transfer of grant funds from one category of expenditures to another is limited, and funds must be spent within a specified period. These rules, which are intended to ensure prudent expenditure, often prevent research institutions from adjusting to unforeseen problems or responding to new developments or insights.

Finally, there is the ubiquitous question of indirect costs. When federally financed research is undertaken in academic institutions, two kinds of costs are generated. The first are direct costs, such as salaries and supplies required solely for the conduct of a specific project. The second are other costs arising from the use of common services provided for a number of projects. These include the provision of heat and light, library services, and use of accounting, contract, and administrative offices. University administrators generally hold that the federal government should pay the full direct and indirect costs of the research that it supports. They also argue that highly detailed accounting on such matters as the division of effort of investigators between research, teaching, and other functions is impracticable and unnecessary. At the same time, government auditors and contract administrators point to abuses that have occurred and argue for stricter accountability of investigators' time.

The debate over indirect costs has gone on for more than 20 years, but it intensified during the recent period of relatively scarce federal funds. Much more than an agreement over the technicalities of accounting is involved. Indirect costs have increased, in large part as a result of the increasing complexity of the project system and of the costs of complying with federal regulations and accounting procedures. Satisfactory resolution of the issue requires a consensus on the respective responsibilities and rights of the sponsoring agencies and universities and on what is properly allowable as indirect costs.

ACCOUNTABILITY, MANAGEMENT, AND REGULATION

Public concerns affecting both federal spending generally and research and development expenditures specifically are leading to closer supervision and management of academic science. The sheer size of the federal investment in basic and applied academic science—about $3 billion per year—directs that responsible attention be given to oversight of expenditures and the management of research activities.

Efforts to use federal funds more efficiently have the implicit support of both government and the academic community, but perspectives differ as to how and where these efforts should be directed.

Pressures for fiscal accountability will not decrease and may well increase. Grant and contract administration, therefore, is not likely to become less complex unless the special problems of the universities can be better recognized within government procurement procedures aimed at preserving public accountability.

Distinct from the relatively small individual research grant or contract are the large research efforts carried out within the universities. These large projects, which often use big machines and the entire personnel of research institutes to attack broadly defined tasks, involve management techniques and resources quite different from those of the individual investigator. Improvements should be attainable in both areas through the joint efforts of university and agency administrators.

Finally, there is the area of social goals. In recent years, national efforts to erase ~~racial discrimination~~ have been joined by equally concerted efforts to eliminate discrimination on grounds of sex or age. Greater national attention is being paid to meeting the special needs of the physically and mentally handicapped, and to protecting the rights of those who serve as human subjects for research. Other goals include more astute regulation of pension plans and more adequate provision of health services in every community.

So far as the universities are concerned, the problems and costs of meeting these obligations and complying with federal and state laws and regulations are similar to those faced by other institutions. However, some regulations, such as those relating to laboratory safety and protection of human and animal research subjects, relate to problems that differ significantly between universities and other institutions and directly affect the cost and quality of scientific research.

Both the universities and federal agencies face complex obstacles in establishing reasonable regulations and administrative processes. Recent legislation directed toward social goals has demanded more in the way of new skills, attitudes, and resources, and more understanding by both parties, than could be generated in a short time. The consequence, too often, has been an adversarial relationship. Moreover, a number of existing regulations and their administration generate costs—in consumption of time and money—that are higher than necessary for both the federal agencies and the universities.

There needs to be a better forum in which the philosophical differences between the universities and the federal government can be considered, free from the atmosphere of adversarial negotiations and tight deadlines that have increasingly characterized discussion of these differences. A basic difficulty is that government funding of academic research is generally short term, while many of the university commitments are long term. More general is the question of whether the government is merely a purchaser of university services rather than a partner in a joint venture in research, sharing responsibility for the institutional health of the universities. If these problems and if differences are to be resolved, the next five years should see a major national effort on the part of both parties to achieve the purposes of each.

OTHER ISSUES

STRUCTURAL ADAPTATION

During the past three decades the growth in the size and complexity of scientific endeavor has required some modifications in the structure of academic science. Today, effective research often requires close collaboration among investigators from different disciplines, highly specialized equipment, and administrative arrangements that do not fall easily into the patterns of traditional university administration. As a result, the last 30 years have seen the rise of special research units—institutes, centers, laboratories, and others—within the universities.

These special units often function in a semi-autonomous way, with their own hierarchical structure, goals, and administrative arrangements. Each of the major research universities has a set of these special research units, operating within the university framework but with somewhat looser ties to the parent organization than the individual scientific departments, and differing in structure from one university to another. This adaptation to new directions in scientific exploration will continue into the indefinite future.

UNIVERSITIES AND INDUSTRY

Prior to World War II, industry provided the major portion of funding for academic science from external sources. After the war, however, industry's place as the chief outside supporter of academic science was assumed by the federal government. As of 1977, industry support for academic science amounted to \$139 million in current dollars, about 3 percent of the total support from all sources.[14] This sum, provided for both general scientific research as well as work on specific projects of special interest to the industrial firms supporting them, does not take into account the large amount of work done by academic scientists working as occasional consultants within industry to help solve specific problems. Most of this consulting work in recent years has been done by members of chemistry, physics, and engineering departments.

It is clear that the nation would benefit from a closer and more extensive relationship between the universities and the industrial world. Basic scientific research often serves as a foundation for later technological advances that can be utilized by industry;

industrial problems, in turn, often offer fertile ground for innovative scientific exploration. The need for a closer relationship between the universities and industry is now becoming more widely recognized. Administrators of federal agencies, university scientists, and corporate officials have all begun to realize the necessity for it and its potential benefits.

OUTLOOK

The following outlook section on academic science and graduate education is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

The major universities of the United States play a vital role in the progress of American science. In addition to educating young scientists, they are responsible for more than half of the nation's basic scientific research. Having had the benefit of strong national attention and a large commitment of federal funds, academic science during the past three decades has developed a momentum that will allow it to continue to make major contributions to the world's store of knowledge during the next five years and beyond. This period, however, will also be a time of adjustment for universities and academic science as they attempt to deal with such new circumstances as more stringent financial conditions, declining undergraduate enrollments, a sharp reduction in the number of new faculty positions, and the need to meet new regulatory and other administrative requirements. Adjusting to these new demands will require time and effort on the part of all those concerned with the vitality of academic science.

SCIENTIFIC RESEARCH AND GRADUATE EDUCATION

The close tie between scientific research and graduate education in the sciences can be expected to continue over the next five years. While most academic scientists will continue to perform their dual roles, teaching graduate and undergraduate students while supervising or conducting research, the number of specialized research institutes and centers may well increase.

ENROLLMENT

Total undergraduate enrollment in American colleges and universities will enter a period of decline during the next five years, chiefly because of the birthrate decline. Graduate school enrollment, however, will continue to grow, in both scientific and nonscientific fields. More college graduates will pursue master's degrees and doctorates in order to bolster their ability to find professional positions. In addition, the sizeable number of foreign students in graduate schools is expected to continue to rise.

DOCTORAL DEGREES

Somewhat paradoxically, the total number of doctorates to be awarded over the next five years will gradually decline, despite an increase in the number of those in graduate schools. As in the past, a substantial percentage of those seeking doctoral degrees will relinquish that goal. Furthermore, declining undergraduate enrollments will mean a reduction in the need for new faculty members with doctoral degrees. A substantial part of the decline in doctorates will be accounted for by a reduced number of doctoral awards in physics, chemistry, mathematics, and engineering.

JOBS

The relative unavailability of university faculty openings will cause a greater number of persons with doctoral degrees or other advanced education in the sciences to seek positions in industry, government, and nonprofit research organizations during the next five years. The small number of academic positions available during the next several years may also require special measures to assure the retention of the best young academics in science.

WOMEN AND MINORITIES

The large increase in the relative number of female graduate students and in the number of female holders of doctoral degrees that occurred in recent years will slow somewhat in the next five years. There will be slow but continued growth in the numbers of women in the sciences and engineering. The declining need for faculty members will also limit the job prospects of women with recent doctoral degrees. The outlook for minority students, particularly blacks and those of Latin American ancestry, is for continued slow improvement in the number of doctoral recipients over the next five years. Continued efforts to overcome the effects of disadvantaged background will be required.

FUNDING OF ACADEMIC SCIENCE

Funds for academic science come from a number of sources—the federal government, the universities themselves, private philanthropic foundations, and industry. The states, through their contributions to public universities, can also be considered a direct source of funding. Although federal funding for academic science has increased modestly in recent years, and may continue to do so, the universities will be under increasing pressure to increase their share of the funding, now more than 20 percent. The decline in federal support during the period 1968–74, combined with inflation and the rising number of scientific investigators within the universities, assures that the years immediately ahead will nonetheless be a time of emphasis on frugality within academic science. Universities and government will have to find ways to cope with the problems of aging buildings, obsolescent equipment, and rising support costs.

PROJECT FUNDING

During the next five years, most of the funds for academic research will continue to be distributed in the form of individual grants and contracts. Although this system has a number of strengths, attention should be given in the years immediately ahead to relieving the procedural rigidity that is now its chief weakness.

Continued attempts will also be made to resolve the differences between the universities and the federal government over the apportionment of indirect costs and over accounting procedures.

ACCOUNTABILITY, MANAGEMENT, AND REGULATION

Growing federal regulation in such areas as health, safety, financial security, and equal opportunity will also continue to affect the universities. In the recent past, the rapid proliferation of regulation has created strains between the federal government and academic sciences. During the next five years the efforts of both parties should be directed toward achieving the goals of regulation in a positive manner that enhances understanding and reduces total costs as well as the time diverted from scientific research.

OTHER ISSUES

The relationship between academic science and industry should become a closer one in the years immediately ahead.

REFERENCES

1. *Science Indicators 1976* (National Science Board, NSB-77-1). Washington, D.C.: Government Printing Office, 1977, p. 293.

2. *Current Population Reports* (Bureau of the Census). Series P-20. 1978, pp. 41 and 61.

3. *Counselor's Newsletter* (Institute of International Education) 11:1, 1978.

4. *Doctorate Recipients from U.S. Universities: Summary Report* (NRC Board on Human-Resource Data Analyses). Washington, D.C.: National Academy of Sciences, 1975 and 1977.

5. *Outlook and Opportunities for Graduate Education* (NAS National Board on Graduate Education). Washington, D.C.: National Academy of Sciences, 1975, p. 1.

6. Anderson, C. (ed.). *A Fact Book on Higher Education.* Washington, D.C.: American Council on Education, 1977, p. 77.114.

7. *National Patterns of R&D Resources, 1953–1977* (National Science Foundation, NSF 77-310). Washington, D.C.: Government Printing Office, 1977, pp. 37–38.

8. *Ibid.*, p. 34.

9. *Basic Data Relating to NIH* (National Institutes of Health). Washington, D.C.: Government Printing Office, 1977, p. 36.

10. *National Patterns of R&D Resources, 1953–1978-9* (National Science Foundation, in press). Washington, D.C., Table 12.

11. *National Patterns of R&D Resources, 1953-1977, op. cit.*, p. 30

12. *Science Indicators, 1976, op. cit.*, p. 217.

13. *Science Indicators, 1974* (National Science Board NSB 75-1). Washington, D.C.: Government Printing Office, p. 181.

14. *National Patterns of R&D Resources, 1953–1977, op. cit.*, p. 23.

15. National Research Council. *Personnel N[illegible]s for Biomedical and Behavioral Research, 1978 Report.* Washington, D.C., National Academy of Sciences, 1978.

16. *Federal Funds for Research and Development.* Vol. XXVI and XXVII (NSF 78-312). Washington, D.C., National Science Foundation, 1978.

11 Institutions for International Cooperation

INTRODUCTION

The vigor and development of science and technology are intimately related to the interaction of ideas and activities of its many participants. This process takes place in several ways: personal contact, publications, symposia and workshops, coördinated but independent research, and closely coupled collaborative research. The process may be highly informal and unstructured, or it may be carried on within formal national or international institutions.

Here we emphasize formal international institutions. It is important to also recognize, however, the network of scientists and technologists that has operated effectively as an "invisible" international institution for more than a century.

Immediately after World War II, emphasis was placed on institutional arrangements that would help to rebuild the scientific infrastructure in Western Europe and Japan. Now, however, international institutions facilitate the traditional role of the scientific and technological enterprise in developing understanding and extending knowledge, and also in responding to global societal problems that have gained prominence since World War II. Since issues such as environmental quality, energy, population, health, food, water, urban settlements, poverty, and peace have global aspects, international institutions are important in coping with them. Clearly, we are still in an experimental phase of adapting existing institutions or fashioning new ones to address such problems, but their urgency requires a rare combination of imagination, wisdom, and discrimination in deploying the intellectual resources of increasingly interdependent nations.

This chapter attempts to put into perspective the rationale for international science at both the academic and intergovernmental levels and to do so in terms of the short-range prospects for international institutions of science and technology. Rather than attempting a catalog, we have concentrated on a selected few institutions of science and even fewer of those dealing with technology, and to draw attention to those that operate largely outside the established and well-studied framework of intergovernmental bodies. We have also tried to concentrate on institutions whose prime interest is in the fields of science, agriculture, and medicine, and have given comparatively little attention to institutions whose contribution to international cooperation is secondary to other missions. There are hundreds of ways in which scientists, engineers, physicians, agronomists, and social scientists share knowledge and cooperate with each other to discover new knowledge. Indeed, the diversity and richness of this interchange makes it virtually impossible to estimate the total scope of scientific cooperation taking place in the world at any given time.

This chapter, then, will briefly discuss international institutions concerned with science, engineering, agriculture, and medicine whose future seems to us to be particularly important in the near term. It will also try to show how they have evolved and the basis of their importance in today's world. The building, care, and maintenance of international institutions is a difficult and still evolving art. Our analysis has led us to the conclusion that much more needs to be done to make international institutions truly effective. We have tried to explain why this is so and to offer some glimpses of future possibilities.

THE RATIONALE FOR INTERNATIONAL COOPERATION

International cooperation in science is essential today, as it has been in the past. It contributes both to the vigor of U.S. science and to the health and well-being of our planet and its inhabitants. Through cooperation we can help to meet our global responsibilities, and we can address issues that transcend the concerns of any one nation.

In order to be shared, knowledge needs to be transmitted in some common language. Early observers of natural phenomena recognized that for others to see what they had seen required that they be able to describe their observations in words and numbers that meant the same thing to everyone. Thus one of the first, and even today one of the most important, aspects of international science is the development of common standards, units of measurement, and nomenclature, most of which is done by the scientific and professional societies and their international analogs.

Science is an international enterprise and communication is central to its existence. Today many hundreds of scientists move across national boundaries to conduct research in the laboratories of colleagues who are working on like problems.

Some idea of the dimension of the person-to-person contact is suggested by a 1978 survey[1] of a sample of 203 U.S. doctorate-granting institutions and medical schools that showed that three-fourths of these institutions had some faculty who collaborated on research with their foreign colleagues—principally in Western Europe, Latin America and the Caribbean, the Near and Middle East, and Japan. Moreover, during 1977–78 nearly 3,800 faculty members from doctorate-granting institutions participated in cooperative scientific and technical activity with developing countries.[2]

By contrast with basic research, institutionally directed and supported applied research, the totality of which also is difficult to measure, may well involve even more people and require more funds because of the scale of the projects. Institutionalized cooperation in technology is infinitely more complicated than it is in science. Technology embraces so many topics and has so many ancillary attributes, many of which are related to commercial profit-making and national security, that a meaningful examination of these institutions is not possible within the scope of this chapter. We would note here, however, that there are a number of areas in which the engineering and technical community of the United States has pioneered in cooperative ventures, space exploration being probably the most dramatic. Equally important, of course, are the intergovernmental relationships that regulate the movement of aircraft and ships and the agreements that permit our current level of telecommunications. Other vital technological cooperative ventures include the setting of industrial and commercial standards, agreements with respect to patents and their licensing, and the relatively free exchange of engineering information upon which our collective safety and progress depends. Primarily for simplicity, we will pursue a discussion of the history of scientific cooperation as illustrative, while recognizing that cooperation in engineering has had a parallel course. The two systems of international cooperation complement each other and are essential to the advancement and application of knowledge.

A scientific observation should be susceptible to duplication (excluding of course the observation of one-time or one-place phenomena). Thus if a scientific paper is properly written, the experiment or observation it describes can be reproduced in the laboratory of a distant colleague. This obviously provides for a desirable redundancy that produces independent verification or refutation of alleged advances of scientific knowledge. In recent years this facet of global science has been deliberately constrained by the scientists themselves (with some encouragement from legislators and budget-makers) when building the huge and very expensive machines of physics. Regular meetings of the directors of high-energy physics laboratories have enabled joint planning so that each major new machine has the capacity to do something the others cannot do and unnecessary duplication is avoided. Instead of building machines in many countries, the practice is to bring scientists from other countries to the machine, at obvious savings of resources.

Another element of the rationale for international science is the globally designed and executed experiment utilizing many observers throughout the world who follow common and agreed-upon instructions. Reaching back to the earliest astronomers who shared their observations about the course of the stars, we have come to such extraordinarily productive projects as the International Geophysical Year 20 years ago and today the Global Atmospheric Research Program (GARP) and the World Climate Program (WCP). In these programs highly sophisticated research is being planned by scientists with the understanding that major and minor powers of the world will contribute according to their resources of manpower and facilities, but they will all receive the full results of these efforts to understand the world's weather and climate.

The techniques of sharing our constantly expanding reservoir of knowledge have not changed very much in a long time except in scale. The basic medium for exchange is still the journal article, but because it often takes many months, sometimes years, to assure publication and because the number and specialization of journals has proliferated, international scientific meetings and symposia have become a complementary medium of great importance to scientists. Further, informal information acquired about a colleague's research in progress frequently leads to personal and group collaboration, avoidance of unintentional duplication, and modification of research plans, techniques, or instrumentation. Conducted today—largely in English and drawing specialists from all over the world, the scientific conference serves several purposes. Here the common language of a discipline receives its regular infusion of new terms; and the process of agreeing upon names, values, and unit descriptions is begun. Also here, results and ideas are exchanged and new research strategies are often planned and sharpened. The significance of a recent experiment or theory can be examined by others similarly engaged, and plans for collaboration of mutual benefit can be worked out, whether the task be that of mapping the world's oceans or sharing the instrumentation of space facilities.

In recent years the scientific conference has been used as a model by the United Nations and its family of agencies for the organization of intergovernmental debates on population, the environment, food and nutrition, resources, and energy, albeit in a largely political rather than scientific atmosphere. These partially scientific, partially political conferences constitute public recognition of the need to apply science and technology to pressing public problems, most of which require international cooperation for their amelioration. The scientific community has felt their impact in terms of the redirection of research funding and priorities and in newly discovered interests and challenges.

SOME CHARACTERISTIC INTERNATIONAL INSTITUTIONS OF SCIENCE AND TECHNOLOGY

A very rough approximation of the division of labor among the international institutions of science and technology is that the nongovernmental entities tend to support the advances of science as an end, while the intergovernmental organizations tend to be devices to convey technology as a means. But it quickly becomes apparent that most nongovernmental and intergovernmental organizations are attempting both.

The broad sweep of scientific work of major international agencies such as the World Health Organization, the World Meteorological Organization, the International Atomic Energy Agency, the Food and Agriculture Organization (FAO) of the United Nations, the U.N. Development Program, or the U.N. Environmental Program, to list the most prominent, is generally well known. Each of these institutions has an objective for the furtherance of which the United States contributes in funds, ideas, people, and policy direction. We have chosen other lesser-known models of international cooperation where the work of individual scientists and scientific organizations has been and can be, we believe, particularly productive in the future. In the discussion that follows, we will, however allude to certain of the specific activities of some of the U.N. agencies as they are also illustrative of models to be pursued.

In the academic world the oldest, best-known, and most effective of the international nongovernmental organizations of science is the International Council of Scientific Unions (ICSU), a sort of umbrella organization for many specialized scientific bodies.[3] ICSU is composed of 18 autonomous scientific unions and more than 60 national members, such as academies of science, research councils, or like institutions. The main purpose of ICSU is "to encourage international scientific activity for the benefit of mankind."

In addition to fostering the efforts of the specialized unions and their component commissions serving the broad range of the physical and life sciences, ICSU initiates and coordinates multinational research programs in such areas as space, oceanographic, and polar research, and also serves as a vehicle to communicate scientific information and develop standards in methodology, nomenclature, and units.

Starting with the voluntary membership of individual scientists in their home country's professional societies, the institutionalization of international science flows through national committees for each discipline to national academies and similar bodies to the international unions of scientists (e.g., the International Union of Pure and Applied Chemistry), and then via the collaboration of these unions within the ICSU to form special project-oriented multidisciplinary committees and commissions that attempt to bring order and purpose to international science (Figure 32).

ICSU itself was formally created in 1931, but its origins can be traced back to the nineteenth-century associations of astronomers and earth scientists that were created, as are most institutions, to meet specific needs (in this case those whose work manifestly transcends national boundaries) and to the International Association of Academies (1900, London, Paris, St. Petersburg, Washington). Today it encompasses all of the natural sciences. Organizations comparable in terms of effectiveness or recognition do not yet exist in either the professional fields of medicine and engineering, or the disciplines of the social sciences, although there have been nominal analogs in existence for some time.

In the case of ICSU, the support and assistance of the major world science academies has been a decisive undergirding element. Neither medicine, engineering, nor the social sciences has equivalent national institutional bases, although strong and internationally minded engineering academies have been established in several countries in recent years. The first international convocation of these academies was held in October 1978 under the auspices of the U.S. National Academy of Engineering.

The strengths of the international nongovernmental science unions are in a way also their weaknesses. They are sustained by the efforts of unpaid scientist volunteers, often supported by small and poorly paid secretariats. The work they do is subject to funding constraints that are always austere at best. Because of other priority calls upon the time of their leadership, projects move slowly and are largely dependent on the tenacity of a few very dedicated individuals.

A strength of ICSU, as distinct from the work of its member unions, lies in its eleven special and scientific committees, five interunion commissions, and three permanent services, which provide for shared disciplinary approaches to space, climate, water, science information, and other topics.

Three examples of ICSU scientific committees follow:

- The International Biological Program (IBP) was organized to coordinate research on the biological basis of productivity and human adaptability to environmental changes. American scientists studied ecosystems of major economic importance such as forests, grazing lands, and arid lands. International collaboration made possible the collection of comparable data from several sites around the world.
- There has been increasing recognition of the needs for interdisciplinary approaches to environmental problems. Accordingly, in 1970 ICSU established the Scientific Committee on Problems of the Environment (SCOPE). It has two main tasks: to advance knowledge about the influence of human activities on the environment and to serve as a nongovernmental source of advice on environmental problems.

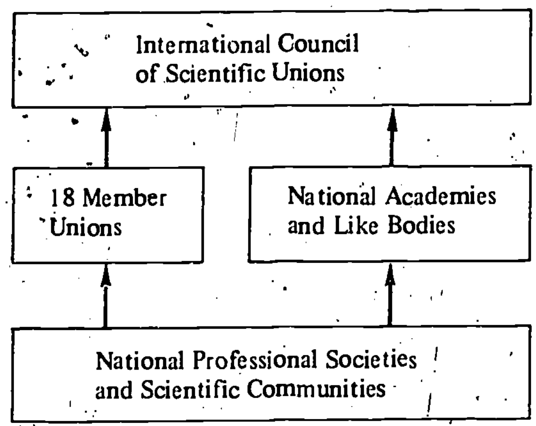


FIGURE 32 Organization of the International Council of Scientific Unions.

SCOPE has seven activities at present: biogeochemical cycles, dynamic changes and evolution of ecosystems, environmental aspects of human settlements, ecotoxicology, simulation modeling of environmental systems, environmental monitoring, and communication of environmental information and societal assessment and response.

- More recently, the Committee on Genetic Experimentation (COGENE) was established by ICSU to serve as a nongovernmental, interdisciplinary, and international council of scientists and as a source of advice on recombinant DNA for governments, intergovernmental agencies, scientific groups, and individuals. Its functions are: to review, evaluate, and make available information on practical and scientific benefits, safeguards, containment facilities, and other technical matters; to consider environmental, health-related, and other consequences of any disposal of biological agents constructed by recombinant DNA techniques; to foster opportunities for training and international exchange; and to provide a forum through which interested national, regional, and other international bodies may communicate.

COGENE is currently concerned with: guidelines for research on recombinant DNA; risk assessment experiments; and benefits and applications of recombinant DNA.

In addition to these ICSU examples, there are joint activities with intergovernmental agencies.

In what has come to be viewed as a model in international institutionalized scientific cooperation, GARP began in 1961 when advances in understanding the theory of atmospheric processes and expanding computational capabilities led President Kennedy to propose to the United Nations that the ICSU and WMO mount an international program of atmospheric research. A joint organizing committee of 12 world-renowned meteorologists was appointed by the two organizations and accorded great autonomy in planning a comprehensive program of unprecedented scope.

By 1974, a large coordinated research effort was conducted in the eastern Atlantic Ocean with the objective, among others, of examining the relationship between convective activity and the large-scale wave-like disturbances that are frequently the progenitors of hurricanes. Several thousand scientists and technicians from 72 countries, aided by a thousand land stations, 40 ships, 12 aircraft, and 6 satellites, brought under observational surveillance 20 million square miles of land and ocean. The research was successful and mathematical models have now been developed that are capable of predicting the development of tropical atmospheric wave disturbances, their westward propagation, and of simulating their rainfall pattern.

This success led to an even larger experiment launched in December 1978.[4] Embracing the entire globe and involving 5,000 scientists and technicians from 150 countries, the Global Weather Experiment will last for 1 year and involve 5 geostationary satellites, up to 4 polar orbiters, more than 40 oceanographic research vessels, thousands of miles of specially instrumented aircraft flights each day, more than 300 instrumented balloons circulating freely about the equator at a height of 15 kilometers, and over 300 ocean buoys in the oceans of the Southern Hemisphere, each capable of transmitting observations for 9 months. In addition, thousands of commercial ships and aircraft have been instrumented to report data from wherever they are located.

Large-scale computers process the immense quantities of data received daily and put it in a form useful for the research effort that will extend over the next decade. Plans were laid at the World Climate Conference in February 1979 to use this unique data base in a World Climate Program aimed at understanding natural climatic variability as well as change induced by human activity.

Testimony to ICSU's value is the continued donation of volunteer time and the continued use of ICSU to stimulate new scientific ventures. Within its formal institutional framework, it is estimated[5] that, on the average, each year about 40,000 scientists come into contact with colleagues from other countries through ICSU-sponsored conferences.

Among institutions where science is pursued as an end, there are portions of the U.N. family of specialized agencies—such as WMO and WHO—that support and sustain basic research (both directed, or applied, and nondirected). Almost invariably, however, the research now supported by U.N. agencies is linked to some social or economic goal and frequently it is tied directly to a technological means. Thus, the S in UNESCO, which used to be science *qua* science, is now labeled "natural sciences and their application to development."

Within the realm of the health sciences, three programs of the WHO deserve mention for the way they illustrate international cooperation. That smallpox has essentially been eradicated in the world is directly attributable to a program initiated in 1966 by the WHO. In this instance, international cooperation was required, first in the production of vaccine, second in its donation by producing countries and its eventual production in developing countries (which in a mere three years produced vaccine that was acceptable internationally for its potency and stability), and finally in an improved vaccination technique that required the development of a delivery system that was usable in all sorts of cultural environments by people of varying technical abilities. The program hinged on the cooperation of governmental health departments and private (in many cases profit-making) concerns who gave willingly and generously to the goals of the program.

A complementary WHO program, also begun in 1966, is international collaborative research on tropical diseases. In this instance, six diseases (malaria, schistosomiasis, filariasis, trypanosomiasis, leishmaniasis, and leprosy) have been selected for concentrated research by scientific working groups in a number of locations around the world under the general guidance of WHO steering committees. The heart of the program is the linkage, via formal and informal networks, of research being undertaken in these diseases in different centers in different countries and the agreement among the participants to exchange information and collaborate in the design and operation of the research projects.

WHO's third program, again more than 10 years old, is the special program of research development and training in human reproduction. The program is designed to meet the needs of developing countries and has three main goals: developing contraceptive technologies, designing technologies that are acceptable and that can be easily implemented, and enhancing developing countries' increasing self-reliance in basic and applied research. The program, which is funded by both developing and industrialized countries, has met wide acceptance throughout the world.

The International Atomic Energy Agency (IAEA) is a well-known example of a multilateral intergovernmental organization devoted to the solution of problems generated by high technology. Coping with safe levels of irradiation of food, the disposal of nuclear waste, transport of hazardous nuclear materials, ionizing radiation, nonproliferation, and the world's long-term energy needs, this agency collaborates closely with WHO, the FAO, and with national nuclear regulatory agencies. The IAEA also supports research and training in nuclear sciences and engineering, as well as serving as the focal point for intergovernmental agreement on one of the most politically complicated technologies of our age.

Among the international institutions where science is precisely targeted is the Consultative Group on International Agricultural Research (CGIAR). This multinational and multi-institutional group stems from the agricultural research programs originally sponsored by the Rockefeller Foundation in the 1940's and later also by the Ford Foundation. CGIAR was started in 1971 to finance a revolution in the agricultural productivity of less-developed countries. It now supports research and training activities through 11 centers or programs located in the developing nations where almost 600 senior staff and a total of over 8,000 persons study the major food crops and animals in virtually all of the ecological zones of the developing world. The annual combined budget of the centers now exceeds $100 million.[6] Jointly sponsored by the World Bank, the FAO, the U.N. Development Program, and the U.N. Environment Program, it has 18 donor governments, 4 private foundations, 3 regional development banks, the Commission of the European Communities, the International Development Research Centre (IDRC), and the Arab Fund for Social and Economic Development affiliated with its program.

The centers began with a strong orientation toward purely biological research, but they became the home of the green revolution. Centers are now developing plant types and practices to benefit small farmers who must operate without much help from controlled irrigation, fertilizers, and pesticides. The centers are also taking into account the social customs, consumer habits, and other societal habits affecting agricultural productivity.[7]

The CGIAR family consists of the following centers and programs with their locations and dates of founding:

CIAT

International Center for Tropical Agriculture
(Colombia) 1967

CIMMYT

International Maize and Wheat Improvement Center
(Mexico) 1943

CIP

International Potato Center
(Peru) 1971

ICARDA

International Center for Agricultural Research in the Dry Areas
(Egypt) 1976

ICRISAT

International Crops Research Institute for the Semi-Arid Tropics
(India) 1972

ILCA

International Livestock Center for Africa
(Ethiopia) 1974

ILRAD

International Laboratory for Research on Animal Diseases
(Kenya) 1973

IITA

International Institute for Tropical Agriculture
(Nigeria) 1968

IRRI

International Rice Research Institute
(Philippines) 1960

WARDA

West Africa Rice Development Association
(Liberia) 1971

GENES Board

International Board for Plant Genetic Resources
(Italy) 1974

In September 1978 the CGIAR produced "An Integrative Report" on the Consultative Group and the International Agricultural Research System. Assessing the impact of high-yielding varieties of rice and wheat (the only two major innovations introduced so far by the CGIAR that have been in use for sufficient time to permit study) on agricultural production and in terms of social and economic impact, the CGIAR drew a number of positive conclusions. For example, in the 12 years since the high-yielding varieties were introduced, they have come to occupy over one-third of the total acreage being sown in these two grains.

One major conclusion of the review, concerning an issue in controversy, was that the principal research strategy of the international centers, that of developing yield-increasing technology, contributes both to more equitable distribution of the benefits of economic growth and greater agricultural productivity. A non-controversial finding was that the returns on the dollars invested in agricultural research have been truly impressive.

Complementing this remarkable illustration of international cooperation in the application of science and technology to a specific global goal are other efforts more modest in scale and scope. For example, one of the most useful institutions for scientific cooperation with Western Europe is the Science Committee of NATO, which manages a $9 million program directed toward basic fundamental sciences rather than programs related to the military and defense objectives of the alliance. The bulk of the funding is used to provide grants or fellowships and for the support of specialized scientific meetings. In the past 20 years, some 75,000 individual scientists and engineers have participated in NATO's science activities.

Another NATO innovation was the Advisory Group for Aerospace Research and Development (AGARD), the history of which goes back almost to the founding of the NATO alliance. This group, which has a fundamental military orientation, is nonetheless a remarkable example of international scientific cooperation. AGARD has many of the characteristics of a professional society and has made contributions very like those made by the professional societies in other disciplines.

AGARD is but one example of very high level technical cooperation in practical fields that have received great impetus from the international military and security requirements of industrialized countries. The civilian applications of these technologies are in many cases as important as their military uses. Certainly NATO- and AGARD-initiated research on aircraft manufacture, safety, and space applications has equaled comparable contributions by other professional, technical, and scientific bodies.

The International Institute for Applied Systems Analysis (IIASA) near Vienna has been successfully launched by an *ad hoc* consortium of scientific bodies, supported by public funds, to explore new ways to understand and manage complex systems in areas such as energy, food, environment, water, and urban centers. It brings to bear the experience and insights of scholars and institutions from all over the industrialized world—from both market and planned economies.

IIASA, first organized in 1972, brings together scientists of different disciplines, cultures, and nationalities to work on problems of concern to mankind and to improve analytical techniques and their usefulness for decision making. IIASA's scientific staff consists of approximately 70 persons, supported by member dues, who spend varying periods of time, ranging from a few months to a few years at the institute, plus another 20–30 persons supported by grants and contracts from other sources. The institute's 1979 budget has been set by its 17-member international council, its governing body, at 130.6 million Austrian schillings (about $9.6 million).

During 1979 the institute expects to complete work on the final report of five years of research on global energy prospects. Approximately two years later work will be completed on its second major program, that in food and agriculture.

The Stockholm Institute for Peace Research (SIPRI) and the International Institute for Strategic Studies (IISS) in London are modest but scholarly and productive efforts addressing the problems of conflict. Despite the highly political and controversial nature of the issues with which they deal, they have acquired respect derived from their objective and scientific approach. SIPRI is an independent research institute with a particular focus on disarmament and arms control. Financed by the Swedish government, but with an international board, staff, and scientific council, it publishes an annual authoritative yearbook of world armament and disarmament, as well as reports and books on special topics in arms matters. Its funds (of approximately 4.5 million Swedish kroner, or about $1 million) come from the Swedish Parliament. By contrast, the IISS is an organization whose international membership is dedicated to scholarly analysis of all types of issues related to international security and specializes in the production of the *Adelphi Papers*—a series of essays. The IISS income, privately subscribed, is $240,000 per year.

The Norwegian government supports the International Peace Research Institute-Oslo (PRIO) which is a somewhat newer institution devoted to study of different types of violence. PRIO publishes studies and two journals, the *Journal of Peace Research* and the *Bulletin of Peace Proposals* under international auspices.

Another institutional innovation, modest in size, is the International Foundation for Science (IFS), which gives grants to young scientists for research relevant to development. Founded in 1972, IFS is a nongovernmental organization sponsored by scientific academies and research councils of 42 countries, two-thirds in developing and one-third in industrial parts of the world. At present 10 countries contribute to the foundation's budget, largely by government grants through academies or research councils. The budget for 1978 was $1.4 million. IFS supports the work of ung grantees who must be native to, and carry out research in, a developing country, and where the institutions of the grantees contribute the salaries and basic support, often at amounts several times the foundation grant.

The IFS is *sui generis* in the sense that it is stimulating a network of individual researchers in several fields of applied biological research. The fields chosen have substantial economic and developmental impact potential and the research is of a sort especially appropriate to the needs of developing countries. The seven principal topical areas are: aquaculture; animal production; vegetables, oil seeds, and fruit; mycorrhiza studies, afforestation problems; fermentation, methods for food preparation; natural products; and rural construction.

Another institutional innovation is the International Development Research Centre (IDRC) established by the Parliament of Canada in 1970 to foster research in development problems and in the ways to apply knowledge to economic and social advancement in developing regions.[4] The IDRC enlists experts from Canada and other countries to help developing regions build research capabilities, innovative skills, and institutions to solve their own problems. It encourages coordination of research and fosters cooperation between developed and developing regions.

The goal of IDRC is to emphasize the role of the scientist and engineer in international development, and it encourages developing countries to use their own scientific communities. It is funded by and reports to the Parliament of Canada, but it is administered by an international and autonomous board of governors.

It is a governmental aid organization created to support research projects originated and conducted by developing country researchers in their own countries and in terms of their own priorities. It has helped to create research networks that allow developing countries to share experiences and conduct studies with a common design in areas of mutual concern.

IDRC's areas of concentration are: agricultural and nutritional sciences, health sciences, information sciences, publications, and social sciences and human resources. Since its inception, over $100 million of IDRC support has been provided for 550 research projects, of which more than two-thirds were in science and technology.

Perhaps the best known of the basic scientific operating organizations financed by governments is CERN, the Conseil Européen pour la Recherche Nucléaire, formed by collaboration among 12 Western European nations. It is an outstanding example of international collaboration both on scientific and political levels. Its laboratories serve a community of more than 2,000 physicists, including many from the United States, some in permanent residence, but most of whom come for short working periods. CERN's present facilities represent an investment close to 1 billion current dollars by the European governments that support it. Its accelerators and support facilities are among the world's best. CERN has led major advances in our knowledge of the fundamental properties of matter and of the fundamental forces governing the behavior of matter in the universe. A proton–proton interacting storage ring is currently in operation, and a massive proton–antiproton interacting storage ring is proposed.

The important role of the large multipurpose funding institutions which support international scientific and technical activity deserves recognition. These institutions support research and applications in many ways, both directly and indirectly, as part of their broad programs. From the World Bank to the U.N. Development Program and the U.N. International Children's Education Fund (UNICEF), to the major private foundations of the world, there are many large and small patrons of science and the application of technology. Without the support these institutions provide to national and international research and its technological applications, many scientific and technological institutions would not survive.

SOME SPECULATION ABOUT FUTURE NEEDS AND PROSPECTS

Over the next five years the debate on how science and technology can contribute to a lessening of tensions between East and West, how science and technology can ameliorate conflict between North and South, and how they can solve the problems of the poorest of the poor will continue. How and to what extent can science and technology provide solutions to a wide range of essentially political problems? With accumulating experience we hope for increasing convergence between scientific and technical solutions and political problems, but the difficulties in achieving such convergence are manifest in the lengthy—and still not very fruitful—conferences on the Law of the Sea and in the thin margin of agreement and rather large residue of disagreement that remained after the 1977 U.N. Conference on Trade and Development. In these forums the consensus form of decision making and crosscurrents of national interest tend at best to reduce the crispness of definition of issues while simultaneously blurring the boundaries of acceptable compromise.

The international institutions that are now and will in the future be regarded as constructive and worth saving or encouraging are those that are perceived to reflect the interests of a number of nations, each of which, in fact, may have quite diverse motives and concerns.

Many of the major international intergovernmental organizations based on a one-country—one-vote principle appear to be coming under the effective control of developing country representatives. The result of this control—most notable in some agencies of the United Nations—may be a shift toward regional and bilateral arrangements where national interests are believed to be more precisely matched. In a sense, this return to free bilateralism may be likened to the rationale that underlay the initial arrangements for scientific cooperation in the early part of the last century.

We perceive three generic clusters of problems with which international institutions will be called upon to deal. These clusters of problems are perceived in the environment of East–West and North–South competition. In some respects they are manifestations of the differences that exist in natural resources and geographic limitations. In other respects they represent the differences in scientific and technological capacity and productivity that exist among countries.

- Interdependency problems such as the global effects of an increasing build-up of CO_2, multinational management of water basins, and exploitation of the seabed under some yet-to-be-agreed-upon international arrangement.
- Distributive issues of equitable access to natural resources, including assured food supply, reliable sources of energy at affordable prices under acceptable environmental constraints, and vastly expanded delivery of health care, adequate housing, and educational opportunity—all of which have science and technology components. These issues of distributive equity relate both to a nation's domestic concerns about meeting the needs of its people and to its international aspirations and the obstacles it faces in global competition.
- Vulnerability concerns such as those that have to do with radioactive wastes and the development of nuclear power, levels of acceptable environmental risk, natural catastrophes such as drought, and the worries of some that our technical capabilities are outpacing our political management skills.

THE INTERNATIONAL BACKGROUND

American interest in international institutions relates directly to our domestic experience. In the 1950's and 1960's America was the unchallenged leader of world science. Today, after monumental, postwar U.S. reconstruction assistance, science and technology have steadily expanded in Western Europe and Japan. America still contributes the single largest and most successful share of the total global scientific exercise—about one-third—but its role is changing.

One of the manifestations of the change in our scientific status is the decrease in postdoctoral exchanges between the United States and Western Europe. Reasons for this are complex and not fully understood, but it is clear that the young scientists of Europe no longer perceive a postdoctoral fellowship in the United States as critical to their career development.

At the same time, Americans are not going abroad for postdoctoral work to the extent they once did, because funding for this purpose is relatively scarce and because they may experience problems in reentering the U.S. job market. This is paradoxical because now, when our colleagues abroad have more to give, our postdoctoral community engaged in research at foreign institutions is at its lowest in years. We know, for example, that the number of those with new doctorates in science and engineering who had firm commitments for postdoctoral study abroad at the time of their degrees peaked in 1971 at 409. By 1977, that figure had dropped by 55 percent to 186. The proportion of this group with firm commitments for postdoctoral study abroad also decreased from 2.4 percent in 1971 to 1.2 percent in 1977, a 50 percent drop. This decrease occurred during a period when the proportion of doctors of science and engineering with firm postdoctoral study commitments increased slightly.[8]

The U.S. educational system is being utilized by increasing numbers of foreign students in our institutions of higher education. Their needs and the obvious importance they assume as the U.S. population of college-age youth declines have not escaped attention. The 235,000 foreign students currently enrolled in U.S. institutions of higher education now are expected to increase to a million within 10 years, reflecting the growing needs of the developing nations and their ability to provide funds. Today, 27 percent of all engineering doctorates and 45 percent of all engineering degrees go to foreign students.[9]

The United States is also still a favored country for advanced training by scholars from the Soviet Union, Eastern Europe, and now the People's Republic of China. The United States has some 16 bilateral scientific and technical agreements with the Soviet Union in addition to the long-standing bilateral exchange arrangements conducted by the National Academy of Sciences and International Research and Exchanges Board. Exchanges with the People's Republic of China have been conducted on a formal delegation-for-delegation basis since 1972, but are now entering into more conventional and productive exchanges between individual students and scholars, lecturers and researchers, which may become a truly significant means of reestablishing the much-desired personal relationships so important to sustained and fruitful international scholarly relations. The value of these exchanges is hard to measure, but the individual exchanges with the Soviet Union have gone on now for 20 years and surveys of participants indicate that in their judgment the exchanges are scientifically productive as well as politically useful.[10]

Competition

European strength in basic research and the impressive Japanese capacity for technological innovation have clear and obvious implications for American industrial concerns. In addition, the more advanced developing nations are determined to carve out a share of the technologically based Western consumer and capital goods markets. Thus, in the United States and in many other industrialized countries there are some fears about loss of jobs and business to foreign competition. While the magnitude of concern is difficult to gauge, there appears to be reasonably strong, albeit limited, protectionist sentiment developing with respect to the U.S. export of high technology.

In the developing world there is a strong pull toward technological self-reliance and a negative reaction to all forms of dependence on former colonial powers. There is also in developing nations a resurgence of interest in their traditional cultures and in the spiritual values embodied in those cultures. Congruent with this concern for their cultural heritage is a search for individual equity and international social justice.

Human Rights

Furthermore, a broad range of humanitarian concerns have assumed singular importance in the eyes of the scientific and technical community. These are the questions of human rights and repression, freedom of inquiry, travel, open immigration, dissemination of research results, and free circulation of scientists and ideas. None of these issues is new, but their enhanced visibility and the expectation of an accepted set of international standards of national behavior have opened new avenues for personal expression of concern by scientists. The revulsion against repression has caused some disruption of scientific exchange. For example, 2,400 U.S. scientists have recently announced their intention to severely restrict cooperation with the Soviet Union as a protest against prison terms meted out to certain Soviet scientists. To the extent that human rights are issues between individual members of national communities of scientists, they will affect and complicate the ways in which the scientific and technical community interacts in almost all of its global institutions.

Even in the most dire circumstances, scientists who are at bay in those countries where freedom of thought is proscribed insist that the exchange with their colleagues abroad is a more effective measure than the practice of boycotting. They make the point regularly and consistently that their professional lives as scientists depend on the ability to communicate with their colleagues abroad. They also take the long historical view and realize that scientific relations are fragile and take time to build, and, if they are upset by political concerns, these relations can wither quickly and at considerable cost.

WHAT DO WE NEED TO KNOW MORE ABOUT?

We have come to the realization that we still have much to learn about transferring technology to developing countries effectively. We are also concerned that these countries appear to be having substantial difficulty adapting, absorbing, and utilizing some of the technologies that have been or are in the process of being transferred.

Second, highly imaginative and creative efforts are needed to fashion new modes of collaborative research. With the resources for research concentrated in the industrialized world and pressing needs for research concentrated in the developing world, there is a gap that must be bridged. In part this issue is being addressed by some of the institutions described in the early part of this chapter and reference will be made later to other institutional innovations.

Third, far more knowledge is needed about the resources and availability of scientific and technical labor. We are witness to over- and undersupplies of scientists, engineers, managers, and technicians throughout the world. India ranks third among the countries of the world in terms of numbers of scientifically and technically trained people, but is first to admit it has yet to find effective ways to employ all or most of them.

Fourth, we also need to know more about how to induce change in international institutions. As a case in point, the major countries of the world for various reasons have long resisted the establishment of more agencies in the U.N. family. This resistance has been based on three considerations: First is the perception that the U.N. agencies now encompass just about anything one wants them to, so why create another agency if a new task can be undertaken in an existing agency? Second, there is no desire on the part of the funding nations to increase their annual dues. Third, as the *ad hoc* Working Group on Policy for Science and Technology within the U.N. system has said, the

United Nations already contains a wide spectrum of activities in science and technology which are not at all well coordinated.[11] These issues must come to the surface and they will demand solution. Fourth, many of the global problems (e.g., climate) that require international cooperation and an appropriate institutional framework are interdisciplinary and we are still at a rather primitive stage in constructing institutions that are effective in crossing scientific-technical and social-political boundaries.

Finally, how far can the networking formula used so effectively by the CGIAR and some other highly directed institutions be applied? Networking is the organized and structured cooperative arrangements that are put in place between institutions to facilitate their communications and exchange of data to enable mutually reinforcing activities.

WHAT WE'VE LEARNED

Having alluded to the strengths and weaknesses of international institutions, it is important to identify the strengths upon which we can build more effective international scientific and technical cooperation.

Creativity and flexibility characterize the volunteer nongovernmental institutions. They can communicate very rapidly, move ideas quickly around the world, and develop unofficial communication arrangements of great sensitivity and usefulness. Theirs is a highly idiosyncratic and individualized activity based on people, not organizational momentum and mission. However, resources are usually insufficient for follow-through and they are dependent on external financial support.

In recent years the United States and partner nations have experimented with multinational technical assistance organizations and with bilateral science and technology agreements, joint commissions, and the like. At present the United States is engaged in 38 government-to-government bilateral agreements in areas that include agriculture, space, environment, energy, the oceans, natural resources, health, housing, defense, transportation, development assistance, and science and technology in general. A much larger number of agreements exists between U.S. governmental agencies and their counterparts abroad. There is no question in the minds of observers that, while both multilateral and bilateral approaches are needed, unless the goals sought are well designed and clearly articulated and unless adequate financial and human resources are dedicated, the experiments will be faulty. We have learned from sad experience that applying science and technology to political problems does not solve the problems.

As we have noted before, this chapter is not the place to analyze the bilateral and multilateral scientific and technological arrangements of the U.S. government. There are, however, a very few of these arrangements that might be cited for their unique characteristics with respect to scientific cooperation.

Bilateral support and networking of research institutions, the sharing of information, and the mutual recognition of priority scientific research targets can be extended into sectors of economic and developmental effort as well. A number of experiments in regional collaboration are getting under way and more should be expected, as should new forms of disciplinary collaboration and cooperative research.

For example, in 1970 the United States and South Korea agreed to collaborate in the establishment of a Korean Institute of Science and Technology (KIST). Starting from its incorporation in 1966, when little industrial R&D was being performed in Korea, KIST has become an organization that at the end of 1977 employed nearly 1,000 persons, (including 300 research scientists) with contracts worth about $13 million, 57 percent of which came from industry. The principal areas of R&D (by total contract value) were chemistry and chemical engineering, mechanical engineering, electricity and electronics, metallurgy and materials, and food science and feeds. KIST has supported Korean industrial development through its R&D and development of pilot-scale operations, analysis of advanced foreign technologies, development of new products and processes (from 1967–73, 131 patent applications were made), and repatriation of overseas Korean scientists and engineers. In addition, KIST has fostered the establishment of several promising new research organizations such as institutes for electronics and for ship-building and design.

The U.S. National Academy of Sciences and the Brazilian National Research Council undertook 10 years ago to modernize Brazilian chemical research resources. The result, after 17 young U.S. chemists had spent two to three years each teaching in the universities of Rio de Janiero and São Paulo and 15 senior U.S. chemists had made regular trips to Brazil, was the phased introduction of 196 highly trained Brazilian chemists into Brazilian education, industry, and research.

On the international scale, the Swedish International Development Agency has created a Swedish Agency for Research Cooperation with Developing Countries to promote research that helps developing countries achieve greater self-reliance and economic and social justice.

A consortium of 8 Sahelian countries and 12 industrialized donor nations has formed *Le Club des Amis du Sahel* to coordinate research to halt and reverse the problems of desertification and drought in the Sahel.

In the United States the prospective Institute for Scientific and Technological Cooperation (ISTC), within the government, has been designed as a grant-making and coordinating body to improve the availability and application of technology abroad and to expand knowledge and skills needed to meet developing-country problems in a framework of mutual benefit and partnership.

The key to ISTC operational style and technique will be direct collaboration of developing country and U.S. experts from the start. The purpose of this approach is to capitalize on the lessons learned in some 30 years of development assistance activity and simultaneously to recognize the presence in developing countries of a new generation of technically capable people who are determined to set their own domestic priorities. The specifics of ISTC's approach remain to be developed and will be central to its success.

Another U.S. innovation that we believe has promise was signaled by the adoption of Title V of the Foreign Relations Authorization Act of 1978, relating to the role of science and technology in foreign policy. This recognition by the Congress of the growing importance of science and technology in the conduct of diplomacy as well as in the management of foreign affairs is a positive step toward effective international cooperation.

PERSPECTIVES: THE NEAR-TERM ENVIRONMENT FOR INTERNATIONAL INSTITUTIONAL CHANGE

Science and technology have become increasingly visible as a part of foreign policy as well as of trade and academic exchange. The tradition has been that science is a part of world culture and technology is part of world commerce, but today these distinctions are blurred. Our views originate largely from the academic world and we have been concerned primarily with the scientific and technical cooperation initiated by scientists and engineers outside of government and carried out by a mixture of official agencies and nongovernmental voluntary organizations. Here, again, the distinctions are blurred because the scientific communities of individual nations cooperate with each other under a wide array of institutional mechanisms. Underlying these institutions, and fundamental to their existence, is the understanding of the need to extend and share knowledge. This is what scientific cooperation is all about.

In the international environment, most of the time institutional change appears to proceed at a very slow pace. Except for very occasional and sometimes dramatic political events that force rapid change, international institutions for science and technology move slowly.

We note that the growth of new international organizations more or less stabilized in the 1960's and, while the bulk of the new international institutions we have discussed (IDRC, IFS, CGIAR, etc.) were formed within the past decade, we do not predict a surge of new organizations in the next decade. Funds to start new organizations are more difficult to obtain and the realization of the immense effort required to make a new international institution work has permeated the thinking of both innovators and funding sources. We therefore expect a period of consolidation and shaking out as the more obviously successful and effective organizations lay claim to the available funds. How this will affect efforts to fill newly perceived and important gaps is not at all clear, but the existing institutions may not be capable of meeting these needs. Substantial renewal and adaptation will be required if international institutions are to meet the challenges facing them.

We foresee increased interest in process within institutions—a realization that our current ways and means might be modified to accelerate the transfer of information and of managerial technology. At the same time we foresee no diminution in the universal reliance on direct contact and the transfer of accumulated personal wisdom and experience.

As noted previously, for the past decade or so the major powers have sought to curb the institution-creating instincts of members of the United Nations. There are no signs that this mood is changing; therefore we suspect that most of the institutional risk-taking will be arranged outside the U.N. framework. However, with appropriate intervention, the U.N. agencies can work a significant improvement in the use of the large resources at their disposal. These agencies have the staff, funds, and technical means to coordinate their work in development and in other areas of scientific and technical support far more effectively than they are now doing.

A straw in the wind of change was the 1978 Buenos Aires U.N. Conference on Technical Cooperation among Developing Countries (or TCDC as it came to be known very quickly). Billed as a forerunner of the U.N. Conference on Science and Technology for Development (UNCSTD), TCDC may in fact have reached a decisive turning point with the decision of the developing countries to seek ways in which they can help support each other in the uses of science and technology. The meeting, which had little press coverage or participation from the industrialized countries, was highly regarded by the developing-world participants and is seen by them as a milestone in cooperative initiatives.

The August 1979 UNCSTD is a unique manifestation of the crosscurrents affecting the way science and technology will be used in the future. The attempt of the conference organizers to embrace the whole of science and technology in an economic, political, and diplomatic framework is proving to be very difficult. In the meantime, several privately sponsored preparatory meetings have developed constructive, albeit limited, options for the conference to consider. Some of these options are referred to in our text. Whether or not UNCSTD is declared a success is of little account, considering the groundswell of encouragement and the myriad solid technical proposals for human betterment that its planning and design phase has elicited. The conference is not an end in itself but part of a process that, we believe, will be destined to be both long term and productive.

While the trends of international institutions in science and technology are largely in the direction of applications to meet human needs, the institutions of science *qua* science will continue to be needed and to be fostered for the practical reasons of disciplinary order, as noted earlier, but, far more important, for the support they give to the continued search for new knowledge.

Two areas where institutional inventiveness might make a significant difference are the application of science and technology to economic development and the technological element of arms control—despite the argument that neither issue is centrally a science and technology problem and both are political, economic, and technical in nature.

In the case of economic development, we urge prompt and full exploration of the simultaneous creation of three forms of international cooperation linking national and international organizations: a consortium of financial donors, a network of knowledge-generating institutions in developing countries, and a mechanism to match specific scientific and technological skills of the industrialized nations to specific problems of developing countries. The donor consortium (which already exists in some respects) would be located predominantly in the developed countries. The knowledge-generating and implementing networks should be fostered mainly in the developing nations. The United States' proposed ISTC may approximate the beginning of a U.S. response to these needs, but it is too early to tell. The need for a well-articulated donor network of national foundations, interacting with national and international research centers and with the national agencies that actually create and maintain agricultural, health delivery, and other institutions, is apparent. This is a unique opportunity for exploration requiring innovative, vigorous, and effective leadership.

Arms control also presents a real challenge. Where there should be an international core of agencies funding research, virtually nothing exists. Where there should be a network of agencies generating and assembling knowledge at the international level, we find SIPRI, PRIO, and IISS, as well as a handful of more narrowly based institutions. We ask if this is sufficient.

The well-known Pugwash and Dartmouth conferences have been valuable for personal contact and wide-ranging exploration of issues. Pugwash is recognized as one of the earliest vehicles by which Soviet-American dialogues on disarmament were given high visibility by technical experts. Today the number of avenues for exploration of new arms control and disarmament initiatives is vastly more comprehensive than it was in the 1950's when Pugwash began. The question now is how to stimulate and sustain analysis in depth of the security issues that divide East and West and to grapple effectively with the arms races in the developing world. Pugwash interest in economic development has diluted its emphasis on disarmament.

The institutional politicization of the 1960's and early 1970's seems largely to have run its course. It is reasonable to anticipate an increasing degree of sophistication about the pressing needs of developing countries: technology transfer, the application of science and technology to development, and a greater recognition of the need for attention to our international institutional resources. The basic global infrastructure has not been built, but a substantial foundation has been laid.

It is clearly in the U.S. national interest to build and to sustain a community of cooperative institutions, working linkages, networks, and consortia that are sufficiently innovative to contribute to the quest for global accommodation as well as the health of U.S. science and its efforts to meet American technical responsibilities.

As we approach the many issues involved in building, renewing, and adapting institutions over the next five years, it is important that we do not become divided into the "uncritical lovers" and the "unloving critics" of institutions described so vividly by John Gardner. The former shield their institutions from life-giving criticism, literally smothering them in an embrace of death, loving their rigidities more than their promise. The latter are skilled in demolition, but almost willfully ignorant of the arts by which human institutions are nurtured, renewed, and made to flourish.[12]

OUTLOOK

The following outlook section on institutions for international cooperation is based on information extracted from the chapter and covers trends anticipated in the near future, approximately five years.

In the near future, a number of issues will draw increasing attention to the workings of international institutions of science and technology. One of these is the environment in which these institutions work, now beset by two types of polarization: conflicts between East and West and North and South. Both result from differences in scientific and technological capacity and productivity between and among nations. With the resources for research concentrated in the industrial world, and the pressing need for expanded research in the developing world, new methods of technology transfer and collaborative research must be found.

The issues we describe here are in many respects overlapping. Nonetheless, we foresee three major areas for concern:

- Interdependency problems such as the global effects of an increasing build-up of carbon dioxide, multinational management of water basins, and exploitation of the seabed under some yet-to-be-agreed-upon international arrangement.
- Distributive issues of equitable access to natural resources, including assured food supply, reliable sources of energy at affordable prices under acceptable environmental constraints, and vastly expanded delivery of health care, adequate housing, and educational opportunity. These issues have two facets. One is internal to each nation and is a direct measure of the well-being of individual citizens; the other relates to the nation's place in the world and its ability to secure the needs of its people in global competition.
- Vulnerability concerns such as those related to management of radioactive wastes and the development of nuclear power, levels of acceptable environmental risk, and natural catastrophes.

We do not see a surge of new organizations to deal with these problems in the next decade. We can expect a period of consolidation and possibly even thinning out as the more successful and effective organizations receive the available funds. We can anticipate changes in the character of a number of institutions that will require a complex mix of scientific and technical knowledge from more than a single discipline if they are to deal successfully with global problems. Newly perceived institutional gaps remain to be filled, and it is important that a conservative stance toward the formation of new organizations not be used as an excuse to avoid institutional innovation.

The status of U.S. science and technology is changing and will continue to change as a result of international competition and as the more technologically advanced developing nations seek to acquire an increasing share of the technology market. This will cause increasing concern in the United States.

The developing countries are newly aware of the potential for cooperation among themselves. This, combined with their apparent influence in many international institutions, may result in a shift toward regional and bilateral arrangements that more precisely match national interests. Certainly over the next five years we can expect an expansion of the debate on how science and technology can be effectively used to help solve economic, social, and political problems.

Two specific areas, among many, where institutional inventiveness can make a significant difference are the application of science and technology to economic development and the technological aspects of arms control.

We predict increased governmental and academic interest in the process of change within institutions. How can organizations be better evaluated against clear criteria of effectiveness and how can organizational improvement be accelerated?

In applying science and technology to economic development, we believe it will be useful to explore carefully three kinds of cooperation linking national and international organizations: a consortium of financial donors, a network of knowledge-generating institutions, and a mechanism to match specific scientific and technological skills of the industrialized nations to specific problems of developing countries.

To bring the most effective intellectual forces to bear on matters of disarmament and arms control, the small group of existing organizations will have to be markedly strengthened and indeed here new institutions may be needed.

NOTES AND REFERENCES

1. *Scientific and Technical Cooperation with Developing Countries, 1977–78* (Higher Education Panel Reports, No. 40). American Council on Education, August 1978.

2. *International Scientific Activities of Selected Institutions, 1975–76 and 1976–77* (Higher Education Panel Reports, No. 37). American Council on Education, January 1978.

3. ICSU funding consists of $800–850,000 per year and the total expenditures of the whole ICSU family amount to approximately $3.8 million.

4. Gossett, B. First GARP Global Experiment. *WMO Bulletin* 28:5, January 1979.

5. From records maintained in the Office of the Foreign Secretary, National Academy of Sciences, Washington, D.C.

6. CGIAR funds amounted to $86.8 million in 1978 and $103.3 million in 1979 (projected).

7. *CGIAR—Consultative Group on International Agricultural Research.* CGIAR, New York, 1976.

8. *Highlights* (Commission on Human Resources, National Research Council 978). "Trends in Postdoctoral Appointments Abroad Scientists and Engineers."

9. Watkins, B.T. Foreigners' Enrollment Concerns Graduate Deans. *Chronicle of Higher Education,* December 11, 1978, p. 6.

10. *Review of U.S.–U.S.S.R. Interacademy Exchanges and Relations* (NRC Board on International Scientific Exchange). Washington, D.C.: National Academy of Sciences, 1977.

11. Berlinguet, L. "Views of the Transfer of Technology Based on Recent Programmes in the Third World," presented at a Workshop on Scientific and Technical Cooperation with Developing Countries, Organization for Economic Cooperation and Development, Paris, April 10–13, 1978.

12. 100th Commencement Address, Cornell University, Ithaca, N.Y., June 1968.

ACKNOWLEDGEMENTS

Study Staff and Contributors

DAVID PRESCOTT, University of Colorado
DAVID PRICE, Argonne National Laboratory
WILLIAM R. PRINDLE, National Academy of Sciences
C. B. RALEIGH, U. S. Geological Survey
EBERHARDT RECHTIN, Aerospace Corporation
JOHN J. SCHANZ, Resources for the Future
JOHN SCHIFFER, Argonne National Laboratory/University of Chicago
SHELDON J. SEGAL, The Rockefeller Foundation
HARRISON SHULL, Indiana University
RICHARD THOMPSON, University of California, Irvine
W. MURRAY TODD, National Academy of Sciences
KING WALTERS, Rice University
MAX WEISS, Aerospace Corporation
HATTEN YODER, Carnegie Institution

Editorial Consultants

DUNCAN BROWN, National Academy of Sciences
MARGARET M. CARTER
ELIZABETH M. FISHER
ELIZABETH W. FISHER
BARBARA FURST
RUTH HAAS, Resources for the Future
SAMUEL IKER
WILLIAM KIRK, Stanford Linear Accelerator Center
MARY LOU LINDQUIST
MICHAEL OLMERT
KENNETH REESE
PHILLIP SAWICKI
LYDIA SCHINDLER

	LUCIENNE SKOPEK ELIZABETH STEPHENS WALTER SULLIVAN, *New York Times* WINFIELD SWANSON DONNA TURNER PEGGY D. WINSTON GEORGE W. WOOD, National Academy of Sciences
Production	
National Academy of Sciences	BARBARA S. BROWN ESTELLE H. MILLER DAVID M. SAVAGE
Support	
National Academy of Sciences	LINDA CANNON EDNA SAUNDERS CHERYL SULLIVAN ANN THOMPSON
International Business Machines Corporation	PAULINE THOMSEN
Staff	
National Academy of Sciences	JOHN S. COLEMAN NORMAN METZGER SUSAN C. PERRY
Study Chairman	RALPH E. GOMORY, International Business Machines Corporation
Advisory Board	RICHARD C. ATKINSON, National Science Foundation PHILIP HANDLER, National Academy of Sciences FRANK PRESS, Office of Science and Technology Policy
Steering Committee	BERNARD DAVIS, Harvard Medical School DAVID HAMBURG, Institute of Medicine JULIUS HARWOOD, Ford Motor Company JAMES HILLIER, RCA Corporation THOMAS MALONE, Butler University DAVID ROSE, Massachusetts Institute of Technology ROBERT G. SACHS, University of Chicago JACOB SCHWARTZ, New York University CONRAD TAEUBER, Georgetown University GILBERT F. WHITE, University of Colorado ROBERT M. WHITE, National Academy of Sciences
Contributors	MANUEL AVEN, General Electric Company DAVID BALTIMORE, Massachusetts Institute of Technology JOHN BARDEEN, University of Illinois GORDON BAYM, University of Illinois, Urbana WILLIAM BRINKMAN, Bell Laboratories ROBERT BURRIS, University of Wisconsin ALBERT CLOGSTON, Bell Laboratories JOHN COCKE, International Business Machines Corporation MORREL COHEN, University of Chicago

MORRIS COHEN, Massachusetts Institute of Technology
GEORGE E. DIETER, University of Maryland
CHARLES DRAKE, Dartmouth College
RICHARD EASTERLIN, University of Pennsylvania
EDWARD EVARTS, National Institutes of Health
DEBORAH FREEDMAN, University of Michigan
HERBERT FRIEDMAN, Naval Research Laboratory
GERALD T. GARVEY, Argonne National Laboratory
PAUL GLICK, Bureau of the Census
PHILIP HAUSER, University of Chicago
ROGER HEYNS, W. R. Hewlett Foundation
WALTER R. HIBBARD, Virginia Polytechnic Institute and State University
DAVID HUBEL, Harvard Medical School
ROBERT I. JAFFEE, Electric Power Research Institute
JOSEPH KATZ, Argonne National Laboratory
CHARLES KIDD, George Washington University
LEONARD KLEINROCK, University of California, Los Angeles
RICHARD KRAUSE, National Institutes of Health
HOWARD J. LEWIS, National Academy of Sciences
MALCOLM MACFARLANE, Argonne National Laboratory
CLEMENT MARKERT, Yale University
GORDON MOORE, Intel Corporation
NORTON NELSON, New York University
WILLIAM NIERENBERG, Scripps Institution of Oceanography
NILS NILSSON, Stanford Research Institute
ALFRED NISONOFF, Brandeis University
ITHIEL DE SOLA POOL, Massachusetts Institute of Technology

Reviewers and Additional Contributors

The assistance of the following persons is gratefully acknowledged. The contents of the report are the responsibility of the Steering Committee.

ENTIRE REPORT
Governing Board Members and Assembly/Commission Chairmen

JACOB BIGELEISEN, State University of New York, Stony Brook
R. H. BING, University of Texas, Austin
HARVEY BROOKS, Harvard University
ROBERT H. CANNON, JR., California Institute of Technology
W. KENNETH DAVIS, Bechtel Power Corporation
DAVID R. GODDARD, National Academy of Sciences
FREDERIC A. L. HOLLOWAY, National Academy of Engineering
R. DUNCAN LUCE, Harvard University
SAUNDERS MAC LANE, University of Chicago
COURTLAND D. PERKINS, National Academy of Engineering
E. R. PIORE, International Business Machines Corporation
FRANK W. PUTNAM, Indiana University
HARRISON SHULL, Indiana University
HERBERT A. SIMON, Carnegie-Mellon University
MITCHELL W. SPELLMAN, Harvard Medical School
H. GUYFORD STEVER, Assembly of Engineering, National Research Council
FRANK H. WESTHEIMER, Harvard University
GILBERT F. WHITE, University of Colorado

PLANET EARTH

SAMUEL S. ADAMS, Consulting Geologist
ARTHUR G. ANDERSON, International Business Machines Corporation
JAMES K. ANGELL, National Oceanographic and Atmospheric Administration

305

PAUL A. BAILLY, Occidental Minerals Corporation
ALBERT W. BALLY, Shell Oil Company
BRUCE A. BOLT, University of California, Berkeley
MICHAEL A. CHINNERY, Massachusetts Institute of Technology
PRESTON E. CLOUD, University of California, Santa Barbara
DANIEL J. FINK, General Electric Company
PETER T. FLAWN, University of Texas, Austin
ROBERT G. FLEAGLE, University of Washington
EDWARD A. FLINN, National Aeronautics and Space Administration
WILLIAM W. HAY, University of Miami
CLAUDE R. HOCOTT, Gulf Universities Research Consortium
JOHN D. ISAACS, University of California, La Jolla
DAVID S. JOHNSON, National Oceanographic and Atmospheric Administration
DON E. KASH, U.S. Geological Survey
WILLIAM M. KAULA, University of California, Los Angeles
CARL KISSLINGER, University of Colorado
JOHN KUTZBACH, University of Wisconsin
HELMUT LANDSBERG, University of Maryland
REUBEN LASKER, National Oceanographic and Atmospheric Administration
LESTER MACHTA, National Oceanographic and Atmospheric Administration
VINCENT E. MCKELVEY, U.S. Geological Survey
J. MURRAY MITCHELL, JR., National Oceanographic and Atmospheric Administration
JACK E. OLIVER, Cornell University
WILLIAM L. QUAIDE, National Aeronautics and Space Administration
ROGER REVELLE, University of California, San Diego
WALTER O. ROBERTS, Aspen Institute
DEREK W. SPENCER, Woods Hole Oceanographic Institution
JOHN STEELE, Woods Hole Oceanographic Institution
JAMES V. TARANIK, U.S. Geological Survey
WARREN S. WOOSTER, University of Washington

THE LIVING STATE

K. FRANK AUSTEN, Harvard Medical School
HARRY BEEVERS, University of California, Santa Cruz
FLOYD E. BLOOM, The Salk Institute
ALBERT C. ENGLAND, III, Center for Disease Control
LESLIE HICKS, Howard University
EDWARD KRAVITZ, Harvard Medical School
STEPHEN KUFFLER, Harvard Medical School
BEATRICE MINTZ, Institute for Cancer Reseach
HOWARD A. SCHNEIDERMAN, University of California, Irvine
GREGORY SISKIND, Cornell University Medical School
TERRY STROM, Harvard Medical School

THE STRUCTURE OF MATTER

PHILIP ANDERSON, Bell Laboratories
CHARLES P. BEAN, General Electric Company
JAMES BJORKEN, Stanford Linear Accelerator Center
WALTER BROWN, Bell Laboratories
CURT CALLAN, Princeton University
KENNETH CASE, Rockefeller University
SIDNEY DRELL, Stanford Linear Accelerator Center
GORDON H. DUNN, University of Colorado

E. NORVAL FORTSON, University of Washington
FRANK Y. FRADIN, Argonne National Laboratory
ARTHUR FREEMAN, Northwestern University
FREDERICK J. GILMAN, Stanford Linear Accelerator Center
MORTON HAMMERMESH, University of Minnesota
THEODORE W. HANSCH, Stanford University
DAVID S. HEESCHEN, National Radio Astronomy Observatory
VERNON W. HUGHES, S1Bale University
JOHN B. KETTERSON, Northwestern University
EUGEN MERZBACHER, University of North Carolina
JEREMIAH P. OSTRIKER, Princeton University Observatory
C. KUMAR N. PATEL, Bell Laboratories
P. JAMES E. PEEBLES, Princeton University
DAVID SCHRAMM, University of Chicago
GOPAL K. SHENOY, Argonne National Laboratory
MELVIN SHOCKETT, University of Chicago
RALPH O. SIMMONS, University of Illinois, Urbana
RONALD F. STEBBINGS, Rice University
VALENTINE LOUIS TELEGDI, University of Chicago

COMPUTERS AND COMMUNICATION

LEE L. DAVENPORT, General Telephone and Electronics Corporation
JERRIER A. HADDAD, International Business Machines Corporation
C. KUMAR N. PATEL, Bell Laboratories
ALAN J. PERLIS, Yale University
RICHARD J. ROYSTON, Argonne National Laboratory
THOMAS SCHELLING, Harvard University
MORRIS TANENBAUM, New Jersey Bell Telephone Company
VICTOR VYSSOTSKY, Bell Laboratories
FRANKLIN ZIMRING, University of Chicago

ENERGY

BETSY ANCKER-JOHNSON, General Motors Corporation
SOLOMON J. BUCHSBAUM, Bell Laboratories
ROBERT A. CHARPIE, Cabot Corporation
JAMES CROW, University of Wisconsin
LINCOLN GORDON, Resources for the Future
JOHN HOLDREN, University of California, Berkeley
STANFORD S. PENNER, University of California, San Diego
WILLIAM E. SHOUPP, Consultant

MATERIALS

RICHARD H. BOYD, University of Utah
WILLIAM DAUBEN, University of California, Berkeley
MERTON C. FLEMINGS, Massachusetts Institute of Technology
GEORGE R. HILL, University of Utah
ALLEN KNEESE, University of New Mexico
M. EUGENE MERCHANT, Cincinnati Milacron
GORDON MILLAR, Deere & Company
REINHARDT SCHUHMANN, JR., Purdue University
WILLIAM SLICHTER, Bell Laboratories
DAVID TURNBULL, Harvard University
NATHANIEL WOLLMAN, University of New Mexico

DEMOGRAPHY

CAROLYN S. BELL, Wellesley College
JUDITH BLAKE, University of California, Los Angeles
DEBBY BURGARD, Harvard University
ALBERT M. CLOGSTON, Bell Laboratories

ARTHUR A. CAMPBELL, National Institute of Child Health and Human Development
JOHN W. COLTMAN, Westinghouse Research Laboratories
KAREN CORCORAN, Harvard University
H. RICHARD CRANE, University of Michigan
LILLI S. HORNIG, Wellesley College
LLOYD G. HUMPHREYS, University of Illinois
DOROTHEA JAMESON, University of Pennsylvania
DIANE KENNY, Harvard University
JON A. KROSNICK, Harvard University
OLIVER H. LOWRY, Washington University
JACK E. MYERS, University of Texas, Austin
GLENN C. OAKLEY, Harvard University
JAMES F. PEDULLA, Harvard University
I. RICHARD SAVAGE, Yale University
LINDA UJIFUSA, Harvard University
CHEVES WALLING, University of Utah
SEYMOUR L. WOLFBEIN, Temple University

HEALTH OF THE AMERICAN PEOPLE

EDWARD H. AHRENS, Rockefeller University
JULIUS AXELROD, National Institute of Mental Health
THEODORE COOPER, Cornell University Medical College
WILLIAM H. DANFORTH, Washington University
JOHN R. EVANS, University of Toronto, Department of Medicine
JOHN W. FARQUHAR, Stanford University
WILLIAM H. FOEGE, Center for Disease Control
ROGER GUILLEMIN, The Salk Institute
PETER B. HUTT, Covington & Burling
ROBERT I. LEVY, National Institutes of Health
WALTER J. MCNERNEY, Blue Cross/Blue Shield
SHERMAN M. MELLINKOFF, University of California, Los Angeles
FREDERICK MOSTELLER, Harvard School of Public Health
ARNO G. MOTULSKY, University of Washington School of Medicine
CHARLES H. RAMMELKAMP, Case Western Reserve University, Department of Medicine
DOROTHY P. RICE, National Center for Health Statistics
DAVID E. ROGERS, The Robert Wood Johnson Foundation
DANIEL C. TOSTESON, Harvard Medical School
JAMES B. WYNGAARDEN, Duke University

TOXIC SUBSTANCES IN THE ENVIRONMENT

MYRON K. BRAKKE, University of Nebraska
JOHN J. BURNS, Hoffman-La Roche, Inc.
DOUGLAS GRAHN, Argonne National Laboratory
ALFRED G. KNUDSON, The Institute for Cancer Research
GENE E. LIKENS, Cornell University
BRIAN MACMAHON, Harvard School of Public Health
JOHN D. ROBERTS, California Institute of Technology
BERT VALLEE, Harvard Medical School
E. BRIGHT WILSON, Harvard University

ACADEMIC SCIENCE AND GRADUATE EDUCATION

ALBERT M. CLOGSTON, Bell Laboratories
JOHN W. COLTMAN, Westinghouse Research Laboratories
H. RICHARD CRANE, University of Michigan
ROBERT HILL, Duke University Medical Center
LILLI S. HORNIG, Wellesley College

LLOYD G. HUMPHREYS, University of Illinois
OLIVER H. LOWRY, Washington University
JAMES H. MULLIGAN, JR., University of California, Irvine
JACK E. MYERS, University of Texas, Austin
ROY RADNER, Harvard University
ISADORE M. SINGER, University of California, Berkeley
CHEVES WALLING, University of Utah
SEYMOUR L. WOLFBEIN, Temple University

INSTITUTIONS FOR INTERNATIONAL COOPERATION

LAWRENCE BOGORAD, Harvard University
JOSEPH FEINSTEIN, Varian Associates
DONALD FINK, Consultant
KEITH GLENNAN, Consultant
BRUCE HANNAY, Bell Laboratories
DIXON LONG, Case Western Reserve University
ROGER REVELLE, University of California, San Diego
EMIL SMITH, University of California School of Medicine
ARTHUR K. SOLOMON, Harvard Medical School
DOROTHY ZINBERG, Harvard University

THE GOVERNMENT VIEW

Statements by Selected U.S. Government Agencies

Department of Agriculture

The basic mission of the U.S. Department of Agriculture (USDA) is to enhance the economic and environmental value of America's agricultural and forest lands. To this end, basic, applied, and developmental research programs are in progress across the country. The principal agencies in the Department that contribute to improving the nationwide effectiveness of research, extension services, teaching, and information dissemination in the food and agricultural sciences are the Science and Education Administration, the Forest Service, and the Economics, Statistics, and Cooperatives Service. Numerous issues are emerging that reflect a national societal context within which there are concerns and roles for scientific and technological activities in USDA. These include natural resources and the environment, energy, structure of the food industry, crop productivity, animal productivity, food and fiber safety and marketing, human nutrition, and forestry and related rangelands. Opportunities for scientific and technological contributions in these and other areas, and constraints on this research, are described in the following paragraphs. Short summaries of several other emerging and equally important issues are also provided.

NATURAL RESOURCES AND THE ENVIRONMENT

Research in the large issues of natural resources and the environment requires multidisciplinary approaches in science and the development and transfer of technology. In many cases, cause and effect are remote from each other and not easily perceived, nor may all the consequences be easily understood. Often the issues exist or threaten to exist because of shifts in alternatives, more intensive uses of resources, and the indivisible human sharing of the environment for differing purposes. Thus, the opportunities for and the constraints on contributions to resolution of natural resources issues by science and technology may, themselves, be judged in human value systems as part of the issues. The important issues that need to be addressed in the next five years are identified and discussed briefly below.

Environmental Quality Constraints vs. Productivity

Recent environmental legislation and the resulting regulations are placing more constraints on the American farmer in his effort to produce food and fiber. As the regulations are implemented over the next few years, farmers must change many of their farming practices if they are to be in compliance with the laws. Increased emphasis must be placed on developing procedures to assess the effect of management practices on runoff, erosion, and water quality, and on developing new management practices to ensure a continuing high level of production while maintaining or improving the quality of the environment. This includes such practices as conservation tillage and structures, multiple-cropping, more efficient use of water and fertilizer, and integrated pest management.

There is growing concern involving the long-term health effects of exposure to processing dust and chemicals on workers, and the environmental effects of food and fiber processing on air and water pollution. Research is needed to identify and monitor potential hazards and to develop alternative technologies that are safe and environmentally acceptable.

Maintain Supply and Quality of Water

Industrial and agricultural water requirements are increasing. As a result, groundwater withdrawal in many parts of the world vastly exceeds the rate at which it can be replenished. In the United States, one-fourth of the total gross farm receipts come from the 10 percent of the land that is irrigated, and the use of irrigation is increasing. We are fast approaching the point where fresh water will be *the* limiting factor in food and fiber production. An intensive effort is needed to develop such management practices for utilizing water more efficiently as drip irrigation, irrigation scheduling, and recycling wastewater, and to develop farming practices, including plant breeding, which will require less water. We must also develop procedures and strategies for replenishing the groundwater supply.

Weather and Climate

A major cause of variation in world food and fiber production is the fluctuation of weather and climate. Recent cyclical changes in rainfall and snowfall patterns have led to unprecedented drought conditions in Western Europe and parts of the United States. In other areas, excessive precipitation has delayed planting and created poor growing and harvesting conditions. In a recent study the National Research Council stated: "An increased effort must be made to define the basic functional responses of crops and animals to weather and climate so that we can improve our present systems of crop and livestock management."

USDA is working to improve its capability to develop new methods and improve existing methods for earlier forecasting and is investigating the impact of production inputs and weather on crop yields in different regions. With state experiment stations and land-grant universities, USDA will be developing data bases and analytic models on crop yields, plant characteristics, technological factors, economic variables, and weather to improve its capability to predict agricultural production levels.

Other Uses of Prime Agricultural and Commercial Forest Lands

America's land must produce more food, fiber, and timber today than ever before if agriculture is going to continue to meet forecasted domestic and foreign demand for food. At the same time, additional land is needed for urban development, highways, factories, and other nonagricultural endeavors. Almost three million acres of cropland are so diverted every year. As the better land in some areas is removed from the agricultural base, farmers are forced to use less desirable land. Usually, production is lower, costs are up, and the chances of environmental pollution, through erosion for example, increase. We must develop farming practices to utilize the less-than-prime land more efficiently in the production of food and fiber. This will require basic research on the chemical and physical properties of soil.

Economic and legal analyses of Federal, state, and local policies and programs on rural land use are needed, particularly as these policies and programs affect the allocation of land between agricultural and other uses. Identification of institutional factors and variables affecting land use is needed to provide research and economic intelligence with respect to land use planning and policy, consequent distributional impacts, and the economic effects on U.S. agriculture.

Cost, Income, and Environmental Constraints

Producer costs of complying with environmental regulations, given the present state of the arts, may vary substantially among different geoclimatic farming regions and among different farm products. For example, research results indicate that costs of meeting environmental requirements in disposing of wastes from confined swine enterprises are higher for smaller enterprises, and those costs are affected by temperature, water balance, soil characteristics, and topography. Such differential costs may encourage development of more large-scale swine producing enterprises, with consequent shifts in regional location of production and the consequent necessity for adjustments in resource and products markets. The extent of such impacts, including the cost of meeting the environmental regulations, will depend to some extent upon the development and use of new technologies and technological systems, both in production and in meeting standards of environmental quality. Economic research is needed which would estimate impacts by present and prospective technologies as well as the effect of compliance with environmental policies or regulations upon producers' costs and income.

Water Quality

Efforts to develop area plans for water quality control as prescribed by Section 208 of the Water Pollution Control Act of 1972 have clearly shown that there is a gross deficiency of needed technical information on the fate and pathways of chemicals in the soil and associated water. As a result, estimated performances of Section 208 plans have unacceptably wide confidence limits. It has also become clear that with current and projected plant nutrient prices (and consumption of natural gas for making fertilizers), such technical data are not adequate for developing economically sound fertilizer rate and soil management recommendations. Basic research to develop improved technical information on the fate of chemicals in the soil and associated groundwater is urgently needed for use in such developmental research activities. Improved knowledge of the nitrogen cycle is the most pressing need. These same research needs exist for organic farming and recycling of wastes.

ENERGY

Energy needs present a very complex set of national and international issues. Like food and shelter, they pervade nearly every aspect of scientific, technological, social, and economic existence. And, they are very much interrelated with aspects of food and shelter that are either taken for granted or are basic issues in human social organizations. The aspect of the energy issue addressed here is the production of energy and other products from sources other than nonrenewable fossil fuels.

There are opportunities for scientific and technological contributions that will allow farmers or forest enterprises to produce on-site energy from agricultural wastes, produce and utilize specific energy crops or materials, apply energy conservation measures, and utilize such alternative sources as solar or wind energy. Methane generators, for example, represent a possible disposition of agricultural and forestry waste for energy production. Forest and sugar crop biomass productions are examples of specific energy crops that need to be continually reexamined in the future.

In addition, plant sources are the only renewable raw materials that can provide the various kinds of hydrocarbons that we now obtain from petrochemicals. Rubber, lubricating oils, pesticides, fertilizers, plastics, and solid, liquid, and gaseous fuels have all been developed from plant raw materials. Comparative costs do not presently favor plant raw material sources for these products, but the pendulum is swinging in that direction, and time is needed for developing the technology that will be required if plant-derived industrial hydrocarbons are to be considered as replacements for those from traditional sources.

Existing alternative uses of biomass such as livestock feed and the return of biomass to reduce erosion and preserve productivity compete with the conversion of the biomass to energy. When considering those alternatives, together with current and prospective price-cost conditions, the economic potential may not support converting all types of biomass to energy products. Developing new technology for capturing and using solar energy may be more promising as a research thrust. Solar energy use in agriculture is now limited to a few applications such as drying grain or heating greenhouses. Its use in agriculture may remain limited until some dramatic technical breakthrough lowers costs. Prospects for those breakthroughs appear to justify substantial increases in research in solar energy use in agricultural production and processing.

STRUCTURE OF AGRICULTURE AND POLICY ISSUES

The trend to fewer and larger farms, capital needs in agriculture, and renewed economic pressures on traditional family farms are among the vital issues relating to the structure and organization of U.S. agriculture. Underlying these issues are such factors as government programs and regulations; market instabilities; new technologies; changing organizational arrangements; escalating land prices; national, state, and local tax policies; and increased capital requirements for farming. The consequences may be a continuation of the trend toward larger and fewer farms, increased concentration of wealth in agriculture, increased vertical integration and contractual linkages with agribusiness, reduced decisionmaking autonomy and independence of farmers, and increased vulnerability of farmers, marketers, and consumers to inflationary pressures. Analysis of these issues in depth is needed to equip policymakers with a greater understanding of the impacts of Federal programs, market conditions and technological developments affecting the structure and organization of agriculture, and alternative program and policy options.

STRUCTURE OF THE FOOD INDUSTRY

The food industry operates under a myriad of Federal, state, and local regulations. It also has become highly concentrated and characterized by many of the competitive practices of the industrial sector—product differentiation and advertising, nonprice competition, and high product turnover. The high costs associated with the food marketing system and related government regulations are of concern to both the President and the public. Future research is needed that would contribute to:

- Determining the effects of government regulations, technology, and the structure of the competitive behavior in the food marketing and distribution sector on the costs and other elements of performance related to manufacturing and distributing food and fiber; and
- Identifying new marketing structures and alternatives that would enhance small and part-time farmers' access to markets and consumers' op-

portunities to obtain high quality food at reasonable prices.

Inflation, energy, transportation, finance, regulations, taxation, antitrust, market orders, and other problems have heightened the policy debate concerning cooperatives and their impact. These concerns are likely to continue into the 1980s. Also, as capital costs increase, farmers and their cooperatives face problems of acquiring and managing their financial resources effectively. Research is needed to analyze factors influencing individual cooperative growth, capital needs, and access to financial analyses and procedures for long-range planning for medium sized and smaller cooperatives.

WORLD FOOD AND FIBER FORECASTING

The 1970s have demonstrated the need for improved information in the international supply of food and fiber. Examples include the 1972 and 1977 Russian wheat production shortfalls and the reduced 1978 Brazilian soybean crop, all of which had significant impacts on U.S. commodity markets and farm program costs. Aerospace remote sensing has the potential to provide significant improvement in the Department of Agriculture's information system for domestic and foreign situations. Research in cooperation with the National Aeronautics and Space Administration, the Department of the Interior, and the Department of Commerce is needed for developing and evaluating this potential.

TRANSPORTATION

Changes in transportation—deteriorating railroads and poor roads and bridges—and a variety of government regulations and programs are having an adverse impact on rural America. At present, the Department lacks information to accurately evaluate the effects of these changes on agriculture and rural communities. To strengthen its role as a participant in the determination of Federal programs, regulations, and policies, USDA needs to improve its information and research base in the transportation area.

To provide the necessary data and analyses, research is needed to evaluate the efficiency of existing and alternative transport systems with emphasis on improving local and export systems for distribution of grains, soybeans, and other commodities and to provide a wide variety of services to rural people.

Research needs to focus on investigating the growing mobility problems of people living in rural areas. Information about the adequacy and efficiency of alternative transportation policies, as they relate to agriculture, rural areas, and cooperatives, would be included.

CROP PRODUCTIVITY

Crop output is an item of basic concern in the food and agricultural economy of any society. The declining rate of increase in the yields of a number of major U.S. crops in the last decade has become a matter of increasing discussion and concern. In the meantime, several world food and nutrition studies have reached the same general conclusion: world population will increase to seven or eight billion in the next two or three decades. This means an increasing world food demand which U.S.-produced crops can help meet. Also, the United States is placing increasing reliance on crop exports for relief from foreign trade and exchange imbalances.

Opportunities for scientific and technological contributions to increasing crop productivity may be classified into two broad categories. One is the biology of the plant. The other is the control and management of plant microenvironment.

THE BIOLOGY OF THE PLANT

In achieving and maintaining high levels of crop yields in the United States, traditional approaches to plant breeding and improving plant physiology have been important. These are the approaches to be relied on as the principal current efforts to improve and maintain crop yields and to improve nutritional quality through breeding. Some of the newer technologies should be viewed as supplementary to the basic approaches, and some of the scientific opportunities should be viewed in a long-run perspective.

Taxonomy of Economically Important Species

A thorough understanding of the systematic relationships of groups of organisms permits us to make maximum use of genetic variability in striving to reach specific objectives in crop protection and crop improvement. Such new approaches and technologies as comparative phytochemistry, computer-assisted analyses of population dynamics and correlation data in systematics, and the increased emphasis on cytogenetics, breeding systems, and biosystematics, along with the well-known traditional approaches, provide a basis for success.

Developing Plants for Stress Environments

Losses now caused by environmental stresses (associated with temperature, moisture, air pollution, soil minerals, salinity, and acidity) must be reduced. Crops that can be grown economically on low productive land must be developed. It is through both breeding and cultural practices that the genetic-environmental interaction is optimized. Traditionally, environmental modification has been emphasized independently of breeding and vice versa, but there is an opportunity for breeding to contribute much more. The major refinements will come through genetic tailoring of plant species to meet environmental stresses and pest problems. This calls into use the broadest possible array of genetic diversity in germplasm collections of crop plants.

Plant Physiology and Photosynthetic Efficiency

A concerted multidisciplinary effort is needed to better define each of the stages that help to determine a plant's net photosynthetic efficiency and to define interrelationships among the several functional stages in the total physiological system of the plant. Such knowledge of the physiological systems of plants is a prerequisite to development through genetic and plant breeding research of commercial plant varieties having higher levels of net photosynthetic efficiency and higher yields in a given microenvironment.

Genetic Modification and Research

Genetic modification of cereal grain plants so that they acquire symbiosis with nitrogen-fixing bacteria would sharply reduce energy requirements and costs of production, and would reduce the threat of environmental pollution from the application of nitrogen fertilizers to cropland. A concentrated research effort is needed to identify and define areas and methodologies to use our knowledge of DNA and its potentials for advances in the agricultural sciences and technologies.

Research Constraints

There are two major constraints to optimizing classical plant breeding methods: sources of genetic diversity for our crop plants and their wild relatives are rapidly diminishing in many important source areas, and resources are not currently available for adequately evaluating crop germplasm already in hand. Constraints on developing plants for stress environments include the complexity of fitting stress resistance into the array of objectives breeders must take into consideration, the insufficient evaluation of germplasm collections to identify needed genetic traits, and the need for interdisciplinary teams of plant breeders, physiologists, and soil scientists.

CONTROL AND MANAGEMENT OF PLANT MICROENVIRONMENT

Plant Pest Management

Evidence of adverse effects of chemical pesticides upon nontarget organisms, including people, and the cumulative long-term nature of some of those effects have resulted in increasingly stringent constraints on the use of chemical pesticides. This development has created the potential for a large social return from research to develop environmentally safe and effective plant pest management strategies for a wide variety of commercial crops.

The increased movement of people and goods throughout the world increases the risks of inadvertently introducing new pests. Options for reducing these risks include improvement of quarantine procedures, improvement of methods for early detection of foreign pests, and expansion of research programs on foreign soil aimed at establishing integrated management procedures for foreign pests before they become established in the United States.

New types of pest control approaches derived from the study of plants and animals themselves will be needed to reduce the current dependency on industrial chemicals. Some of these currently exist: for example, plant, insect, and nematode pathogens; pest-resistant plants; insect attractants, pheromones, and hormones; and cultural and mechanical methods for disease, insect, nematode, and weed control.

Research is needed on the environmental, economic, and social impacts on agriculture and rural areas of alternative Federal environmental quality standards, policies, and programs, including the impacts on farm production, costs, income, and consumer prices. Improved economic research tools for policy analysis and evaluation of pest control programs need to be incorporated with nationwide information on pest control, including alternative chemical, biological, and cultural control methods. This would contribute to improved determinations of cost effectiveness and economic impacts of alternative pest control strategies, including integrated pest management practices.

Ecosystem Structures

The economic, sociological, and nutritional demands upon the biosphere by expanding populations necessitate the development of the knowledge and technology required to establish and maintain ecosystems which optimize food production while providing other benefits. The knowledge to manage ecosystems optimally through controlled input, selective utilization of natural resources, and controlled harvests is not sufficient. Research is needed to assess the impact on ecosystems of varying management practices, including the establishment of different crop plant to noncrop plant ratios, crop to crop ratios, and crop to noncrop ratios. Cropping patterns, crop production systems, and pest management systems that contribute to ecosystem management must be developed.

Research Constraints

For protection against the potential introduction of new plant pests, present research opportunities to develop management options are constrained by the need to limit domestic research to biologically safe facilities. Also, resources are inadequate to support preventive research in host countries.

Implementation of integrated pest management technology is lagging. Shortages of resources and trained personnel are important factors hampering attempts to achieve widespread understanding of integrated pest management principles as a systems approach for pest management in major agricultural production chains.

ANIMAL PRODUCTIVITY

The production of animals and animal products on farms usually equals or exceeds the production of crops in value. Research on animal protection and production is essential to the continuation of the plentiful and economical supply of animal foods that the American consumer desires and enjoys. Achieving the goal of reducing costs to both producers and consumers while providing safe and nutritious products is dependent upon a wide range of improvements including reducing the $4.6 billion lost annually due to diseases and pests, improving inherent efficiency, and providing optimum nutritional and management regimes.

As in the case of plants, the opportunities for scientific and technical contributions to animal productivity are considered to fall into two general categories. They are the biology of animals and the control and management of animal microenvironment.

THE BIOLOGY OF ANIMALS

Reproductive Performance of Beef Cattle and Swine

The breeding herd in a meat animal industry may be visualized as an overhead cost to the production of meat by their progeny. The longer the gestation and sexual maturation periods, the smaller the number of progeny per gestation, and the smaller and younger the progeny at the time of sale relative to the mature animals in the breeding herd, the higher are the overhead costs. Recent scientific advances indicate that an increase in the number of progeny per gestation in both beef cattle and swine can be expected to result from research on reproductive physiology.

Animal Production

Increased efficiency in animal production is dependent upon improving genetic capabilities for specific purposes, meeting nutrient needs, improving reproduction rates, and developing management systems that optimize inputs. Significant progress has resulted from studies in many of these areas. Some outstanding examples are a halving of fat in pork carcasses in the past quarter century, markedly increased milk per cow, and increased rate and efficiency of gain in broilers. Major opportunities exist for further improvements. Some of these are (1) continuing emphasis on breeding and selection to improve genetic capacity for production: (2) reducing current losses due to reproductive problems; (3) increasing efficiency of feed utilization through improved understanding of factors affecting appetite and digestibility: (4) developing methods for increasing usefulness of range plants, other forages, crop residues, and other feedstuffs not directly utilizable as foods by people; and (5) modifying ratios of fat to protein in animal products by genetic and physiological approaches while retaining desired flavors and other consumer attributes.

Physiology of Turkey Breeder Hens

Sharp advances in turkey feeding technology have not been matched by advances in the technology for producing turkey eggs and poults. Physiological problems that cause the rate of lay of turkey breeder hens to be low and variable have not yet been defined. Breeding and management practices to correct these problems cannot be developed until such research has been successfully completed.

CONTROL AND MANAGEMENT OF ANIMAL MICROENVIRONMENT

Animal Protection

Reproductive, respiratory, enteric, and exotic diseases, mastitis, pseudorabies, salmonellosis, bluetongue, and brucellosis are major concerns in food producing animals. Interdisciplinary teams are making significant progress on many diseases and with sufficient funding can increase basic knowledge to overcome constraints now limiting progress. Vaccines that are being developed for internal parasites appear promising. Blood parasites are under intensive study, but the research requires development of new basic concepts and approaches, for example, antigen fractionation, attenuation in insect tissue cultures, cell-mediated immunity, and genetic studies.

Animal Production

Recent research indicates that swine can be raised in intensive systems, and adequate further research might achieve increases in efficiency similar to those demonstrated in the poultry industry since 1950. In dairy and beef cattle, large advances might also be achieved through careful management research.

Insect Pest Management Systems

To increase animal productivity through insect pest management systems, major emphasis is needed on selective chemicals, biological control, attractants, and breeds of animals that are resistant or tolerant to the attack of various insects. Of prime importance in developing improved control technology is basic research in insect physiology, genetics, and behavior.

Research Constraints

Undecided consumer and nutritional issues involving dietary composition and food safety, and environmental issues involving increasing livestock concentrations combine to present a set of general, social, and economic constraints on levels of research, technological development, and technology transfer. More specifically, constraints include the (1) removal of chemicals from our current arsenal against diseases and insects by action and regulatory agencies concerned with human health, safety, and environmental pollution; (2) enormous cost of getting a new chemical or drug on the market; (3) development of resistance to drugs and chemicals by pests, and the introduction of new harmful pests; (4) absence of basic research on which to build new technology; (5) rejection by humane societies of methods of controlling predators; and (6) spreading of current resources to include new areas of responsibilities, such as human nutrition and energy conservation.

ON-FARM PRODUCT LOSSES

On-farm losses in agricultural production are extensive. Firm estimates of the losses are not available, but several reports indicate them to be serious. They represent losses of both energy and resources. Some of the losses of raw products occur in harvesting, storage, and processing (such as drying); some are caused by equipment failures due to inefficiencies; some derive from adverse environments. Losses occur in livestock production as well as in the production of food and fiber crops.

Improved harvesting methods, machinery, and technology would reduce losses in harvesting crops. There is a need for basic research on the mechanism of stress and damage in the harvested product. Equipment failures probably indicate the need for better materials technology in equipment fabrication. Losses in the on-farm storage of many food crops can occur because of structural failure due to inadequate design, an improper storage environment, or crop conditions at time of storage. The environment in which livestock are kept greatly affects their production and reproduction efficiency. Research by agricultural engineers has led to improved production and reproduction environments for some livestock. There is an opportunity to expand this research to all livestock and to determine economical methods of providing environments that will reduce losses in livestock production.

Under current price relationships, the technology to reduce agricultural losses is expensive when compared to the value of the losses. But this balance seems to be changing, and the application of new technologies to reduce agricultural losses will be increasingly justified. Nevertheless, the current price relationships detract from a sense of urgency about developing the technology to reduce agricultural losses.

FOOD AND FIBER SAFETY AND MARKETING

Postharvest costs make up two-thirds of food prices and a much higher proportion of fiber product prices. As the world population expands, it will be increasingly important to reduce food and fiber losses in the product processing and distribution system.

Reducing Food Losses

Secretary of State Kissinger, at the 1974 World Food Conference, said, "Major priority must be to reduce losses from inadequate storage, transport, and insects." Postharvest food losses for the United States have been estimated at $31 billion, and losses have been reported in less developed countries of over half the food produced. There is a need for improved protection and preservation of food by environmentally acceptable methods to prevent contamination by insects, microorganisms, and toxic materials and to prevent physical loss. Potential improvements in technology include controlled atmospheres, temperatures, hormone bioregulators, low dose atomic radiation, and pH control. These and other technologies need to be examined and improved individually and in combinations.

Reducing Food Costs

Processing, materials handling, and transport are major cost components in domestic food marketing. New technologies are needed that are energy efficient, reduce handling, and improve product protection without adding packaging materials or transportation costs. Such appropriate technologies as ambient cooling of produce and direct marketing of fresh, rather than processed, foods need to be reexamined.

Increasing Food Safety and Quality

Food and feed must be protected against contamination from naturally occurring toxicants and microbial damage as well as from chemical residues and additives which can affect the health and safety of consumers. Continuing research is needed to identify contaminants, develop rapid methods for detection, determine cause and potential for occurrence, and develop means for prevention, removal, or safe disposal of contaminated products.

Improved analysis of food safety and quality issues and of the cost effectiveness and impacts of current and planned programs is needed. This includes research to:

- Study the nutritional status of the population and evaluate the economic impacts of public efforts to improve that status;
- Determine the economic impacts of Federal regulation of drugs and additives in the production, manufacture, distribution, and storage of food;
- Evaluate existing and proposed changes in food grades and standards for consumers, producers, and food marketing firms;
- Evaluate cost effectiveness of alternative nutrition education programs; and
- Investigate and analyze the roles of price, income, family size and composition, advertising, product labeling, health status, and attitudes on consumers' food purchasing habits.

Utilizing Natural Fibers and Renewable Resources

Natural fibers, including leather, and numerous byproducts from plants and animals compete directly with manmade fibers, plastics, and chemicals from natural gas and petroleum. Research is needed to improve efficiency and reduce costs of converting farm commodities and byproducts into useful consumer products as nonrenewable sources of basic materials become more scarce.

HUMAN NUTRITION

Three recent government reports that deal with national priorities in nutrition research are in essential agreement. The overall issue of major significance is the definition of and means for promoting consumption of optimal diets for growth, development, and prevention of disease at all ages. This is of special importance in view of our rapidly changing national diet. It is estimated that a supermarket may now contain upwards of 12,000 items from which the consumer must select his or her food; many of the items change with time, and the composition of the mix is often unknown to the consumer. Lifestyles and eating habits that have nutritional implications are also changing.

Specific problems (opportunities) for scientific and technological contributions include:

(1) Development of improved data on food composition and improved evaluation methodologies to keep pace with changes in the food supply;

(2) Development of better knowledge of the factors that influence the formation of food habits and food preferences, and the means by which these may be modified;

(3) Assessments of the cost effectiveness of various food and intervention programs, nutrition education, and regulatory activities concerned with food safety and quality;

(4) Development of techniques, equipment, and educational methods to help consumers in the home or in institutions select foods that provide a nutritionally adequate diet at reasonable cost;

(5) Research and policy analyses to deal directly with food-related issues. Such research will require a much improved food price, expenditure, and consumption monitoring system. This system should provide the base for conducting policy studies on the effects of food industry merchandising practices, food and nutrition information programs, and the changing number and size of firms on consumer food prices;

(6) Development of a better understanding of the interactions that occur between nutrients in diets, in the presence of drugs, and related to other environmental factors that affect nutrient requirements and availability; and

(7) Development of better methodologies for the determination of nutrient requirements of individuals and populations, and for the assessment of nutritional status.

The central constraint identified in reports is the lack of a sufficient cadre of research workers. Prominent constraints are the ethical and other limitations on the use of human subjects in research, severe current methodological limitations, and fragmentation of research activities.

CHARACTERIZATION AND EVALUATION OF FOREST RESOURCES

Most everyone favors multiple use of forest lands. But which uses and where? Conflicts are growing as more people demand more goods and services from the nation's limited forest and rangeland base while at the same time there is an increased environmental awareness on the part of the public.

First, we need to know how much we have and what the demand is for each of these resources. We must learn how much of each product or service the land base can produce singly and in combination and at what cost.

Finally, we must be able to answer the question, "What are the physical, social, and economic results of each of the various land use strategies?" Such questions are addressed by the characterization and evaluation of forest resources.

Multi-Resource Inventory and Appraisal

Land managers need information on which to base decisions regarding alternative land use. To provide this information, certain kinds of data are needed.

Presently, we know how to design inventory programs to answer some of the questions concerning potential for timber production and, to a certain degree, range and wildlife habitat. But the need is growing for information on all resources embodied in the forest or range ecosystems. There is need to improve techniques for measuring the relevant elements of forest and rangeland resources and services and to develop more efficient methods for assessing and integrating future demands for renewable natural resources and services.

Multiple-Use Potential and Evaluation

How to choose among alternative land use possibilities and how to best allocate management funds are increasingly difficult problems for managers. Under certain conditions, joint production of goods and services from the same land is compatible, while in other situations use for one purpose entirely eliminates other uses or severely constrains them. The problem is further complicated because various interest groups value these outputs differently. Furthermore, shifts occur in value perceptions and priorities over time.

Research itself cannot resolve land use issues. It can, however, provide the factual basis and techniques for evaluating the consequences of choices between complementary and competing use values. Generally, economic and political considerations will have to be taken into account. So, research can and should improve methods for evaluating the socioeconomic benefits and costs of alternative land use and develop more efficient land use planning systems for forests and associated rangelands.

PROVIDING ADEQUATE SUPPLIES OF WOOD

The need for forest products will double by the year 2000, and timber management research must provide ways to boost timber production to satisfy this need. Concurrently, new and increased demands are being made on forest resources from an expanding human population enjoying greater amounts of leisure time.

Timber management research, therefore, must not only provide new knowledge about the culture, breeding, and economics of growing trees. It must achieve its goals in harmony with other forest values.

Future requirements for timber-based materials can be met in part through better use of our wood resources. It is now physically possible to increase per-acre harvest of wood fiber as much as 50 to 100 percent in many areas by removing defective logs, tops, and branches. Similarly, it is possible to increase substantially the yields of materials from logs and to improve greatly the efficiency of wood product use in buildings and manufactured goods. But research is needed to make these possibilities economically feasible.

Biology, Culture, and Management of Forests and Timber-Related Crops

To provide the information needed to meet the demands for additional forest products, we project that timber culture and management research must assess the capabilities of the forest to produce and sustain timber growth. This means intensive treatments—mechanical site preparation, fertilization, thinning, and whole-tree utilization. It means treatment of hardwoods and softwoods—pure stands and stands of mixed species—for both even-aged and uneven-aged forests.

At the same time, the quality of the forest environment needs to be preserved and enhanced. The problem is complicated by increasing concern over energy supplies. There is need to develop techniques of intensive culture to maximize production of wood fiber on selected sites while enhancing environmental quality and using a minimum of fossil fuels, find better ways to convert unproductive land to forests, and produce growth, yield, and related forest stand parameter models for commercially important species.

Genetics and Breeding of Forest Trees

Genetic gain has the potential of shaping the timber resource to help meet multiple demands of the public. Significant gains have been made, even though much is not known about genetic mechanisms. The results have encouraged the establishment of 13,000 acres of seed orchards for the production of superior seed. On other fronts, development of disease- and insect-resistant stock shows promise.

However, the underlying obstacle to greater progress is time. We must develop advanced techniques to reduce the time period required to test seed and produce improved seed and seedlings.

Economics of Timber Production

While considerable research has been done in pure, even-aged stands on some of the major forest types, we know little about the economics of timber production under mixed stand conditions and in uneven-aged stands. Nor can we identify changes in forest stand values over varied rotation lengths, over a range of site conditions, in conversion of sites from one species to another, and in response to various cultural treatments. In short, we need to improve our capability to determine the costs and benefits resulting from alternative timber management practices and determine the effects of public policies on the production of timber from public and private lands.

Harvesting and Forest Engineering Systems

Lack of appropriate harvesting equipment and systems limits wood production in many areas. Past research has increased our capability to determine operating costs, choose among alternative systems or equipment, and determine the effects of timber harvesting on other forest values. Expanded research is needed to determine what equipment is needed to harvest trees under difficult slope and soil conditions, find the most economical machines and systems for harvesting small trees, determine how to minimize the impact of harvesting and transporting wood on other forest uses, and develop more efficient harvesting operations.

Properties, Processing, and Protection of Wood

One way to help meet timber demands is to put the available wood supply to better use. Research has already paid dividends—for example, some woodpulp now is made from defective logs and formerly unused species. But this is just a start. Expanded research is needed to find economical ways to use low-quality trees and residues, extend the resource base and conserve energy by improving processing efficiency, devise more effective and efficient methods for end-use application of wood, improve wood protection techniques, and develop more effective and economical methods of using wood as a source of energy and chemicals.

Economics and Marketing of Wood Products

New developments in forest product utilization, from harvesting to final product uses, have little chance of practical application unless they are more profitable than existing techniques and are acceptable to potential users.

One example of improved technology is whole-tree utilization, which increases the usable harvest of wood from a given area and reduces the problem of logging residues. Another example is improved structural applications of lumber. Homebuilders are unlikely to use new wood-saving designs unless these designs provide consumer satisfaction at competitive prices. Clearly, there is a need to produce more information on the economics of forest products utilization techniques; provide more accurate marketing information to meet the needs of timber owners, wood processors, and product distributors; and increase our capability to evaluate trends in forest products consumption and trade.

FOREST PROTECTION

Insects, diseases, and wildfires kill an estimated 2.4 billion cubic feet of timber annually—an amount equal to one-fifth of the annual harvest. Insects and diseases cause approximately the same amount of annual growth loss. In addition, the three factors together impair esthetic values, recreational opportunities, watersheds, and wildlife habitat. In rural and urban areas, insect and disease damage can significantly reduce real estate values.

Forest protection research is focused on minimizing these losses. Increasingly, research is concentrating on preventive measures as well as control methods.

Control of Insects Affecting Forests

We have a good understanding of the biology and habits of various bark beetles and some of the major defoliators, such as the Douglas-fir tussock moth, the spruce budworm, and the gypsy moth. Cone and seed insects, tip and shoot moths, hardwood borers, and insects of young stands have received less attention. Additional research is needed in a number of areas—for example, use of behavioral chemicals as attractants or repellants, biological control agents, pest-host relationships that determine population trends, and the influence of various silvicultural practices. We also need more information on insect-disease relationships—namely, insects as disease vectors, or one agent predisposing the host to attack by the other.

Control of Diseases, Parasites, and Nematodes Affecting Forests

We know that stresses from soil compaction, moisture deficiency, air pollution, improper site selection, and mechanical wounding increase a tree's susceptibility to disease. Tree vigor clearly is a vital factor in disease resistance. For this reason, it is important that we strengthen basic research to maximize the growth and maintenance of healthy trees; improve our capability to detect incipient diseases; identify those diseases registering the greatest biological, social, and economic impacts on forest values; and develop more integrated pest management programs.

Prevention and Control of Forest and Range Fires

Mathematical models have been developed for predicting spread rates and intensities of fires burning in simple, single-strata fuels. These models provide the basis for the National Fire Danger Rating System, and for selecting suppression strategies. But, they are of limited use for predicting the spread of fire in multi-strata fuels. To accomplish that, we need to devise models that will evaluate the amount and rate of heat released for crowning and spotting fires, and fires burning in complex, multi-strata fuels; develop methods of inventorying live and dead wildland fuels to determine chemical and physical properties of plants and plant litter; and increase our understanding of prescribed fire effects on the total spectrum of physical, biological, economic, and social resources.

PROVIDING ADEQUATE SUPPLIES OF HIGH QUALITY WATER, PROTECTING THE SOIL RESOURCE, AND ALLEVIATING POLLUTION

Forests and rangelands are the source of two-thirds of the nation's high quality water. These watersheds also produce timber, forage for domestic livestock, habitat and food for wildlife, and recreation opportunities for people. All these uses influence water quality and yields; soil properties; site productivity for timber, forage, and wildlife; air quality; and quality of the land for recreation. Because these influences promise to become greater as the intensity of forest management increases, research is needed to develop ways to minimize their impact on air, soil, and water quality.

Watershed Protection and Management

Increasingly, we will have to rely on intensive management of the more highly productive forests and rangelands for the bulk of the nation's timber and forage supplies. This will involve the maintenance of range vegetative types, complete utilization of fiber on short rotations, fertilization, and a host of other practices.

Research on the effects of intensive management on watersheds will help determine the short- and long-term effects of intensive management on water quality and on the yield from different kinds of forests and associated rangelands.

Soil, Plant, Water, and Nutrient Relationships

Soil and its nutrient qualities, plants, and water are interrelated in complex ways. These interrelationships determine the productivity of forest and associated rangelands for timber, forage, wildlife, and recreation. They also influence the quality, quantity, and timing of water yield and greatly influence the severity of site disturbance and pollution resulting from management practices. These general concepts are understood, but if we knew the specific interrelationships for different regions of the country, it would help answer management questions posed in other research areas. For example, we could determine specific fertilizer and pesticide prescriptions for forests and rangelands, and the ability of soil systems to produce continuous fiber or forage harvests.

Alleviation of Soil, Water, and Air Pollution

Many possibilities lie ahead for using forests and associated rangelands to reduce environmental pollution. For example, forests and rangelands can become an integral part of waste treatment systems through controlled absorption and neutralization of municipal and industrial wastes. The need is to find out how to use municipal and industrial wastes on forests and rangelands while protecting soil and water quality and human health, and to develop ways to alleviate the impacts of past and present mining practices on forests and associated rangelands and restore these areas to productivity.

IMPROVING RANGE AND HABITAT CONDITIONS FOR DOMESTIC LIVESTOCK, WILDLIFE, AND FISH

Integrated management of forest rangelands is needed for enhancing the range production of beef, lamb, and wool, in addition to furnishing forage and browse for wildlife and providing water, recreation, open space, and natural beauty.

Forests and their waters are an important habitat for many kinds of wildlife and fish. They must be managed to benefit both game and nongame species, with particular attention to endangered species. While past research has developed guidelines for tailoring forest and range management practices to the needs of certain wildlife and fish, especially game species, the need now is to develop information that is more widely applicable to nongame species and a greater variety of habitats.

Management of Range Resources

Increasing public pressures require multiple-use management of forest rangelands for watershed, recreation, and wildlife habitat, as well as for domestic livestock.

Earlier research developed information on vegetation responses to livestock grazing and range improvement practices. Research has also provided rehabilitation techniques using mechanical. fire, and chemical control of vegetation. But little is known about the effects of these practices upon the mix of domestic and wild species that an area might accommodate. To substantially increase the use and value of forest range, we need to increase our capability to integrate various forest range management techniques to enhance a variety of resource uses, while minimizing adverse environmental effects.

Wildlife and Fish Habitat

There is a need to know more about the life histories and behavior of fish and wild animals. We need to identify and quantify habitat requirements for a wide range of fish and wildlife species, predict the impacts of various land uses on fish and wildlife habitats, and develop ways to improve forest and range habitats for a wide variety of fish and wildlife species.

ENHANCEMENT OF FOREST RECREATION AND ENVIRONMENTAL VALUES

Trees and forests provide a variety of amenity or noncommodity values that are steadily assuming greater significance to Americans. The goal of forest recreation and environmental values research is to enhance these values and widen their availability.

Outdoor Recreation

Land managers need to compare the worth of providing recreation opportunities to that of producing wood, wildlife, and water. It is much easier to value timber and forage than recreation, wildlife, and quality water, but our ability to evaluate recreation benefits is needed to identify tradeoffs with other forest and range output alternatives. To aid in making resource allocation decisions, research could produce methods to quantify social and economic benefits from recreation on forests and rangelands, investigate the factors that determine trends in demand for various kinds of forest-based recreation activities, and determine factors underlying wilderness attractiveness and wilderness concepts.

Rural and Urban Environmental Enhancement

Whether found in urban or rural settings, forests can modify and enhance the immediate living environment. Research has provided some useful guidelines for improvement of rural and urban wildlife and for selection of appropriate trees to improve thermal comfort, reduce noise, and enhance landscape values.

It is evident that forest resources can do much more to enhance rural and urban living than has been the case to date. To expand the environmental enhancement of forestry, research could devise methods to breed, select, establish, maintain, and protect forest vegetation for various urban uses and produce guidelines for establishing and protecting shelterbelts to meet the needs of changing agricultural and environmental conditions.

RURAL AND COMMUNITY DEVELOPMENT

Development of rural areas is affected by numerous social, economic, and institutional factors. Much of the research on rural development has been partial and "micro" in nature. More effort is needed to develop comprehensive models that systematically relate development and economic variables as a basis for analyzing, projecting, and forecasting development of rural areas. Specific research areas include demographic trends, manpower programs for rural labor, indicators of rural well-being, revenue sources for local governments, adequacy of community services (such as water, sewerage, health, education, housing, energy, credit, and transportation), and the impact of Federal programs and policies on rural growth.

IMPACTS OF THE ADOPTION OF NEW AGRICULTURAL TECHNOLOGIES

The very essence of mission-supported research and technological development in agriculture is to achieve new and improved technologies. Yet the adoption of new technology is seldom neutral. Almost invariably, there are differential social and economic effects on individuals and firms, and on economic and social groups.

There are opportunities for economists and social scientists to join forces with scientists of other disciplines to evaluate the differential impacts of new technologies before they are adopted, or when new technologies are projected but not yet available for adoption. While there is generally a strong belief in this country that there are net social gains from technological advance, there is a real need to be able to anticipate the distributional impacts of each major new technology under consideration.

In addition to learning about the distributional effects of new technology among users, more research is needed for developing anticipations of the side, indirect, and delayed impacts of new technology, some of which may be adverse, and some beneficial. It is common knowledge that past agricultural technologies have had adverse effects on the environment and employment of rural labor but beneficial effects on food prices and national economic growth. But that is a hindsight view of the impacts of technology. What is needed is foresight about the impacts of technology before completion of development, where feasible, to minimize any prospective adverse impacts.

Department of Commerce

Programs in science and technology at the Department of Commerce are aimed at advancing the economic health and well-being of the nation. The Department's science and technology efforts foster U.S. industrial development and economic growth through the conduct of basic research, the advancement of the scientific and technological infrastructure of the United States, and the enhancement of the process of technological innovation in U.S. industry.

Technological innovation has a positive influence upon economic growth, productivity, employment, and the balance of trade. Advancement of the technological capability of U.S. industry is essential to produce the goods and services requisite to preserving and improving the quality of life of the nation. The role of the science and technology programs of the Department of Commerce is to assist in the realization of these social goals through the advancement of the technology of production. Its specific concern is for promoting the efficient production of goods and services, for improving the competitiveness of U.S. industry in domestic and foreign markets, and for enhancing the ability of industry to meet other social goals. The Department as a whole touches upon all aspects of the industrial innovation process—from the exploration of fundamental ideas, through the provision of direct support, to the diffusion of technological innovations in the marketplace.

The first item in our Five-Year Outlook report, entitled "Domestic Policy Review on Industrial Innovation," concerns the general problem of the nation's reduced level of innovation. The Department's approach to the problem reflects its view of science and technology as an instrument to develop and promote economic growth and industry. Other items in the report focus on specific national problems which may be resolved through our science and technology programs.

DOMESTIC POLICY REVIEW ON INDUSTRIAL INNOVATION

President Carter's concern for the status of industrial innovation in the United States resulted in a major domestic policy review that gave the issue of industrial innovation the highest level of policy attention by the executive branch. The domestic policy review resulted, during 1979, in a report to the President, setting forth recommendations for Federal action for enhancing the status of industrial innovation in the United States. In recognition of the concern of the Department of Commerce for science and technology as a means for advancing industrial technology, the Secretary of Commerce, Juanita M. Kreps, was designated chairperson of this wide-ranging interdepartmental study.

Industrial innovation is a process. It proceeds from an idea, through research, development, and demonstration phases, to the commercialization and diffusion of a new product or process through the market. Research, development, and demonstration are integral parts of the process insofar as they lead to the commercial utility of a process, a product change, or an invention. However, research, development, and demonstration make no economic contribution to industry unless at least some part of the knowledge generated is successfully commercialized; that is, unless the results are either integrated into the production process or manifested in the form of a new or improved product or service.

This review of industrial innovation was based upon a recognition that actions of the Federal Government can have, and have had, significant impact upon most firms' abilities and decisions to innovate. Through careful, systematic consideration and precise formulation of specific policies and programs, the Federal Government can influence the future rate and direction of industrial innovation in a way that will provide positive benefits to the economy and to society.

As a result of the review, the President is considering programs in the major areas of procurement, direct Federal support of research and development, environment, health and safety regulations, regulation of industry structure and competitiveness, economic and trade policy, patents and information.

Need for Government/Industry Program of Cooperation to Advance Basic Industrial Technology

The need for programs within which government and industry can work collaboratively to solve national economic and social problems by advancing basic industrial technology is recognized by a body of public opinion (indicated by, for instance, a 1978 Harris poll) noting the importance of technological innovations to our society; a body of academic thought noting the need for public action to complement private investments in research and development; a recognition by government of opportunities to stimulate technologies which cross-cut the operations of many firms rather than those typically pursued by industry which offer firm-specific, proprietary advantages; a body of industrial thought noting the importance of strengthening the system for innovation in the United States; and an awareness by industry and government of the challenges raised by the massive investments of other nations in new technologies.

One means of meeting this need would be through the cooperative development of generic technology. The goal of such a cooperative technology program would be to stimulate the environment for private development of technologies critical to the vitality of the American economy. Specifically, the program would be designed to spur the production and accumulation of scientific and technical knowledge essential for industrial investment in the commercial development of critical technologies.

A one-year study of the cooperative technology program received the approval of the President and Congress. The Office of Management and Budget recommended that five steps be undertaken in assessing the desirability and feasibility of a cooperative program and subsequently, if appropriate, in developing a program plan. These steps were study of precedent programs; assessment of the need and value of the program; design of a program structure; solicitation of public comment on the proposed program through seminars, hearings, etc.; and refinement of proposed program design and selection of target industries.

ENERGY

Alternative Energy Sources

Dwindling of oil and gas reserves requires the study of alternative energy sources and increased utilization of existing resources. The Department of Commerce currently is involved in programs to apply science and technology to the development of such possible alternative energy sources as the sun, the oceans, and geothermal reserves.

Transportation of Fuels

Oil, gas, coal, and other minerals exist in large quantities in the U.S. arctic, especially Alaska, and also in offshore areas north of Alaska; yet their availability and their economic usefulness are inhibited by lack of adequate transportation. Dependence on pipeline links increases costs, inhibits movement of products other than liquid petroleum or derivatives, decreases system flexibility (as new sources are located), and requires completion of an *entire* system before any product can be moved.

Shipping systems capable of moving cargoes directly from ice-bound ports can alleviate these problems. The Department of Commerce could develop a joint industry program to develop the technology for constructing arctic ships that can travel either through heavy ice or under it and will carry oil, gas, and bulk materials.

Atlantic Coast Refineries

A need for additional petroleum refining and storage capacity exists on the U.S. Atlantic coast, but fear of possible spills from VLCC (very large cargo carrier) operation and unloading, and further degradation of air quality from expansion of local refining capacity, thwarts our ability to import and refine crude oil in that area. The exclusion of U.S. east coast oil port development has resulted largely from the periodic recurrence of collisions, groundings, fires, and explosions attributed to human error, as well as technological limitations dictated by the priorities placed on short-term economies in the selection of design criteria and equipment.

The feasibility of overcoming these limitations should be investigated by designing a new generation of tankers and terminals as fully integrated systems devoted exclusively to the goals of safety, dependability, and productivity. The initial high cost will be offset by the long-term economics of placing a major, deep-water east coast terminal off-shore but close to the center of demand.

MATERIAL SHORTAGES

Serious difficulties and disruptions will arise in future years in the supply of materials for the U.S. economy. Among the factors contributing to the shortages are increasing depletion of high-grade ores, rising dependence on imports susceptible to political disturbances, and the high cost of process energy and of petroleum-based feedstocks. To alleviate these restrictions, scientific and technological solutions should be pursued along the following routes:

Use of Substitute Metals

Replacements for scarce metals used for alloying should be sought by systematically investigating phase diagrams and properties of multicomponent alloys.

Use of Composite Materials

Composite materials (heterogeneous structures consisting of a support matrix strengthened by reinforcing fibers) show higher strength, lighter weight, and greater durability than conventional materials and can, therefore, significantly contribute to materials and energy savings. Increased industrial utilization is inhibited by lack of confidence and high cost. Failure modes and the effect of manufacturing processes, environment, and operational parameters on failures need to be explored. This requires a detailed understanding of the properties and structure of the matrix, the reinforcing fibers, and the matrix-fiber interface. Effective and reliable test and inspection methods must be developed.

Reducing Materials Losses

Significant materials savings can be effected by reducing the losses of $70 billion/year in the United States caused by corrosion, fracture, and wear. Existing technology for testing, controlling, and reducing risks and costs associated with deterioration and failure must be transferred to the appropriate user. The fundamental physical and chemical mechanisms that affect the durability of various classes of materials must be studied, and predictive theories for materials performances must be established.

Waste Recovery

Large quantities of potential resources are discarded as waste: 20 percent of tin and aluminum requirements, 10 percent of ferrous metals requirements, and enough organic materials to supply 2 percent of current energy needs. A market for these materials can be established by developing specifications for the composition and properties of waste materials, by establishing the essential equivalence of recovered materials to virgin materials, and by developing or improving resource and energy recovery processes.

CLIMATE

Alleviating climate conditions that may adversely affect economic development is a major issue.

Susceptibility of Industry and Society to Climate Fluctuations

Seasonal and interannual climate fluctuations adversely affect food production, water resources availability, land use, energy supplies, and other factors that are vital to national security and human welfare. Repeatedly in the past decade there have been examples of climate-caused social and economic tragedies. The Department of Commerce is employing many new and existing scientific and technological capabilities to alleviate adverse effects of climate conditions. Programs are underway to assess the effect of climate on our industrial and social activities, to improve our understanding of climate processes, to improve methods for climate forecasts, to improve dissemination of climatological data, and to improve international cooperation in climate research. Specific program areas involved are:

(1) Identifying potential increases in atmospheric carbon dioxide which may cause global warming leading to severe regional or national climate changes;

(2) Increasing efforts to formulate the link between physical fluctuations of the oceans and gross changes in continental weather on the one hand and fisheries resources on the other;

(3) Identifying climate-sensitive components of the Gross National Product and the Consumer Price Index, and closely examining those cases where the economists have attributed significant changes to the weather;

(4) Evaluating the potential effects of climate fluctuations on national and global grain supplies;

(5) Collecting data on the effects of weather and climate on an individual's physical and psychological well-being;

(6) Investigating the impact of the overall solar-terrestrial environment on societal systems;

(7) Investigating the influence of climate on design

criteria for residential housing so as to minimize fossil fuel consumption;

(8) Developing expertise and measurement standards to measure the influence of anthropogenic activities on the environment;

(9) Improving, refining, and expanding the forecasting techniques used to predict storm surges in coastal waters contiguous to the United States;

(10) Improving our understanding of the mechanisms that generate catastrophic natural events such as tsunamis, high winds, and wind-driven flooding in coastal locations from severe storms; improving our ability to predict such events; applying engineering and design technology to increase the resistance of coastal structures to such events; and improving early warning systems of such disasters;

(11) Developing a prototype operation in the high-risk Appalachian region to test and evaluate techniques for improving regional Federal/state/local flash-flood warning services;

(12) Using new technology in the Agricultural Weather Service program to increase accuracy and reliability of weather forecasts, thus assisting Federal decisionmakers in the management of domestic renewable resources;

(13) Developing the Next Generation Weather Radar (NEXRAD) system to increase accuracy of weather forecasting;

(14) Increasing the potential for weather modification as a means to augment water resources and mitigate natural disasters and weather hazards; and

(15) Utilizing new instrumentation and technology to measure crustal motion with high accuracy and over long distances to identify movements of the ocean floor as a means of insuring the safe and efficient development of ocean resources.

MARINE TECHNOLOGY

Especially during the 1970s, there has been a growing national commitment to the exploration, utilization, and management of the oceans and their resources. National programs in ocean energy, marine minerals, fisheries, climate and weather, pollution control, and environmental monitoring, assessment, and stewardship have increased tremendously. As these activities continue to increase, the need for effective development of supporting marine technology becomes more urgent. There are explicit Federal roles in this ocean technology development.

An ocean engineering organization could be established and fostered within the Federal Government to support and coordinate the long-range engineering research and development required by the U.S. Government's current and future marine activities; to collect, generalize, and disseminate ocean engineering data and information; to avoid unintentional duplication of engineering development in Federal programs; and to provide a forum and focus to coordinate, as appropriate, the research, the long-range engineering development, and the immediate short-range engineering development in the Federal, state, and private sectors.

The Department of Commerce, through its National Sea Grant Program, is sponsoring research in the following ocean technology areas:

(1) Reducing the recurring economic loss due to corrosion of materials in the marine environment by increasing our understanding of the mechanism of corrosion and related phenomena;

(2) Studying the process by which biological fouling is initiated upon material surfaces exposed to natural sea water;

(3) Increasing our understanding of marine sediment transport in nearshore water; and

(4) Increasing our understanding of the circulation on, and exchange across, continental shelves.

OCEAN POLLUTION

The increasing use of the ocean for resource development as well as for waste disposal has produced multiple use conflicts that pose a serious threat to the health of the oceans. There is a lack of knowledge about the effects, fates, and transformation products of the new varieties and increasing quantities of chemicals being introduced into the marine environment. The uncertainty about acceptable levels of pollutant exposures results in very large economic costs for control and regulation.

The Department of Commerce has three programs in this area. They are concerned with expanding our knowledge about the capacity of the ocean to assimilate waste materials so we can make informed decisions on what types and what levels of waste may be safely disposed of in the ocean, research on oil spill concentration and trajectory prediction, and monitoring and research related to ocean dumping.

SEAFOOD PRODUCTION

U.S. food processors have been unable to produce a wide variety of seafood products and market them domestically and abroad. Even though a major share of the world's fisheries are in U.S. waters, much of the harvest is processed abroad because the U.S. fishing industry cannot efficiently catch and process it. The problem can be alleviated by gaining a thorough chemical and physical understanding of products suitable for large foreign markets and determining ways of producing and preserving them.

In the United States most of our traditional fishery resources are already being harvested at or near maximum sustainable yield levels. The demand for traditional seafoods in the United States will become critical within the next decade, resulting in shortages and increased prices for many products.

Production of fish through aquaculture is one means of supplementing traditional stocks. The development of new technologies in closed system recycling of water will eliminate many of the land and water use problems that have inhibited aquaculture development in the United States. Utilization of thermal effluent from power plants and other sources will enhance the efficiency of production of aquaculture installations. New advances in applying the science of genetics similar to those made in agriculture will efficiently provide a healthful and nutritious product.

IDENTIFICATION OF MARINE PATHOGENS

There is a lack of a rapid, efficient, and specific method for assaying the potential of marine waters and marine animals to communicate disease to humans. We currently deny ourselves the use of fishery resources from large areas based upon infrequent determination of fecal coliform counts. There is an assay method available, but it is time-consuming and does not really indicate the presence or absence of viable human pathogens. Utilization of science and technology could develop rapid, accurate techniques for determining the presence of human pathogens in the marine environment and lead to better decisions on the use or nonuse of marine areas for sport and commercial fishing and recreation.

AUTOMATED MANUFACTURING

The annual rate of growth of industrial productivity in the United States (2.7%) seriously lags behind that of highly developed economies such as those in Germany (5.7%) and Japan (9.7%). The low rate of productivity growth has contributed to the nation's balance of payment problems and to the inflation rate. A major opportunity for improving the rate is through widespread use of automation in manufacturing.

Barriers to the use of automation on a larger scale include the difficulty of obtaining capital, resistance by unionized labor, and technological problems. Among the latter is a lack of adequate on-line, precise, and rapid measurement techniques. Improvements in measurement methods utilizing multiple sensors coupled with sophisticated computer process-control techniques would improve automation capability, remove technological barriers, and enhance economic benefits.

REDUCING FIRE LOSSES

Fires in the United States cause approximately $3.4 billion in property loss, 7,500 deaths, and 200,000 injuries annually. Of equal concern are the indirect losses due to fire. These include the cost of local fire protection, the construction cost of building in fire safety, and the product-safety engineering costs that industry recovers from the marketplace.

The physical and chemical dynamics of fire need to be determined so that new or improved technology can be used to reduce fire losses. The optimum way to apply this technology, once it is developed, is through standards, codes, and regulations.

BROAD BAND COMMUNICATIONS TO THE HOME

Today roughly 95 out of 100 American households have telephone service and 97 out of 100 have television sets. However, the technological methods of delivery of the electronic signals—namely, twisted pairs of copper wires, over-the-air broadcast, and coaxial cable—limit the information-carrying speed of the telephone and the number of television channels offered. Demand for a wider choice of communications operations may arise in the next five years as information access takes on greater importance in national life. The technology of optical fiber transmission may be able to supply such communication capacity, or bandwidth, at a reasonable cost per channel.

IDENTIFICATION OF TOXIC SUBSTANCES

About 65,000 chemical substances are presently used in U.S. industry. Synthetic organic chemical production amounted to 165 billion pounds in 1976. Many of these materials are toxic or cancer causing. The ubiquitous presence of a variety of toxic or cancer-causing substances in the environment—asbestos fibers in air or halogenated organic compounds in drinking water—poses a grave risk for the health and safety of the population. A mechanism is needed to assess the potential hazard of these chemicals and to control their risks where appropriate.

Chemical risk assessment requires extensive physical, chemical, and biological data for each substance produced, and such data must be supplied to gain approval for the production of the estimated 1,000 new chemicals developed every year and for the continuation of the production of currently used substances. This burdens industry, particularly small innovative companies, with $300 million annually to test new materials and $2 billion to test existing chemicals. Rational decisionmaking and economic efficiency of abatement methods require that the available data and analytical methods be accurate.

Measurement quality assurance can be carried out efficiently through the use of Standard Reference Data (SRD) and Standard Reference Materials (SRM) provided through the Department's National Bureau of Standards. Investigation of pathways of migration, distribution, and transformation of chemicals in the environment can form the basis of evaluation techniques for environmental persistence and fate modeling. The development of valid, predictive models for classes of substances could help substantially to reduce the total number of measurements required. Development of data base management techniques will ensure the availability of existing information to industry.

RADIATION HAZARDS

The recent explosive growth in the sources of electromagnetic radiation, such as 30 million citizen's band radios, 40,000 miles of extra-high-voltage transmission lines, and widespread use of microwave ovens and microwave communication links, to name only a few, have led to concern about the potential health hazards from nonionizing radiation. In addition, the pervasive use of electronic controls in transportation, consumer products, industrial machinery, and products affecting health and safety (such as heart pacemakers) has led to the proliferation of devices subject to electromagnetic interferences (EMI).

The first step in alleviating hazards from nonionizing radiation is to identify, access, and quantify the hazards. Research results to date have identified thermal effects associated with heating of biological tissue by incident radiation of sufficient intensity. Research results have also identified effects in the absence of appreciative thermal heating, but neither the mechanisms nor the human health hazard of these effects have been ascertained. Thus additional biological research is needed to provide understanding of the mechanisms of interaction in the absence of dominant thermal effects. Lacking theoretical knowledge, empirical animal studies of long-term, low-level exposure are needed to answer pressing public questions as to the existence of cumulative exposure hazards. Epidemiological studies of selected exposed human populations will establish whether nonionizing radiation above a particular level is a probable cause of morbidity or mortality. After such research provides a more definitive assessment of hazard as a function of conditions of exposure, safeguards of an engineering and regulatory nature can be adopted that balance risk and cost far more satisfactorily than our present limited knowledge allows.

Rapid growth in applications of ionizing radiation, coupled with an increasing awareness of hazards associated with naturally occurring radiation, has resulted in a growing number of actions designed to protect the public from hazards of ionizing radiation. These actions are principally of two types: research into the biological effects of radiation and pathways by which it reaches humans, and development and promulgation of regulations by appropriate governmental agencies.

The number of regulations relevant to protection from ionizing radiation has grown at a rapid rate, and their economic impact is significant. It is important that determinations of compliance with these regulations be based on accurate, reliable measurements and data. Equitable and effective enforcement of existing regulations requires measurements that are accurate, reliable, and inexpensive. The various regulatory agencies must be on a common basis to avoid conflict and confusion. Those being regulated must also be on this same basis to avoid conflict with the regulations. Thus, the need is for a common basis of reference and ready access to it. Although the development of measurement dissemination and quality assurance mechanisms should be emphasized, concurrent research efforts aimed at improved standards and measurement science are required.

SCARCITY IN RADIO SPECTRUM

Rapid increases in domestic and global demand for telephone, broadcasting, mobile radio, and navigational services are causing increasing congestion and scarcity in the radio spectrum.

Engineering methods of spectrum management allow us to limit the interference with communications that accompanies congestion. Computerized data bases of equipment characteristics and radio transmission models provide the means for achieving electromagnetic compatibility of the operating frequency and locale of systems. Refinement of these data bases will allow the closer packing of systems in congested spectral regions.

Applied research on how to utilize little-used spectral regions, such as the millimeter wave or infrared regions, will make opportunities for additional communications systems available for use. However, constraints on relieving spectrum scarcity are set by the physical characteristics of wave transmission in different frequency regions. Therefore, the potential for additional spectrum usage is limited.

COMMUNICATIONS IN RURAL COMMUNITIES

Rural communities in the United States in many cases do not enjoy the same extent and variety of telephone, television, and other telecommunications services as urban communities.

Several available technologies can make a contribution to the resolution of this problem. Television translators, low-power and low-range television stations, mobile radiotelephone systems, cable television, regional videotape banks, small satellite earth stations with modest up-link capacity, and direct broadcast satellites may all be employed in suitable combinations to upgrade the rural communications system.

The results would create opportunities for rural populations to increase the speed and quality of health care, share in educational instruction the equal of any in the country, receive continually updated information that would assist agricultural operation, and enjoy cultural and entertainment programs on a par with those of urban areas.

ELECTRONIC TRANSMISSION OF SOME OF THE MAIL

The electronic communications industry estimates that electronic collection, transmission, and delivery of many types of information that previously were carried by the U.S. mails are technically feasible and could be implemented on a large geographic scale in the near future. Types of information amenable to electronic transmission are computer-originated notices and bills, machine-readable data, typescript, and facsimiles of graphic or handwritten materials.

A system of electronic mail can be designed and constructed using currently known technologies and infrastructures such as satellites, data communications, and microprocessors. The system could probably be made economically viable and integrated with the mails and the telephone system.

Department of Defense
Defense Advanced Research Projects Agency

The Defense Advanced Research Projects Agency (DARPA) operates as a corporate research and advanced development center for the Department of Defense (DOD). Its prime objective is to select and develop revolutionary technologies designed to ensure maximum national defense capability. DARPA manages long-range, high-payoff technologies which do not fall within the programs of the individual services. All research is conducted at industrial, academic, or other government laboratories.

DARPA's future priorities will be focused on material sciences emphasizing electronic and electro-optical materials and techniques for sharply reducing the cost of structural components, and on information-processing techniques, with emphasis on those applicable to the problems of command, control, and communications technology.

Current and emerging problems, and opportunities for and constraints on the use of new and existing scientific technological capabilities, are described in the following sections.

CURRENT AND EMERGING PROBLEMS OF NATIONAL SIGNIFICANCE

Land Combat

The problem of imbalance in the quantity of opposing forces is becoming exacerbated as enemy forces improve and modernize, introduce sophisticated weapons into their inventories, and erode the U.S. technological lead in critical defense capabilities. Of particular importance are the advances in tank kill mechanisms, which create an urgent need to improve the protective quality of U.S. armor.

Antisubmarine Warfare

An urgent need exists to improve our means of detecting and localizing diesel submarines operating on battery power and nuclear submarines at extended ranges. In addition, an accurate assessment of the vulnerability of our submarine forces to detectors and their localization by advanced surveillance systems is also of critical importance.

Aircraft Affordability

The United States must develop highly sophisticated airframes, avionics, and weapon systems of equal or superior performance as a counter to the quantity and quality of threat aircraft. Skyrocketing costs in manufacturing place a high demand on new technology for achieving advanced performance at less than or, at a minimum, equal to, the cost of today's aircraft.

Survivability of Space Assets

The value of space assets is continuing to increase but will be limited by the ability of these assets to survive threat environments. New scientific and technical approaches, such as hardening, distributed multiple nodes, and active defense, are needed to achieve the required degree of survivability.

Space Defense

There are technologies on the horizon that could make possible space-related uses of high energy lasers and particle beam weapons. Such weapon systems in the hands of an enemy could threaten our vital satellite network and strategic deterrence by almost instantaneous, long-range target kill and extremely rapid refire capabilities.

Space Surveillance

Costs and survivability considerations are placing accelerating emphasis on distributed space-based infrared and radar sensors to provide an all-weather capability for early warning and assessment of intercontinental and submarine launched ballistic missile attacks, air vehicle detection, and theater ocean and space surveillance. To be credible, such surveillance systems must have the required sensitivity, target discrimination capability, geographical coverage, and survivability. These capabilities require new scientific and technical concepts for infrared and radar sensing, lightweight large optics and supporting structures, and on-board, high-throughput signal processing.

Nuclear Test Verification

Agreements that restrict nuclear testing pose difficult verification problems, and ongoing negotiations for a comprehensive ban on nuclear testing pose additional and even more difficult problems. Over the course of time, these problems could prove to be influences tending to undermine the stability of the agreements. Advances toward solution of many of the verification problems require both broadening the base of fundamental scientific knowledge and the development of highly sophisticated remote monitoring methods. International political trends that reduce U.S. access to friendly foreign bases aggravate existing difficulties in monitoring remote areas of the globe.

National Decisionmaking

Important decisionmaking variables, such as those connected with cognitive processing, problem structuring, and stress, are becoming ever more relevant to effective national decisionmaking. When groups must produce sound defense decisions in time-constrained, information-poor situations, the problems are exponential.

Crisis Management

Given the predominant position that the United States occupies in the global arena and the corollary depth of its global participation, the United States will continue to be confronted with important intranational (for example, Iran) and international (Middle East) events and conditions which will have enormous implications for our national interests. No longer do our indications and warnings (I&W) systems serve us well in all possible situations. Indeed, since 1950, we have only forecasted 54 percent of all the inter- and intra-national crises that occurred. We can no longer rely on single-track, military indicator systems or purely intuitive methods. We are also only partially prepared for the demands that the management of a terrorist incident may pose. Beyond terrorism, other forms of unconventional warfare, including economic warfare, international perception management warfare, deception, advanced propaganda, chemical and biological warfare, and genetic engineering, will require more sophisticated indicator and warning systems and a better fundamental understanding of how economic, political, and military variables interact to "drive" international affairs.

Training

Today, weapon systems are under development that simply cannot be operated or maintained by the "average" operator without prohibitively expensive training and retraining. The situation in this area is critical. Given the projections regarding the quality of recruits in the 1980s, the situation—without remedial action—may worsen.

Command and Control

Military commanders are encountering increasing sophistication in the computer "machinery" they are required to use in managing their resources. Problems in the human factors associated with data base management, information display, and cognitive processing in cooperation with computers are growing at alarming rates. More attention must be given not only to the human-user dimension in the design of information processing systems, but also to operational methods required to effectively process the information brought to "users" in even faster, more reliable, and secure ways. Unless these needs are filled, valuable computer and information resources devoted to command, control, communications, and intelligence operations will be wasted.

Energy-Fuel Substitutes

Diminishing supplies of petroleum, together with the increasing costs of petroleum-based fuels, have resulted in conservation programs and planned storage of strategic reserves that could contribute to national survival during conflict situations. In the event that an enemy should stop the flow of petroleum for an extended period, alternative means are required to supplement existing supplies.

Failure Prediction

The availability of weapon systems certainly varies. Therefore, we are not obtaining all of the military capability achievable if all our systems were failure- and maintenance-free. Clearly, if we could predict the failure of systems and extend system lifetimes by corrective measures, we should be able to increase the availability of weapon systems and achieve a significantly enhanced military capability.

Materials Conservation and Substitution

The supply of critical defense materials can be jeopardized by future political and economic sanctions or direct military action. New materials science and technology are required to extend the useful lifetimes of components containing critical materials, reduce waste in manufacturing, provide substitutes through innovative design and processing technologies, and find improved techniques for recycle, rejuvenation, and rework.

OPPORTUNITIES FOR THE USE OF NEW AND EXISTING SCIENTIFIC AND TECHNOLOGICAL CAPABILITIES

Land Combat

The continuing quest to achieve greater mobility, firepower, and survivability in land combat systems at decreasing defense costs poses difficult challenges to the scientific community. Advances in the understanding of the hydrodynamic behavior of complex armor systems and the development of new combinations of materials with increased ballistic efficiency are uncovering new concepts in improved armor protection. The realization of improved armor protection will significantly degrade the effectiveness of antitank weapons and will lead to the development of lighter weight, air-transportable, more highly mobile and agile combat vehicles. These opportunities will be achieved not just by the synthesis of new materials but by the development of improved performance models and design concepts that will direct the optimal use of materials in greater combination and specificity.

Recent advances in radar technologies, data processing, dispersed sensors, wide-area munitions, and precision guided missiles are permitting the development of integrated, near real time target acquisition and strike systems that can operate long range against multiple targets moving against a broad front. These capabilities are increasing the potential of friendly forces to overcome a quantitative superiority in an enemy's land combat forces by effectively striking and weakening the exploitation or second-echelon forces at long range, before they can have a major influence on the outcome of the battle.

Air Vehicles and Weapons

Significant new capabilities in air warfare and missile systems are emerging with the investigation of revolutionary new aircraft concepts, the application of infrared space surveillance technology to tactical missions, and the development of missiles with multimode sensor systems.

X-wing and forward swept wing aircraft concepts offer new opportunities in improved performance and significant reductions in cost and weight over conventionally designed aircraft. The X-wing aircraft offers the potential for combining, in one flight vehicle, the advantages of the efficient vertical take-off and landing performance of the helicopter with the high subsonic speed capabilities possessed by fixed wing aircraft. This capability is accomplished by a unique circulation control blowing over the edges of the rotor blades to furnish lift and control for all modes of flight. The forward swept wing configuration offers significant aerodynamic improvements over conventional aircraft. The advent of practical composite materials and structures is the key to overcoming aeroelastic instability in the X-wing rotor blade and to achieving suitably stiff, light-weight wing structures that will sustain divergence loads in the forward swept wing. The development of very high specific stiffness, high specific strength aluminum alloys by both rapid solidification powder technology and advanced casting techniques is offering major weight savings and performance improvements for advanced tactical aircraft.

Successes in fabricating high-density monolithic focal plane arrays for storing infrared images offer the possibility of fire-and-forget seekers which are low in cost and permit central impact of ground targets immersed in high clutter. The same basic technology permits the development of very high performance target acquisition devices for potential use with forward looking infrared imagers.

Advances in multimode seeker technology will increase the operational modes of advanced missiles in searching for and locking on targets and in reducing the circular error probability of surgical strikes. However, bandpass requirements for dual-mode seekers, minimum distortion of the transmitted signals, minimum self-emission in the bandpass of interest for infrared domes, and minimum intrinsic scattering at high temperatures pose stringent optical requirements. Furthermore, material removal due to erosion, ablation, and evaporation at high-mach velocities is generally not tolerable in the case of optical windows and infrared domes and must be limited to small amounts in the case of radomes. Multicomponent and multiphase ceramics offer some degree of freedom for extending the transmittance spectra to longer wavelengths and for increasing fracture toughness, erosion resistance, and thermal shock capacity through microstructural control.

Advanced sensors and digital signal processing technologies have potential for providing extremely accurate terminal guidance for second generation cruise missiles. If combined with techniques for providing increased penetrability, range, and payload, the increased accuracy could effectively respond to possible future upgrading of opposition defenses.

Power Systems

Projected improvements in the specific thrust, specific power density, and specific fuel consumption of advanced reciprocating and rotary engines require higher compression ratios, higher combustion and turbine inlet temperatures, higher speeds, and more efficient internal energy conservation and transfer. Radical improvements in the cycle pressure ratio and turbine inlet temperature for gas turbine engines will dramatically reduce the airflow requirement and specific fuel consumption for the same power output and will provide additional freedom in designing efficient turbofan engines for specific military systems. It is now conceivable to project a dramatic upswing in the historical trends in these parameters through the emergence of metal-matrix and ceramic-matrix composites, structural silicon ceramics, improved oxidation- and hot-corrosion-resistant coatings and rapidly solidified powder technology. The full development of these advanced materials will permit dramatic improvements in hot-section reliability at current turbine inlet temperatures, extension of turbine inlet temperatures to 2800°F (1538°C) and above, and a reduced dependence on cobalt and other strategic elements.

Fuel conservation and the ability to employ multiple fuels in military systems is also a critical scientific objective for the future. Technologies being pursued for advanced cruise missile engines will directly benefit the general aviation industry by providing a high-bypass, low specific fuel consumption turbofan engine in the 300- to 1000-pound thrust class. This new engine would offer a reduction in specific fuel consumption over the existing small turbofan engines. Further improvements in fuel conservation and management will result from the development of fuels with higher gravimetric and volumetric BTU contents, substitute aviation fuels in the event of extended periods of petroleum crises, advanced fuel misting concepts, and improved combustion modelling and analysis.

Dramatic increases in inlet temperature and speed for high-performance gas turbine engines will require major advances in bearing materials and design. The performance and lifetime of current high-load, high-speed bearings are limited by wear and fatigue life, thermal breakdown limits of liquid organic lubricants, the efficiency of heat removal, and various failure modes for the bearing race, retainers, and separators. Bearing reliability will become a more stringent safeguard requirement for high-performance aircraft engines and high power density machinery as speeds increase, since a bearing failure can result in catastrophic failure of the engine and surrounding support structures.

Potential technology advances in bearings include the development of improved bearing materials, improved surface processes for wear and corrosion control, and the development of solid lubricants either as the principal or as a backup mode of lubrication. Solid lubricants would overcome many of the inherent chemical limitations of liquid lubricants at high temperatures, under oxidative environments, and over long dormancy times. They can also eliminate weight and volume penalties imposed by liquid lubricant circulation and heat removal systems and reservoirs.

Advances in superconducting and acyclic, segmented-magnet, homopolar (SEGMAG) electric propulsion concepts offer significant reductions in weight and volume (by factors of 4 to 10) over conventional electric drives. The current collection system dominates the geometry and size of new electrical machine concepts having the greatest potential value to the Department of Defense. It is now possible to project the attainment of 3 mega amperes per square meter (Ma/m^2) current collection capacity for monolithic, metal-graphite composite brushes and 7 to 10 Ma/m^2 for multi-element brushes within the next five years.

The successful development of high current density, current collection systems for SEGMAG and superconducting electric machines will make possible several technical revolutions in military systems. Tank and vehicle electrical transmissions with peak power ratings double those of present levels for hydromechanical systems will be developed. The use of distributed electrical drive motors on tanks will not only increase survivability and maneuverability, but will also facilitate the development of novel, articulated tank concepts. Electromagnetic launchers for aircraft, drones, and cruise missiles will lead to more advanced ordnance launchers (such as hypervelocity rail guns) which will be quiet, efficient, and have greatly reduced infrared signatures. Pulse power sources and high energy storage machines for laser and particle beam weapon systems and inertially confined fusion power generation will be investigated.

Antisubmarine Warfare

To aid in countering submarine and submarine surveillance threats, technological advances in acoustic arrays, processors, and multisensor correlation to increase possible acoustic detection ranges are essential. Of potentially great value against the current threat, but also essential to countering a significantly quieter threat, are emerging concepts that are not dependent upon the radiated noise characteristics of the submarine for detection. These include improved active sonar processors, significantly increased source level and lower frequency projectors, nonacoustic detection methods, and precise location techniques.

As the technology of reducing sound levels of the submarine advances, the requirements for more sensitive acoustic sensors for towed arrays, bottom-mounted sonar ranging arrays, and sonobuoys will likewise increase. The new materials developments, namely, optical sensors and biphasic composite materials, offer the potential of order-of-magnitude improvements in acoustic response and greatly extended depths of application. The optical sensor is composed entirely of optical fiber and senses variations in pressure associated with the acoustic field through stress-field changes in the refractive index of the glass. Biphasic, composite materials employ the tailored connectivity of two individual phases to achieve an optimal series of parallel coupling of sensor property coefficients in order to greatly extend the performance figure of merit beyond that possible in a single-phase material. Such an approach in seeking the best combination of materials and ways to process them is greatly extending the performance horizons for piezoelectric, pyroelectric, and magnetoelectric materials. Rapidly advancing developments in strong, ultra-low-loss optical fiber-guides along with advanced sensor materials and devices offer revolutionary advances in towed sonar arrays.

Space Systems

Advanced space defense and surveillance systems have a number of challenging requirements for which new technologies will be required. Among the systems that are projected for operational status over the next 25 years are: high-energy, highly efficient infrared-chemical and visible-eximer lasers to include precise pointing systems and large adaptive-optic beam expanders; large, space-based, infrared, staring surveillance systems to include large adaptive-optic reflectors, high-density focal-plane infrared detector arrays, and on-board data processing systems with massive (10^{14} bits) storage and subnanosecond data processing capabilities; and large parabolic and phased-array radar antenna systems for high altitude, radiometric, strategic target surveillance. Most of the spacecraft required for these applications must be highly tailored to unique structural and system logic concepts, and the customization must be reflected in the earliest stages of design.

Large-scale, deployable, and erectable space structures will require highly innovative structural design and materials engineering concepts to satisfy an increasingly complex mix of requirements for maximum rigidity and minimum weight; control of structural, thermal, and environmental loads; high survivability (against nuclear, laser, and particle beam threats); elimination of creep and relaxation during storage; and ease of space erection, joining, and proof testing. Optimal structural design concepts for zero gravity are likely to be dramatically different from those employed on earth, and new methods for employing stored energy and memory materials (reversibly transformable) will be sought to ease the space erection burden.

Among the emerging materials technologies, metal matrix composites have a number of desirable characteristics for space structure applications. Some promising next-generation materials for postulated environments for large space optics are carbon-graphite, magnesium-graphite, and glass-matrix composites, primarily because of their potential for achieving high stiffness and for compensating or nulling interphase thermal expansivities to near zero over a significant temperature range. In essence, the engineering of these complex materials for large space optics is an attempt to substitute adaptive feedback materials design on the ground for adaptive feedback control of the final mirror in space. Recent advances in nonlinear optics will also simplify the space optics function through wave-front reconstruction techniques.

Passive space surveillance systems are being developed to exploit charged-couple device (CCD) technology in the form of many densely packed chips arranged over an integrated focal plane to detect a wide range of military targets. Such systems will ultimately have sufficiently sophisticated sensors and filters to discriminate background clutter from the energy radiated from the earth, its atmosphere, and stars. They will also integrate many signal processing and detection functions on the same chip to facilitate the use of on-board computers to achieve clutter suppression in real time. This same basic technology has the potential of employment with advanced seekers to develop a new generation of tactical forward-looking infrared imagery which will scan and discriminate high-value ground targets immersed in background clutter.

Some of the critical obstacles to be overcome in realizing high-density focal plane arrays pertain to the growth of device processing technologies for the infrared detector materials. Both extrinsic and intrinsic compound semiconductor monolithic and hybrid arrays are undergoing extensive research to determine their ultimate limits of performance. Intrinsic detectors, such as mercury cadmium telluride, offer several advantages over easily adjustable cutoff wavelength and, most importantly, an operating temperature in degrees Kelvin two-to-three times that of an extrinsic silicon detector of comparable spectral response. Hence, the cooling burden for a surveillance spacecraft would be substantially lower for an intrinsic detector focal plane array. In addition to advanced infrared detector array technology, advances are also required in spectral filters for target discrimination.

For applications requiring lasers in space, visible lasers offer significant advantages because of optical system design penalties imposed bv longer wavelengths. However, technical problems in developing a high-power, ultraviolet-visible, noble gas, eximer laser system include the development of mirrors and windows for control of the laser beam. The laser pulses are so intense that antireflection and reflection-enhancement coatings are frequently blown off with a single laser pulse. If they are not, the components may optically distort and defocus the laser beam during the one microsecond duration of the laser pulse.

Although performance goals for diffraction-limited optics represent significant state-of-the-art advancements in laser materials technology, window and mirror substrates can be commercially obtained with near theoretical limits in absorption and with low-scatter surface finishes. However, coating technology has not progressed to the same degree. Typical coating absorbance is several orders of magnitude greater than corresponding bulk values due to anomalous absorption attributable to hydrogen. The degradation of these coatings in space radiation environments may impose special protection requirements.

High energy laser technology also offers the promise of propelling small bulk payloads into the earth's orbit. Technical and economic feasibility questions need to be answered through an appropriate scientific research and development program to determine if advantages exist compared to conventional approaches.

Nuclear Test Verification

The U.S. ability to verify nuclear testing limitations will be enhanced by advances in basic knowledge of seismic source mechanisms, both for explosions and for earthquakes; geophysical characteristics of the earth that affect seismic propagation, especially structural and absorptive features that induce distortions in seismic signals; complex interactions between explosions and the geophysical environment that produce potential observables (e.g., electromagnetic pulse, density changes at the earth's surface); and new sensing principles or differences in characteristics of signal and noise that can be exploited for increasing detectability. Advances in sensors for measuring direct emanations from nuclear explosions in space and for detecting direct or indirect effects of explosions (acoustic, seismic) from systems operating in international waters or space will also improve U.S. verification capabilities, as will advances in analytical and data processing methods.

Communications

Communications is one area in which it is easy to see how to exploit the rapid decrease in the cost, size, and power requirements of digital logic. There are a variety of ways to trade computing for communications bandwidth. Examples are bandwidth compression, error detection and correction, multiplexing, dynamic scheduling of channel capacity, and dynamic routing to balance the load on available circuits.

Packet communications technology provides a basic model for systems that are extremely flexible and that maximally exploit these tradeoffs in data communication. The value of the packet communications concept has been proven in a variety of fixed land line networks, in a mobile radio environment, and in the Atlantic packet satellite experiment. The feasibility of sending speech over a packet network has been demonstrated at rates as low as 2.4kb/s using linear predictive coding techniques. As a result of these successful demonstrations, industry is placing high priority on the development of all digital communications systems that support multimedia transmission of intermixed speech, data, and pictorial imagery. The proposed systems are all variants on the basic packet communications concept formulated by DARPA and its contractors in the mid-sixties, and first demonstrated in a substantial way in the ARPANET in the early seventies.

The transition to all digital communication will greatly simplify the problem of providing secure communications for U.S. forces scattered all over the globe. Data encryption systems operate in the digital domain, so secure voice systems must currently convert the speech from analog to digital, encrypt it, and then convert it back to analog for transmission. With an all digital transmission system, the encrypted speech can be sent directly at substantially lower cost.

DARPA is participating in the development of the technology for end-to-end encryption of information on packet networks, including the capability to distribute encryption keys over the network. While the intended use is for computer-to-computer data communications, the extension of this capability to voice and images will probably be underway by 1985.

There will be a large number of packet communication networks in place in 1985, with a variety of uses and transmission characteristics. Technology is being developed for interconnecting these networks to achieve a high level of fault tolerance and survivability. Gateways and internetwork software protocols have already been demonstrated that allow communication among radio, satellite, and land wire networks. The advanced networking technology under development will provide smart gateways and adaptive routing algorithms which ensure that the highest priority traffic gets through if any path exists through a combination of land, air, and space circuits. This ability to dynamically reallocate communications capacity in a crisis and guarantee the ability of the National Command Authorities to communicate will be an important element of strategic deterrence and will have a major impact on all future military operations.

Signal Processing

Our ability to pack information on semiconductor substrates increases by a factor of 2 in each dimension every two to three years. That means, for example, that the capacity of memory chips is likely to increase by more than an order of magnitude, from the current 16,000 bits per chip to 256,000 bits per chip by 1985. Similar increases in the density of processing logic can be expected. New design technology will be needed to make effective use of that much logic on a single chip. Advancement in the processing of silicon and compound semiconductors is greatly facilitating the development of solid state power supplies and the development of very large scale integrated circuits for high throughput signal processing missions. The development of laser-switched electronic devices will greatly facilitate millimeter and submillimeter radar applications in space and will bring about a high level of sophistication in the integration of optical and electronic device technologies.

A number of new materials processing technologies are emerging which will bring about dramatic new capabilities in fabricating submicron-scale micro-circuits and in achieving more sophisticated two-dimensional and possibly three-dimensional circuit architectures. Among these are the following: (1) the deposition of semiconductor and insulator films on semi-insulating substrates; (2) direct circuit writing with directed ion beams of the required dopant species; (3) the use of focused beam processing to lay down semiconducting films, induce epitaxial regrowth, provide electronic activation, and anneal implantation damage at low substrate temperatures; and (4) the development of high-speed, high-registration techniques for multiple-beam lithography at several tenths of a micrometer level of resolution. The feasibility of fabricating hybrid chips containing both silicon and compound semiconductor circuits should be demonstrated within the next five years. The rapidly growing complexity of large-scale integrated circuits during this time period will require improved computer models for device design and fabrication process definition and control. Such models are now being introduced into practical application and are being actively improved.

Military Uses of Computers

The rapid decrease in the size, weight, energy consumption, and cost of computer processing power, in combination with the availability of economical and survivable packet communications techniques, will lead to a dramatic increase in the use of computers for national defense. Information processing costs are falling toward key cost thresholds below which the uses and management of computers will change drastically.

One threshold is the cost of storing paper files and microfilm. Currently, storage on computer processable media is relatively expensive. For example, it would probably cost between $100 and $1,000 to keep a typical book on-line for a year. The same book could be stored on an existing archival storage device for between $1 and $10. Optical disc technology currently under development shows promise of permitting that same book to be stored in computer processable form for a few pennies per year. Furthermore, optical disc technology is being developed for the home market, so it will not be necessary to have a large facility to exploit this technology. Optical disc readers should be available for a few hundred dollars, and once they are widely available, it will make economic sense to store all information in computer processable form.

A second major threshold will be crossed when a $25,000 computer can run programs written for large mainframe machines. Once microprocessors can support mainframe data management and networking capabilities, there will be a massive shift toward distributed processing. The importance of being able to run the same software on a personal computer and on a large and much faster mainframe computer cannot be overestimated.

The last threshold is the point at which control logic for a physical process can be implemented more cheaply with a microprocessor than with special purpose analog or digital circuits. This threshold has already been crossed for many applications, including temperature controllers for buildings, missile guidance systems, avionics control systems, and automobile fuel injection systems. Once the threshold is crossed, advances in computer technology are often used to increase performance rather than to reduce cost. Since microprocessors will be available in 1985 which can execute artificial intelligence algorithms for image processing, signal classification, and speech understanding, we can anticipate widespread use of heuristic algorithms which control devices and physical processes in much the same way that a person would.

Training and Expert Systems

The impact of having all information in computer processable form will be revolutionary because computers can take an active role in helping people to use information effectively. One result will be to make possible computer-based instruction systems capable of holding complicated conversations comparable to a Socratic dialogue about a subject. Computer-aided instruction systems developed to date have for the most part been simplistic and of use only in special situations. The major cause of these limitations has been a lack of adequate storage and processing power. A few successful artificial intelligence systems, however, provide a window into the future. The DENDRAL system for physical chemistry, MYCIN for infectious disease diagnosis, and MACYMA for calculus and other types of symbolic mathematics all demonstrate the ability to answer questions the same way an expert in the field would answer them. By 1985, there will be several instructional systems of this type in daily use, and numerous efforts will be underway to expand the range of topics covered and the depth of understanding possessed by the systems.

Other opportunities in the advanced training area include the capitalization on videodisc technology coupled with the establishment of a geographically dispersed instructional system. This will eliminate the need for soldier students to travel to the "school house" and enable us to bring the school house to the students. This development, feasible by 1985, coupled with the automated authoring and instructional materials systems will enable us to reduce training costs dramatically. Beyond this are opportunities in multimedia training technology which represent high-payoff initiatives in an already critical area.

Expert systems to help people perform specific tasks are much like instructional systems. In fact, the three systems listed above to show the feasibility of a Socratic teaching system were actually developed to be expert assistants rather than teachers. DOD has an enormous need for such computer tools to maintain its complex weapon systems in a hostile environment, and in the face of rapid personnel turnover. By 1985, DOD should have developed and demonstrated a knowledge representation scheme for aircraft maintenance data that a contractor can provide at reasonable cost and that is suitable for driving a training system for aircraft mechanics, providing a diagnostic aid for special problems, and printing any hard copy maintenance documentation required. Once this knowledge representation technology is demonstrated, it should be rapidly applied to a variety of other systems during the latter half of the decade.

The major problem caused by improved computer and communications systems will be the increased tendency toward centralization of all decisionmaking authority. There is little doubt that organizational policies and decisions can be strictly enforced and rapidly implemented using a highly integrated decision support system built on the information processing and computer communications capabilities we expect to have in 1985.

Cybernetics

Advanced decision technology in its broadest sense requires a number of technology options including the development of user-oriented computer-based problem structuring aids, options/actions prompters, and the large empirical data bases that should drive such aids. New insights in this field should come from exploring more fully the cognitive processes that determine individual and group decisionmaking behavior via content analyses of written and spoken communications, voice stress analyses, and other psychophysiological techniques. Opportunities exist in the biocybernetics area as well as in the area of artificial intelligence. On a more basic level, emphasis is needed first on understanding neural processes and then on modeling them for explanatory and predictive purposes. Progress can be expected in modeling decisionmaking expertise and correlating the effects of stress on cognitive decisionmaking performance.

Within the next five to seven years, progress should be made in calculating the probabilities of intra- and international crises based upon the development of computer-based indicator and warning systems and "intelligent" multitrack quantitative indicator systems consisting of economic, political, and military indicators. Manual processing of the data necessary to drive these systems will also give way in three to five years to self-generating data systems, capable of, in effect, feeding themselves. When we couple this capability with knowledge-based algorithmic capabilities, we will have generated an intelligent indicator and warning system capable of generating alert lists, country/situation profiles, and information with a minimum of manual input.

New technologies and opportunities in man-computer interaction include the development of spatial data base management systems, which will enable users to store, retrieve, and process information spatially and without keyboards; speech input and output systems; graphic input and output; and personalized user-computer interaction and modeling. Efforts will focus on developing commander-oriented information management systems which should be available by 1985. Opportunities in applied robotics (of all natures) and the biocybernetic operation of command and control information management systems also exist.

Lower Defense Costs

New emerging technologies will lower the cost of national defense over the next five years. New theories of wear, methods for detecting and analyzing the nature and source of wear particles, and surface treatments (such as surface alloying, plasma deposition, and ion implantation) for resisting wear are being developed. The formation of amorphous layers and highly homogenized, polycrystalline layers on surfaces by rapid solidification with lasers and electron beams is imparting greatly improved corrosion and oxidation resistance in structural alloys. New nondestructive techniques are being brought into practical application for providing quantitative flaw signature data which, when coupled with fracture mechanics, yield a rational basis for making accept/reject decisions and also provide an analytical means for predicting the remaining life in a component. The application of advanced sensors, microprocessors, and adaptive feedback to such manufacturing operations as machining, casting, and welding will provide a higher degree of self-optimization, increased product quality and uniformity, and improved overall productivity.

CONSTRAINTS IN THE USE OF NEW AND EXISTING SCIENTIFIC AND TECHNOLOGICAL CAPABILITIES

Nuclear Verification Methods

The primary constraint on verification methods for agreements limiting nuclear testing is that the methods must be as nonintrusive as possible and, generally speaking, must be effective at ranges of hundreds to thousands of kilometers (tens to hundreds of thousands of kilometers in space).

Information Processing

There are six major constraints on the rapid exploitation of information processing technology:

(1) There is a shortage of highly trained computer scientists, especially scientists with experience in artificial intelligence;

(2) The ability of people and organizations to adapt to change is questionable;

(3) The cost to implement software grows exponentially for very large systems; whether the logic is loaded onto a single chip or a large mainframe is irrelevant—all that matters is the number and complexity of decision elements;

(4) The dependability of existing computer systems does not justify our reliance on them; the technology for designing error resistant software that can recognize and compensate for hardware failures is primitive at best;

(5) Proven and widely accepted communications interface standards are needed to realize the potential benefits of interconnecting packet communications networks; and

(6) Communications bandwidth is becoming a scarce resource; existing frequency allocations make relatively poor utilization of available communications capacity, and new policies and allocations are needed to facilitate the rapid introduction of packet communication technology.

Cybernetics

There is a prevailing lack of appreciation and understanding of the impact that nonphysical science approaches can have on defense problems. It is still the practice to overlook human factors and the problems that can arise from operational complexity when allocating funding for new weapon systems. Training research and development receive far less than one percent of the overall training budget. There is a shortage of interdisciplinarians seeking to seize upon the opportunities inherent in the field of cybernetics.

Lower Defense Costs

National emphasis on new scientific and technical concepts to simplify design, increase productivity in manufacturing, and extend the life of high-cost hardware is inadequate to overcome escalating defense costs. There is insufficient "eliteness" attached to these fields by top universities and government funding agencies to attract top investigators and students. Public policies do not yet provide adequate incentives to assure the conservation of critical materials, and procurement practices seldom specify warranties based on total life-cycle costs to reduce operating, maintenance, and component replacement costs. Holistic approaches are needed to more effectively exploit advances in materials processing, cybernetics, nondestructive evaluation, fracture mechanics, corrosion and wear control, and computer-aided design, process control, and manufacturing.

Department of Defense
Department of the Air Force

The science and technology program of the Air Force provides the foundation for technological superiority of military weapon systems and capabilities vis-a-vis our potential adversaries and, thereby, is a major factor in the maintenance of our national security. The Air Force technology base program is the source (1) of ideas and innovation that lead to new and advanced weapon systems, (2) of improvement in existing systems, and (3) for the demonstration of equipment and subsystems for integration into coherent systems.

In satisfying those three major objectives, the Air Force conducts basic research, and exploratory and advanced development in seven major defense-related technologies: propulsion and power, materials, aerospace vehicles, weaponry, electronics, geophysics (environment), and life sciences. The Air Force uses its inhouse laboratories, the industrial R&D organizations, and the academic community to accomplish these objectives. Each plays an important and unique role. The Air Force labs provide essential coupling between participants and maintain a spectrum of technological skills essential to the coupling process. Industry contributes additional innovation and production know-how. Universities provide the primary source of knowledge for long-term scientific and technological advances.

Making sure that the Air Force has the right technology at the right time is a continuous, complex process involving inputs and reviews from many organizational levels of the Air Force, the sister services, the Department of Defense (DOD), other Federal agencies, the private sector, and Congress. Therefore, a great deal of emphasis is placed on research and technology planning by the Air Force and DOD.

The quantitative and qualitative gains of the military capability of the Soviet Union in the last 15 years, relative to ours, make it imperative that Air Force science and technology investments lead to the development of superior military equipment and systems. In addition, the growing rate, breadth and complexity of technology, coupled with the spiraling cost of force modernization and the declining real dollar budget, place even greater pressure on top-level Air Force managers to make intelligent planning and programming decisions. To better accomplish this task, the Air Force Systems Command is developing a major planning initiative called Project Vanguard. Vanguard is intended to provide an integrated planning continuum across the entire spectrum of Air Force research and development activities.

A brief overview of the near- and far-term science and technology outlook for the Air Force is given in the following text.

NEAR-TERM RESEARCH THRUSTS

The Air Force basic research program covers a broad range of scientific disciplines. Some significant projects are listed below.

Charged Particle Beams

Current research is concerned with obtaining a basic understanding of the mechanisms of production and propagation of intense charged particle beams. (Particle, as used here, means electron, proton, and neutral hydrogen atom.) Expected payoffs include high-intensity laboratory radiation sources to simulate nuclear weapon effects, directed energy weapons, and high-power electromagnetic energy sources.

Long-Life Rechargeable Battery System

The present research is a high-risk and high-potential payoff program which, if successful, will yield a long-life, low-cost, secondary (rechargeable) battery for applications such as aircraft and satellite station use.

New Mathematical Techniques Related to Logistics Problems

This research is investigating reliability and maintainability statistics of systems under actual operating environments, repairable systems, and degradable systems. The potential payoff is in the improved efficiency of Air Force operations and maintenance.

Microwave Tube Science for High-Power Applications

In a combined effort, the Air Force, Stanford University, and the tube manufacturing industry will provide (through a graduate program begun at Stanford in 1977) new ideas and research for microwave tube technology. Microwave tube scientists are a welcome by-product of this program. The Air Force still needs microwave tubes and scientists because solid state devices cannot generate the power levels needed in high-power applications such as radars, communication devices, and electronic countermeasures.

Nondestructive Evaluation (NDE)

Investigation of NDE methods includes research on advanced ultrasonic techniques for flaw detection and for studying intrinsic material properties such as state of stress. Potential payoffs include improved structural integrity and reduced maintenance costs.

Joining Science

This program is concerned with investigating adhesively bonded material systems with a view to improving the mechanical properties, and hence the service life, of components and subsystems.

Damage and Failure Process in Composites

Causes of failure in structural composites, including environmental effects on composite fracture, are being investigated. Potential payoffs include predictable behavior and reduced maintenance and repair costs.

Gravity Models

This research will involve analyzing satellite altimetry data to derive new knowledge of shape and gravity fields over open ocean areas. This will lead to improved gravity models needed for increased missile accuracy.

Size, Shape, and Distribution of Cloud Particulates

Ice crystals, droplets, and snowflakes in clouds are being measured and analyzed. This information is necessary for the design of optical systems and for understanding how cloud species erode missile parts. A potential payoff will be in the improved design of missile nose cones.

Spacecraft Charging

This research is concerned with obtaining a fundamental understanding of the electrical environment in which spacecraft operate and the interaction of the environment with the spacecraft. This interaction can, for example, activate circuits, burn out components, and lead to loss of communication. Research on spacecraft charging could pay off in improved reliability of space systems.

Toxicological Hazards

The main thrust in this area is to find early indicators of the toxic effects of materials of interest to the Air Force such as the hydrogen chloride gas that is released when the Titan III missile is launched, and the hydrazine that is used in the F-16 emergency power unit. Mechanisms of toxic actions and biochemical measures to counter them need to be studied. This research will lead to the development of reliable standards for toxic materials, including permissible tolerance limits.

NEAR-TERM TECHNOLOGY THRUSTS

Brief synopses of the major near-term technology thrusts in each of the seven defense-related technology areas are provided in the following text.

PROPULSION AND POWER

Included in this major technology area are air breathing propulsion for aircraft and missiles, rocket propulsion for spacecraft, air launched and ballistic missiles, aerospace power, aircraft fuels, missile propellants, turbine engine lubricants, and aircraft fire protection.

Turbine Engines

The turbine engine development work sponsored under this program provides the world's best turbine engine technology for U.S. military aircraft. By coordinating with the Army and Navy research and exploratory development programs, the best component technologies are united in jointly funded technology demonstrators which become the preprototypes for future engineering development programs. The emphasis over the last several years has been performance. While we are maintaining a strong performance program, existing contracts and new starts are being tailored to stress durability.

Spacecraft Power

In spacecraft power, great strides are being made in developing lighter, more efficient, and more cost-effective batteries and solar cells. Aircraft auxiliary power unit volumetric power densities are now capable of providing fighter aircraft self-sufficiency during ground operations with the eventual goal of in-flight, all altitude capabilities.

Rocket Propulsion

The rocket propulsion effort for air launched missiles is directed toward reducing production and life-cycle costs and providing significant improvements in performance in range and maneuverability.

Ballistic Missiles

Space launch vehicle payload capability will be increased up to 20 percent through the incorporation of ballistic motor technologies into satellite propulsion systems. The ballistic missile propulsion effort is devoted to demonstrating technology to improve performance, reduce engineering development risk, and reduce the life-cycle cost of future ballistic missiles.

Fuels

The Air Force is starting to spend significant sums of money in studying the changing character of the petroleum-based aviation fuels and the emerging synthetic feed stocks such as shale oil, tar sands, and coal. The billions of dollars in existing Air Force turbine engine inventory demand that the Air Force maintain its position as a knowledgeable consumer by strict yet cost-effective control of fuel specifications.

MATERIALS

Materials technology provides the foundation for advances in performance, reliability, and durability of Air Force weapon systems. Acquisition and ownership costs are also strongly influenced by materials parameters.

Exploratory development resources for materials remained approximately even with inflation from fiscal year 1978 to fiscal year 1979, and 3 percent real growth is proposed from fiscal year 1979 to fiscal year 1980. These resources support efforts directed toward advances in new, more durable materials for aircraft structural applications; new elastomeric materials for seals, sealants, and lower cost-of-processing resins for composites; more durable high-temperature materials for long-life gas turbine engine components; erosion resistant thermal protection materials for re-entry vehicles and rocket nozzles; materials for missile radomes; and detectors for surveillance satellites. This materials technology contributes to new Air Force capabilities, and can also be integrated into existing systems through incorporation into maintenance and repair activities. Materials technology addresses current Air Force problems in areas such as corrosion control, nondestructive inspection, and failure analysis.

Advanced technology demonstration resources, which combine both materials and structures technology, have been reduced substantially over the fiscal years 1977-1980 due to overall funding constraints in this budget category. Demonstration programs have been maintained for missile and space applications of composites, laser-hardened materials for satellites, and low-cost titanium fabrication for airframe structure. Effort has been reduced on application of advanced composites and laser-hardened materials for aircraft, and work will not be continued on improved windshield materials/designs. Demonstration of these and many other promising exploratory developments as options for future systems will therefore be indefinitely delayed. The new materials thrust in the area of metal matrix composites will see moderate funding in fiscal year 1980, with significantly increased activity planned for fiscal year 1981.

FLIGHT VEHICLES

The major objective of this technical area is to provide the flight vehicle technology required for the design and development of future aerospace systems and for improvement of current systems. It encompasses the technologies of structures and dynamics, vehicle equipment, flight control, and aeromechanics.

Current emphasis includes programs on airframe designs using advanced structural materials and concepts, design criteria for hardening aircraft against ballistic combat threats and atmospheric electrical hazards, and improvement in the quality and serviceability of aircraft landing gear, brakes, and tires to reduce logistical and maintenance costs. Work will also be done on efficient cryogenic cooling for aeronautic and space applications, innovative flight-control concepts, simplified cockpit and displays for better man-machine interface, validation of new aircraft and missile configurations by wind tunnel tests, evaluation of emerging technologies to determine their impact on system effectiveness versus development and operating costs, and expanding the capability of ground test facilities to minimize costly flight testing of aircraft and missiles.

Major considerations in all flight vehicle technology efforts are improvement in performance, reliability, and survivability; reduction in life-cycle costs of weapon systems; and energy conversion.

WEAPONRY

Included in the weaponry area are conventional munitions and new, advanced weapon concepts such as lasers and particle beams.

Conventional Munitions

The conventional munitions program will continue to emphasize anti-armor efforts such as self-forging fragment projected shape charges, and millimeter wave (MMW) and infrared target signature and background clutter measurements. It will also generate methodology and vulnerability models for evaluating current and conceptual weapons. In addition, technologies that support development of a future short range, air-to-air missile to be used when the tactical situation warrants replacement of the AIM-9 missile will be continued.

Lasers

The Air Force high-energy laser program will continue to examine the feasibility of developing laser weapon systems that are capable of effectively engaging and destroying selected targets. During fiscal year 1979, the Airborne Laser Laboratory will enter into a series of technology demonstrations to investigate not only the integration and operation of high-energy laser components in a dynamic airborne environment, but also the propagation of laser light from an airborne vehicle to an airborne target. Emphasis will be on new laser devices, beam control, and pointing and tracking technologies.

Nonconventional Weapons

Included under advanced weapon concepts are both nuclear and nonconventional weapon technology. The joint Department of Energy/U.S. Air Force nuclear weapon phase studies are to improve the state of the art and ensure military effectiveness of Air Force nuclear weapons. Advanced nonconventional weapon studies are conducted to identify new and emerging technologies, such as particle beam technology, to determine if they have potential for development as weapon systems.

ELECTRONICS

For the past twenty years, electronics has been the most rapidly changing sector of technology. This rapidly evolving field has become a force multiplier since it enables the Air Force to do much more with one aircrew and aircraft than could be done with several aircraft in the past. Advances in electronic components and techniques should provide even greater capabilities at an affordable cost in the future. Advances in command, control, and communications will result in all-digital systems which emphasize survivability and interoperability with the other armed services and NATO.

Computer Technology

In computer technology, software costs are being limited by standardizing the use of only two controlled computer languages. A tri-service program is underway to take the next step forward in high-speed integrated circuit technology. This program will give us circuit speed and circuit complexity, will produce circuits qualified in terms of reliability, radiation hardness, and temperature range, and should keep us ahead of our adversaries in computational power and signal processing capabilities. Major efforts in electronic warfare (EW) technology should provide the capability to negate, disrupt, delay, and deceive threat radar communication and electro-optical systems while at the same time developing EW-resistant equipment to allow us to operate in an EW environment.

ENVIRONMENT/GEOPHYSICS

The need to locate satellites accurately, communicate reliably, and pick up targets at greater ranges makes it necessary for us to better understand the environment. In many cases the environmental effects on surveillance, communication, and detection systems are limiting factors in system performance. It is, therefore, necessary to know more about the environment to assist system designers and operational decisionmakers. There is also a need to develop technology for environmental quality and alternate energy sources to comply with Federal, state, and local environmental standards and regulations.

SCATHA and OPAQUE

A Spacecraft Charging at High Altitude (SCATHA) satellite was launched to measure the spacecraft charging environment at geosynchronous altitude. Intensive studies of solar activity as the solar maximum approaches are being conducted to develop solar activity prediction techniques for use in support of communication systems. The Optical Atmospheric Quantities in Europe (OPAQUE) measurement program, which documents the atmospheric transmission and scattering properties in Europe, will be completed. Stratospheric environmental impact assessment models will be enlarged to include additional pollution reactions and used to assess the impact of planned Air Force conversion to JP-8 fuel. Refuse Derived Fuel (RDF) and oil cofiring feasibility studies will be initiated.

During fiscal year 1980, the reduction and analysis of the SCATHA and OPAQUE data will continue, as will the solar activity studies. The RDF and oil confiring studies will be completed.

PEOPLE

The all-volunteer environment has placed new challenges before the Air Force. Procurement of quality personnel is the cornerstone. These personnel must be properly trained and utilized. A highly productive force that meets the demands for maintaining our sophisticated weapon systems and the personnel, manpower, and logistics systems that both strengthen and support our day-to-day operation are a must.

Job Placement and Training

By the end of fiscal year 1979, improvements in the effort to match recruits to jobs are expected. Better measures of aptitude and attitude are the key. Efforts are underway to expand our knowledge of the use and design of computer-based instruction systems. Plans call for the improvement and evaluation of methods to apply people-related data very early in the design of weapon systems and to determine their effect on operational and support costs.

In fiscal year 1980, data should be available to support the use and design of simulators for electronic-related maintenance training. Model development is underway to simulate response characteristics of the personnel system as it responds to policy changes in a dynamic environment. Continued emphasis on the development of flight training methods and media specifically directed at transition and continuation flying is planned.

Similar people-related research and development will continue into fiscal year 1981. The technology spin-off to the user is dynamic. The users include all major Commands, the Air Staff, and separate operating agencies.

Biomedical and Health Programs

The Air Force people program also includes extremely important technology efforts on health-related subjects through its biomedical programs. The biomedical mission is to enhance man's capabilities to function safely and effectively as an integral part of Air Force systems and operations. One area of emphasis for fiscal year 1980 will be the chronic effects of low-level radio frequency radiation, with research designed to answer questions concerning both public and occupational exposure from Air Force radars. The relative biologic consequences of pulsed versus continuous wave radiation will also be studied.

Another area of emphasis is predictive toxicology, which must be developed to enable reasonable risk versus benefit tradeoffs in employing new chemicals on a large scale. The methodologies under development are pertinent to occupational and public health and safety standards and developing environmentally safe limits for other forms of life.

The Health Evaluation and Risk Tabulation (HEART) program is designed to develop and demonstrate an integrated biomedical system to identify and reduce the risk of preventable cardiovascular disease in Air Force personnel. With the $1.9 million requested for fiscal year 1980, the program will complete the development of an operations research model that will allow cost versus benefit analyses of risk intervention at varying levels of risk probability. Plans for fiscal year 1981 include the evaluation of prototype systems.

LONG-TERM TECHNOLOGY ISSUES

The increasing concern over the diminishing U.S. technological advantage is reflected by the number of long-range studies conducted by the Air Force in the last three years. All of the studies have attempted to identify specific areas of science and technology that have the potential to significantly affect the way the Air Force accomplishes its mission.

The projection of these important technology-driven trends can aid Air Force strategists in answering critical R&D related issues. Some of the penetrating ones are:

(1) What percentage of the Air Force budget should be assigned to R&D versus operations and support?

(2) What is the appropriate tradeoff between quality and quantity in future, high-cost systems?

(3) Can we achieve a level of survivability in our space-based systems commensurate with our increasing reliance on them?

(4) How should the Air Force allocate its R&D budget among basic versus applied research, offensive versus defensive systems, tactical versus strategic systems, and high-risk, high-payoff programs versus low-risk, medium pay-off programs?

Undoubtedly, certain technologies, such as the microelectronics computer technology and sensor developments will have major impacts on all the Air Force system capabilities, from improved intelligence, surveillance, and reconnaissance, to greatly improved weapon delivery capabilities in both tactical and strategic systems.

Directed energy beam technology offers a significant potential for offensive and defensive weapons, communications, and target detection capabilities. New advances in propulsion, power sources, fuels, and materials offer exciting possibilities across a spectrum of systems. The fundamental issue is, "Can we develop and exploit these technological opportunities in a timely manner to maintain our deterrence posture? Can our basic research programs provide new technology breakthroughs to prevent technological surprise and a threatening imbalance in military capabilities in the future?"

Successful answers to these questions depend not only on better, wiser planning and programming decisions by the Air Force and DOD management, but on the resolution of some larger issues outside the control of DOD.

Department of Defense
Department of the Army

U.S. Army science and technology goals in the next five years have been planned to emphasize the enhancement of national security by improving ground combat capability. Key problem areas being addressed include the following: armor development and counterarmor capability; improvement of command, control, communications, and intelligence; enhancement of the Army's ability to conduct combat operations under adverse atmospheric and environmental conditions; improvement of battlefield mobility and firepower; and improvement of the survivability of the combat force against all threats. Further, inasmuch as Army resources will be severely constrained in the next five years, it is imperative that the cost effectiveness of Army personnel utilization be optimized by reducing lost time, reducing the need for retraining, and enhancing the effectiveness of a soldier's initial training. Scientific and technological opportunities exist in all of these areas.

ARMOR

In armor development and counterarmor capability, technological advances are required in new material fabrication, analysis of penetration mechanics, and development of superior penetrator materials and techniques.

COMMUNICATIONS

The improvement of command, control, communications, and intelligence is a key multiplier of combat force effectiveness. Dramatic technological advances are necessary to improve the ability to handle and process increasingly large volumes of data and communications circuits in the harsh environment of jamming, communication disruption, and electromagnetic interference.

Very High Speed Integrated (VHSI) Circuits

VHSI circuits may overcome significant shortcomings in data processing. The VHSI program is a 6-year, tri-service/industry development program to provide a giant step forward in integrated circuit performance and production capabilities for future military needs. VHSI is aimed at establishing a new plateau of performance that will be orders of magnitude above what can be accomplished with today's large scale and very large scale integrated technology. A tenfold reduction in size, weight, power consumption, and failure rate, with accompanying savings in both initial and life-cycle costs of military computer processing systems, is envisioned. The requirement is for small subassemblies to provide 100 times the processing throughput of present integrated circuits to perform an almost unlimited amount of real-time data processing within small mobile weapon systems, fire-and-forget missiles, remotely piloted vehicles, and mobile analysis centers. New or improved chip architecture will be developed to permit chip design at an affordable cost and minimize customization to reduce supply and logistic costs. The program will provide for the commercial availability of submicron lithography equipment and an industrial capability for producing VHSI circuits. The program is aimed at accelerating the advancement of microcircuit technology by at least five years to firmly reestablish our national leadership in this rapidly changing field. It will also assure that the military will have the industrial capability and sources for advanced military quality devices and subsystems for the next generation of computers, missiles, radars, and intelligence processing centers.

Radio Spectrum Management

Deliberate jamming and frequency spectrum conservation that cause radio interference present us with two contradictory requirements. Spread spectrum and frequency hopping techniques are applied to overcome interference due to jamming and to discourage interception. The present state of the art in these new techniques lacks an experience backlog regarding electromagnetic compatibility with other systems. The already crowded FM spectrum dictates that we develop new capabilities to optimize spectrum usage and efficiency.

Low-Cost Optical Fibers

Low-cost fiber optic communications will undoubtedly replace much of the present wire and cable systems. National security interests require consideration of survivability; optical fiber purification techniques will overcome the vulnerabilities of communication systems to nuclear radiation.

Management Information Systems

Modern organizations, including the Army, are extremely large and complex structures that severely tax traditional concepts of management and leadership and strain organizational cohesion. The concurrent information explosion and increasing speed of information transfer add to this strain. The use of a multidisciplinary approach to integrate management information systems and information processing concepts, combining with organization effectiveness concepts, cognitive theory, and concepts of artificial intelligence, will improve communication and control in large, complex organizational structures.

COMBAT UNDER ADVERSE ATMOSPHERIC AND ENVIRONMENTAL CONDITIONS

To enhance the Army's ability to conduct combat operations under adverse atmospheric and environmental conditions, significant scientific and technological advances must be made. Current knowledge of the characteristics and behavior of microscale atmospheric conditions is inadequate to empirically determine the propagation of electro-optical energy for military systems.

One of the most promising opportunities for operating successfully in an obscured environment lies in exploiting the general portion of the electromagnetic spectrum lying between 100 and 1000 GHz near-millimeter-wave (NMMW) technology. This portion of the spectrum has general characteristics of the very-far-infrared optical waves and also of the extremely high-frequency microwaves. It penetrates obscurants reasonably well and offers acceptable resolution. The technology base is very tenuous; a major effort is required to develop sources and detectors, and to provide knowledge of propagation characteristics.

BATTLEFIELD MOBILITY AND FIREPOWER

Battlefield mobility and improved firepower also require technological advances. We need significant improvements in energy-efficient means for mobility. The economic and national security implications of the petroleum shortage require near-term mitigation. Gasohol fuel and engine technology, methane fuel and small engine technology, and development of alternative methane sources, particularly by means of low-risk technologies, are easily transferable to all sectors of the economy. Expedited development of engine technologies that are workable alternatives to internal combustion is an urgent requirement.

Improvements must be made in the use of new materials that are lightweight and provide cost-effective alternatives to critical metals. The delivery of firepower requires cost-effective solutions to problems of wear and erosion of gun tubes as well as the development of effective propellants and explosives with low vulnerability.

Rechargeable Batteries

High density rechargeable batteries having good cycle life need to be developed to replace the heavier nickel-cadmium and lead-acid batteries presently used in Army aircraft and ground combat vehicles respectively. The weight of present batteries is excessive, and they take up valuable space that could be used to increase combat operational capabilities. In aircraft, the nickel-zinc rechargeable battery will be capable of providing the equivalent capacity with one-half the size and weight of a nickel-cadmium or lead-acid battery and will provide for cranking and emergency power operation down to minus 40°C. In ground combat vehicles, the reduced size and weight will alleviate overcrowding and provide for longer silent watch operation. With a renewed, sustained effort in cell design and the testing of cells containing the improved components and battery controls, it should be possible to develop an inexpensive nickel-zinc power source with a useful life of 800-1000 cycles, an energy density of 40 watt-hours/pound, and a cost of six cents per watt-hour in kilowatt-hour package designs.

Combat Simulation Models

The U.S. Army has been using "engagement simulation" as a way of measuring and maintaining increases in combat readiness of small combat units. When incorporated within performance-oriented unit training models, realistic engagement simulation exercises of increasing tactical complexity possess a unique capacity to promote learning. Such learning facilitation is particularly successful with use of the "After Action Review" in which soldiers themselves describe how they were able to engage an enemy soldier, destroy a target, or how they were "killed" themselves. They thus reinforce and extend what they learned during the engagement.

This development and its subsequent extension should be viewed against a background of what used to be the case. Historically, in Army tactical training exercises in the field, the units met "aggressors" in a manner that almost totally lacked tactical realism. As a result, combat arms units could obtain effective training in only a limited number of tactical techniques; they gained little if any proficiency in making adaptive decisions necessary in the face of an active opposing force. Tactical training and evaluation were particularly hampered by the lack of a realistic credible method for simulating battlefield casualties. The development of low-cost techniques for the simulation of weapon effects (casualties) and weapon signatures provided the breakthrough necessary to simulate the battlefield in a credible fashion. Casualty assessment techniques were developed for the range of tactical weapons available to infantry and armor units; these techniques include simulation of the effects of the M-16 rifle in the hands of the infantryman as well as the effects of tank armament, anti-tank weapons, and indirect fire. The Army has found that when engagement simulation techniques are used in training small combat arms units, extremely effective and motivating tactical training can be achieved to a degree not previously realized. The learning of individual and group tactical skills is enhanced, and troop motivation and interest increases significantly. Moreover, the method is useful for diagnosing training needs and evaluating training, doctrine, weapons, and total combat readiness.

The Army also sees the use of videodisc, holographic, and microprocessor technology to provide display and control of simulated "real world" spatial environments (such as battlefield views and land navigation experience for inaccessible areas) as a further enhancement of engagement simulation techniques. Applications of engagement simulation to urban warfare and to nonmilitary areas, including law enforcement training and possibly sports training, are seen as possible.

HEALTH AND SURVIVAL OF THE COMBAT FORCE

Research applicable to dealing with military disease hazards is currently undergoing a series of dramatic advances in a trend likely to continue throughout the foreseeable future.

Disease Control

Progress in immunobiology is spectacular and is opening new avenues for rapidly identifying microorganisms, understanding disease mechanisms, and developing preventive measures. Over the next five years, concentrated efforts will be directed toward developing new concepts for immune enhancers and the potential for in vitro synthesis of immune globulins. The ability to propagate microorganisms and tissue culture cells in large quantities continues to improve rapidly as do the basic techniques for isolating, purifying, and characterizing components of organisms and their toxic products. Utilizing the new technology of molecular biology, trends indicate the potential for synthesis of artificial gene products which may be used to develop vaccines, or the use of surrogate compounds which mimic the structure of microbial antigens.

Genetic manipulations of organisms and their plasmids and nucleic acids through reassortment or recombinant techniques are opening exciting new fields of knowledge. At the same time, these new tools introduce a potential threat from genetically restructured organisms with de novo changes in their virulence, toxigenic capacity, antibiotic resistance, and range of host infectivity. The recent appearance of infectious diseases such as swine influenza, Legionnaires' disease, and Ebola fever suggests that additional, previously unrecognized diseases may emerge during the next several years.

Long-recognized disease threats may reemerge in areas of the world that experience a breakdown of public health practices. To counter these threats, emphasis will continue to be placed on the development of more effective vaccines and drugs. Methods such as liposome delivery systems will be utilized for safe and efficient distribution of drugs to appropriate target cells. A wide variety of adjuvants will be evaluated for their ability to accelerate the immune response and intensify the efficacy of vaccines and toxoids.

The resurgence of vector-borne diseases currently observed in many developing countries is being attributed to the growing resistance of many disease vectors (especially mosquitoes) to available pesticides. This resistance coupled with economic and environmental considerations, dictates the need for directed research over the next five years to evaluate and develop integrated methods of control for target pests, emphasizing nonchemical measures.

Biological Research

Ongoing and anticipated technological advances in microbiology increase the level of threat from aggressor nations that might elect to engage in offensive biological warfare. To defend against this possibility, we must continue to maintain a broad and ever-advancing technological base. Military medical research should continue to emphasize the development and production of safe and effective immunizing agents against a growing list of organisms, including those with extreme virulence. Other research efforts will attempt to minimize the threat by improving prophylactic and therapeutic methods and by accelerating diagnostic capabilities. Emphasis in the next five years will include applications of research technology for use in the field. All efforts in military disease hazards technology will continue to be coordinated with the work of other military services, Federal agencies, and international organizations.

Emergency Care

Combat casualty care research is currently directed toward the development of a comprehensive field medical system. Capabilities for far-forward resuscitation and rapid return to duty are emphasized. In addition, the medical system must be capable of operating in environments containing chemical and nuclear hazards. This research addresses two problems of national significance which are dependent upon scientific and technological solutions: clinical management of shock and trauma patients utilizing minimal resources, and adequate protection in nonconventional environments. Specific studies will utilize recent findings in immunology to determine the contributions to sepsis of altered immune response. Blood research based on recent accomplishments will be multifaceted. Having completed purification of stroma-free hemoglobin, a potential blood substitute, researchers will emphasize improvement of oxygen offloading in tissues and prolongation of intravascular half-life. The shelf life of banked blood has been improved by modifying the storage medium; follow-on studies are expected to further increase useful shelf life. Efforts will also be directed toward cryopreservation of platelets for immediate field use. Isolated perfused organ models will be used to evaluate resuscitative fluids.

A major new trend in shock and trauma research will be defining wounding mechanisms of improved conventional munitions to include inhalation and blast injury. Studies to delineate the effects of anesthetic agents on physiologic sequelae of major trauma will continue. Advances in knowledge of neurophysiology and neurochemistry are required to fully define the mechanisms of action of anticholinesterase agents and candidate antidotes. Development of appropriate animal models is required for advancement of electro-anesthesia techniques. Although we are constrained by lack of knowledge of subcellular effects of ionizing radiation, some progress is expected in the reformulation and development of prophylactic drugs to minimize the effects of such radiation. The effective dose of any new prophylactic or therapeutic drug is expected to be large, resulting in a problem of delivery to minimize harmful side effects. Liposome drug delivery technology, developed in a recent parasitic disease research, will be applied to potentially reduce total body exposure and harmful drug side effects. All combat casualty care research, especially in shock and trauma, will continue to be conducted in cooperation with civilian organizations. Worldwide sources will be screened for information potentially applicable to U.S. requirements.

Health Hazard Prevention

Research in systems health hazard prevention is driven by actual and contemplated future battlefield requirements. As a result of technological advances, increasing complexity of weapon systems, and development of new tactics, doctrine, and training requirements, individual soldiers and crews are exposed to a variety of potential hazards that may impair their health, exceed physical or emotional tolerance limits, or degrade their performance capability. These new developments have prompted a dramatic shift in our research efforts so as to identify and minimize the adverse impacts of weapon systems, materiel, and training or working environments on health and performance effectiveness. This emphasis in our research will continue in the foreseeable future and will expand to meet new medical requirements that stem from newly discovered health hazards or the need to comply with Federal, state, or local health regulations.

A prominent thrust in our efforts is directed at development of health hazard assessment and prevention technology for the toxic effects of the "Dirty Battlefield" and other military environments. Many new military-unique compounds pose health risks in the processes of manufacture, storage, transport, handling, field use, and exposure to combustion products or contaminants. New weapon systems require complete assessment of the crew compartment chemical environment. Presently configured combat vehicle crew compartments and individual protective clothing ensembles may pose substantial risks of heat exhaustion and heat stroke to the user. New air and ground combat vehicles require assessment of the medical and performance effects of vibration, noise, and impact. Low-frequency noise has recently been associated with high-frequency hearing loss. Pulmonary, auditory, and other effects of blast overpressure require further study. Chronic low-level laser exposure has been associated with irreversible vision impairment. Further assessment of laser hazards and development of protective devices are urgently required if low-level laser emitters are to be employed in training. An assessment of high-powered microwave sources to which personnel are exposed during training and field operations is also required. Recent medical concerns have focused on the genetic, cancer-producing, and ocular effects of microwaves.

Current manpower shortages and future battlefield conditions will place the soldier under more stress than ever before. Measures to decrease stress, prevent psychiatric casualties, and enhance performance in the face of lifethreatening circumstances and extended performance requirements need to be developed. The impact of other factors potentially affecting our readiness posture must continually be assessed. These include the adverse effect of drug and alcohol abuse on readiness and the high first-term enlistment attrition rate. Increasingly heavy physical demands on fitness have highlighted the need to establish medically based realistic physical fitness standards and physical training programs to maintain combat readiness.

Military Dentistry

Environmental and toxicological factors associated with the practice of military dentistry are a problem of national significance and need to be studied thoroughly. The high incidence of maxillofacial injury in high-intensity combat, oral diseases with their trend of increasing treatment costs, and the shortage of adequately trained combat troop replacements and professional personnel early in war dictate research targeted toward time/treatment/cost goals to return troops to duty rapidly through improved professional care and disease treatment. Emphasis in the next five years will be placed on utilization of biodegradable and biocompatible materials in combat avulsive wounds, the continuation of CO_2 laser studies in maxillofacial wound treatment, and studies to determine tissue damage caused by various-sized missiles of different velocities and diverse wound pathways. Electrical analgesia techniques will be evaluated for effectiveness, rapidity, and simplicity. Salivary test methods for chemical agent detection during combat will be refined.

FIRE- AND FLAME-RESISTANT FABRICS

The combustion of organic polymeric materials presents serious fire, smoke, and toxic vapor hazards. The growing use of these materials in all segments of society, as well as in the military forces, results in a heightened concern for the safety of human life and property. As people live in more congested conditions, the possibility for fire catastrophies is compounded. The losses in the United States annually due to fires are in the billions of dollars. We cannot count on the emergence of any new industry-produced, inherently flame-resistant fibers. Reliance on conventional polymers rendered fire-resistant by the use of additives or by chemical modification of the polymers themselves must be continued. Research needs to be done to better understand the dynamics of fire, smoke, and toxic vapor production. We also need to improve fire pit testing and to develop laboratory methods that replicate fire situations. Specifically, the technology is available to pursue vigorously the use of the phenomenon of intumescence in textiles. Intumescence is the use of specific chemicals to induce the foam-char decomposition that has been demonstrated to significantly protect the fabric-supporting matrix. More complex, but technologically feasible, is the incorporation of hollow fiber technology and intumescence. In this case, the intumescent chemicals would be incorporated within the fiber, providing flame protection in a fire situation but not affecting the comfort properties of the fabric as do today's flame-resistant textile additives.

STRUCTURES

Several recent catastrophic failures of dams have emphasized the urgent need for precise techniques for monitoring structural movements or deformations of dams and other large structures. The need for monitoring tens of thousands of publicly and privately owned dams throughout the United States, many where sizable populations live in the downstream flood plain, poses problems of national significance. Required is the cost-effective development of continuous monitoring techniques and equipment with sufficient precision and reliability.

Department of Defense
Department of the Navy

Development and maintenance of naval power to preserve national security constitutes the paramount mission of the U.S. Department of the Navy research program. Management of this research program is based on introducing innovations in science and technology into Navy practice and equipment. This policy is implemented by participation in scientific forums and by funding of basic research in areas that are applicable to naval operations. Navy's research managers are also alert to the nonmilitary implications of emerging technologies and scientific advances. Thus, Navy research in genetic engineering to develop cells that produce food and energy has significant implications for the civilian economy.

Topics highlighting the Navy's science and technology objectives are described as they affect both the Navy and the nation. These deal with materials and structures, information and computers, military science, health, the environment and environmental protection, and manpower. For each topic, a brief discussion of problems, opportunities, and constraints is presented.

MATERIALS AND STRUCTURES

Deterioration of Materials

In many sectors of the economy, corrosion presents a significant economic burden. It is a particularly severe problem for the Navy and commercial interests existing in a salt water environment. It has been estimated that corrosion costs to the U.S. economy were around $70 billion in 1975, and costs to the Navy that year were estimated at $624 million. There is a subtle but important degree of dependence on scarce materials from foreign sources for some of the current techniques and methods that deal with corrosion. There is also a close interrelation with wear, which limits the life and the energy efficiency of machinery. Avoidable costs can be reduced by increasing the awareness of designers and builders to corrosion problems and known possible solutions, educating users about corrosion generally, and improving accuracy of specifications on systems and components. Currently unavoidable costs can be reduced through expanded and coordinated research and development on corrosion fundamentals, specifically on new materials and the effects of hostile environments. Surface modification through ion implantation appears to be a promising technique for reducing both corrosion and wear.

Failure of Structures

The dynamics of the failure of structures, especially ships, subjected to forces that vary in strength, direction, and frequency, are not clearly understood. The process of gradual degradation has received little attention. In the presence of storm or battle damage, there is little understanding of why one structure is merely degraded while another fails catastrophically. Although work has been done to model the sensitivity of a water-borne platform (ship) to damage from various causes, the instance of gradual degradation of capabilities has, heretofore, been treated as a boundary condition in such models. There is a need for research and analysis that will permit models to be upgraded to include or to predict the incidence and course of gradual degradation in performance. Of particular interest to the Navy is the ability to predict the effect of multiple, simultaneous, or sequential casualties on platform and system performance.

INFORMATION AND COMPUTERS

Information in Decisionmaking

The human ability to accept, assimilate, and act upon large quantities of information in anything approaching real time is severely limited. When the information comes from a number of diverse sources (sensors), is based upon different physical principles (radar, sonar, intercept, visual), is received in various forms (sight, sound, data readout) with varying accuracies, and may be transmitted simultaneously, the magnitude of the problem increases dramatically. Research is required to extend the work already accomplished in logic analysis. Work is required in all aspects of "artificial intelligence" or, as it might be termed, computer-aided decisionmaking. There is a need to apply advances in this field to the complex problems of situation assessment and decisionmaking. There are many situations in which the Navy could benefit from being able to assess large quantities of multiformatted information so as to be able to react appropriately and in real time. There continues to be a need for more and better algorithms—those capable of assisting in the real time analysis of uncorrelated data. At present, there are too few people trained in the area of logic, and the number of people knowledgeable in both human logic and mathematics is very small. The psychological acceptance of machine-aided decision recommendations must be addressed.

Computers

The continued blurring of boundaries between hardware, firmware, and software provides both new problems and new opportunities. Studies are needed to optimize the roles and interactions of components. In addition, consideration should be given to developing self-correcting hardware and software. Other areas of research include parallel processing and fault-tolerant technology.

MILITARY SCIENCE

Detection Equipment

Action against targets beyond the range of visual detection and identification of the targets has grown to be a significant problem which limits platform offensive capability. There are weapons whose useful ranges are, in fact, greater than can be utilized due to our inability to "see" and identify discrete targets. The problem is even more acute in detecting submarines, because of the volume and nature of the medium to be searched. Solutions will come from two principal directions; namely, evolution in existing technology and revolutionary progress in wholly new technologies currently at the frontiers of research and science. While considerable effort is expended in developing better radar, infrared systems, and the like, much more attention might be devoted to the investigation of basic phenomena affecting detection. The fact that we are able to characterize planetary bodies down to their atmospheric composition and atomic structure from enormous distances certainly provides some measure of confidence that adversary ships and aircraft might be similarly characterized with great precision. Developments in devices which use intense microwave energy in the gigawatt region or intense relativistic atomic particles might be brought to bear on the problem. The payoff is immense but so are the risks. Principal constraints are the translation of the problem to the basic science community and the rapid translation, communication, and exploitation of knowledge so that it can be properly applied.

Fast Turn-on, High-Power Devices

Fast turn-on, high-power devices could lead to a new concept in emitter control-capable systems. Gigabit logic and compact nonvolatile fast memories with large bit capabilities will meet the needs of future systems. A major advantage of gigabit logic and fast memories in electromagnetic support measure (ESM) equipment is the prospect of digitizing at the receiver front end. Areas for research include optical upconverters, solid state physics, superconductivity applications, and energy conversion/high-power microwave devices. Priorities need to be established in long-term, high-payoff research.

Passive Ranging

It is highly desirable to be able to obtain the range of threat emitters by passive means. The techniques developed to date are inaccurate or not useful. Opportunities for research exist in bistatic radar and methods/materials/

concepts to achieve the requirements of spatial, spectral responses in the full spectrum of radio frequency, infrared, and electro-optics. Very few new workable ideas have yet been developed in this area.

Coding

There is a lack of basic knowledge of frequency and phase codes suitable for sophisticated waveforms. The designer has available empirical knowledge of acceptable waveforms, but a unified understanding of this subject does not exist. An investigation of the needs imposed by countermeasure environments and of problems due to electromagnetic interference and spectral occupancy is warranted. The shortage of both money and people with knowledge of the problem is a major constraint.

HEALTH

Malaria

Because of the propensity of the causative parasites to develop resistance to chemoprophylactic and chemotherapeutic drugs, malaria continues to pose a serious threat to activities in semitropical and tropical regions of the world. An opportunity to deal with this problem to the benefit of our own interest, as well as those of the many underdeveloped countries in the affected regions, appears to be offered by the demonstration that immunization with the sporozite stage of the parasite affords partial to complete protection against infection. This, together with the development of techniques to cultivate the parasite in vitro, makes an effective malaria vaccine a potential reality in the next decade.

Effect of Electromagnetic Environment on Humans

The research opportunity here is a broad-based program to define the effects of the electromagnetic environment on health.

Genetic Engineering

Genetic engineering offers both a problem and an opportunity. On one hand, a problem is posed by the potential ability, through DNA recombinations, plasmid transfer, and other forms of genetic engineering, to create, artificially, a wide variety of pathogens including potential adversary development of biological warfare agents. On the other hand, we also have the opportunity to counter not only these threats to health but many others through the development of methods to mass produce broad-spectrum, antiviral, and antibacterial substances (such as interferon), vaccines, and inert avirulent organisms with the proper antigenic structure and other survival properties to make mass immunization by the aerogenic route feasible. DNA or genetic engineering technology offers opportunities to improve waste disposal and pollution (air and water) control techniques, and there is a potential to develop improved single cells for the production of food and energy. DNA also offers promise in tissue culture technology—the development of organs for transplants achieved by regulating growth and eliminating histo-incompatibility. Limitations on DNA recombinant research may result from stringent guidelines imposed on this type of research.

Diseases With Long Latent Periods

The fact that exposure to certain chemical toxicants and physical agents at low levels may result in debilitating disease or death decades later has been recognized. This has resulted in a moral and legal requirement to assess the etiological role of a wide variety of potentially hazardous substances and energy forms encountered in naval and industrial environments. Of particular concern are the potential adverse effects of radiation emanating from equipment and systems, atmospheric contaminants, and chemicals, paints, and insulating materials widely used in shipyards and other industries. Basic research, including retrospective and long-term prospective epidemiological studies, offers the opportunity to quantify the health risk associated with such agents. A constraint on this approach, as well as on many other necessarily long-term studies, is the emphasis on the 5- to 10-year time frame for producing definitive results.

Oxygen Deprivation

A lack of oxygen to a part of the body, as in the case of wounds or injuries, or to the body as a whole, as in the case of exposure to an oxygen-poor climate, has harmful effects that are often irreparable. In recent years, the naturally occurring prostaglandin B1 has been shown to restore oxidative phosphorylation to aged mitochondria. This prostaglandin derivative, designated PGB, has the potential to act as a prophylaxis against the adverse effects of hypoxia; as such it warrants investigation. To date, it has not been defined chemically or isolated, and none of its pharmacological properties has been adequately assessed.

Motion Sickness

There is a lack of basic information from which to establish criteria for the ride characteristics to apply to the design of advanced, high-performance ships and craft. Low-frequency motion is commonly responsible for the decrement in human performance associated with sea or motion sickness; however, inability to perform has also been demonstrated when humans are exposed to high-frequency motion. The need exists for basic and applied research into the subject of human limitations in and adaptation to motion; of greatest interest is motion in the frequency range of 0.07 to 80 Hz.

Impact Acceleration

Impact acceleration is a major cause of injury in both military and civilian communities. The data available on human dynamic response to acceleration and impact are sparse. Although there are many models that attempt to predict human biodynamic response to accelerated force fields, none has been, or can be, validated. A means of evaluating and testing engineering designs is greatly needed. To date, only the dynamic response of the human head and neck to impact acceleration has been studied. The whole area of techniques to employ human test subjects in dynamic response situations and to transfer the resulting data to models in sufficient quantity that the models can be validated is one that urgently needs investigation. The limitations imposed by using human test subjects demand a well-planned methodical program.

Constraints

A general constraint to the pace of progress in health-related studies is caused by increasing nationwide concern over the protection of human subjects utilized in medical research. This has resulted in the establishment of more stringent review procedures than any previously utilized. As a result, some categories of direct experimentation may be precluded. It is important that suitable alternatives to human test subjects be developed for experimental evaluations. Although animals still offer many possibilities, improved techniques for extrapolating results derived from animal experiments to the human must be found. It should also be noted that there is a similar growth in public concern over the use of animals in biomedical research; this is a potential constraint that may adversely affect the pace of progress in health-related research.

ENVIRONMENT AND ENVIRONMENTAL PROTECTION

Monitoring Ocean and Arctic Environments

Information about the oceans and arctic areas, their dynamics and their interaction with and impact upon the atmosphere, is needed in real time for input to predictions of climate and weather, optimum ship routing, efficient and effective fisheries planning and management, and effective search and rescue operations. Research on the proper scale is very expensive, requiring multiagency and multinational commitments over extended periods of time. A sustained effort is required to determine the time-space scales, and the types of measurement, and for concomitant instrument development. An opportunity is offered through the application of satellite sensors, employed in combination with "ground truth" measurements and large computer modeling. A coordinated national research and development program to determine the economic and productive balance of these techniques will be required. Constraints will occur due to the need to solicit international cooperation and establish priorities in support of such programs.

Propagation

There is a need to have an improved fundamental understanding of the ionosphere (including those conditions that give rise to multipath propagation), a better understanding of worldwide meteorological conditions as they affect ducting, further knowledge of propagation in turbulent media and in nonlinear and ionized media, and improved knowledge of the sea surface for sea clutter modeling. Studies of microwave propagation, solar radiations in near space, plasma radiation, remote sensing radar, magnetospheric substorms, and radio wave propagation through turbulent ionized media are recommended. The lacks of money and of people with an understanding of the problems are limiting factors.

Preservation of Wood Pilings

The Environmental Protection Agency (EPA) may soon ban the use of creosote for preserving wood pilings. Creosote is the most effective and most widely used preservative in the world. If the EPA bans its use, the U.S. Navy (and many others) would be faced with a costly problem of replacing pilings. The ban could affect railroad ties and telephone poles as well as marine pilings. Two opportunities are suggested: develop a suitable preservative for pilings that will not harm the environment, or develop a suitable material to replace wood pilings entirely.

MANPOWER

Improving Work Force Effectiveness

The assessment of individuals for selection, classification, training, and advancement is a problem common to industry, educators, government, and military services. The business of testing people is economically significant and has, indirectly, considerable social and cultural impact. Generally, it is important that testing be efficient, accurate, and bias-free. A significant opportunity is offered by improvements in and increased utilization of computerized adaptive testing, based on a theory of personnel assessment that appears to have both quantitative and qualitative improvements over older testing schemes. The application of modern test theory and adaptive testing techniques can result in tests more than 50 percent shorter than conventional tests and with higher levels of measurement precision. Adaptive achievement tests with half the number of items can have validities equal to those of conventional tests, and multicontent achievement tests can also be reduced drastically in length without decreasing measurement efficiency. The design of the tests, as well as the intervention of the computer, serves to decrease bias. Obvious constraints are the initial cost of implementation and large-scale test construction for many discipline areas.

Supply of Manpower

In the next 10 to 20 years, the U.S. Armed Forces face the problem of an insufficient quantity of available young people. Recruitment and retention problems exist, and the costs associated with a rapid turnover in personnel will be high. The problem in the Department of Defense is, therefore, how to obtain and retain sufficient personnel to protect this country effectively. At the same time, society in general suffers from growing welfare costs, high youth unemployment, a hard core of unskilled labor, a rising rate of functional illiteracy, and an increase in the incidence of young and female criminals. Thought should be given to looking at the many facets of the same problem and to seeking a series of integrated solutions rather than individual and perhaps conflicting solutions for each of these facets. A coordinated program of basic research into the cause and effect of human behavior is required. If government is to be a force in the solution, a single, coordinated plan should be formulated, with each of several government agencies and departments prosecuting those parts of the plan within its area of expertise.

Department of Energy

In October 1977, the Department of Energy (DOE) came into existence to deal with the nation's changing energy circumstances, brought most vividly to our attention by the Arab oil embargo in the fall of 1973. The Department's tasks are to encourage the increased supply of energy, to administer regulations required by law, to encourage the dampening of growth in demand for energy by means of conservation in the home and increased productivity in the use of energy resources in industry and transport, and to accomplish all these tasks with due consideration for the health and safety of the American people and environment.

The economic and regulatory functions and the conservation and commercialization functions do not for the most part involve new scientific and technological research. These programs predominantly make use of already-existing scientific and technological knowledge and development, and they try to extend the use of this knowledge and development in the marketplace.

Though the next decade will inevitably see sharp debate about pricing, regulation, and competition among energy sources, there is now a general consensus that world petroleum production will peak around the year 2000 or before. The Department's strategy rests upon the inexorable fact that supplies of abundant, inexpensive petroleum are rapidly running out. An economy and a society based on the premise of abundant, cheap supplies must inevitably change.

Broadly, the alternatives available for exploitation include solar energy, nuclear fission, and fusion. Each poses severe technological and scientific challenges, and the need to bring these sources to a stage at which their promise can be assessed, along with the need for effective development of alternate (if temporary) coal uses and development of conservation measures, dominates the DOE outlook. The Department's goals extend beyond the time period of normal private investment and involve the encouragement and development of energy systems for the twenty-first century.

Unlike the Department of Defense and the National Aeronautics and Space Administration, DOE will not be the final user or purchaser of the products of its own research and development. Success in DOE's research, development, and demonstration programs requires that the systems eventually leave DOE's domain to enter and compete in the market sector.

The seven sections that follow are devoted to solar energy; geothermal energy; fossil energy; fission energy, including nuclear waste management; fusion energy; the environment; and the Basic Energy Sciences, High Energy Physics, and Nuclear Physics programs. The first five categories include technologies having widely varying time horizons for expected results. While fossil and fission systems are working now, continuing research and development are still needed. Most solar systems have yet to be made economically viable, and thus need much technological development. Fusion is approaching scientific feasibility demonstration. While large funds are being committed in an attempt to ensure its timely emergence, its full operational capability may not occur before the end of the first quarter of the next century. The sixth category, environment, addresses generic consequences of all the technologies, while the last encompasses the building of an underlying intellectual base for the technology programs.

SOLAR ENERGY

Solar energy has captured the imagination of the American people because of a growing base of support for developing environmentally safe energy sources and because many recognize the inevitable shortcomings of relying on depletable resources.

In May of 1978, the President gave new impetus to the national effort to develop solar energy by simultaneously directing that support for solar research, development, and demonstration be increased by $100 million and initiating a formal governmentwide Domestic Policy Review (DPR) for solar energy. The DPR concluded that solar energy does offer numerous important advantages over competing technologies and affirmed other recent assessments of solar energy by finding that a significant potential exists for expanding its use during the next twenty years. The President's solar message in June 1979 recognized the findings of the DPR and proposed a number of incentives, such as additional tax credits and a solar bank, to spur accelerated use of solar technologies. A goal was also established for the nation to meet 20 percent of its energy needs from solar and solar-related energy forms by the year 2000. The future solar share of national energy use will depend not only on technological success but on the price of alternatives.

Although some solar technologies (for example, solar hot water heating where the alternative is using high-cost electricity) are competitive today, other uses of solar and renewable energy will require the conception and development of new methods and processes that have the potential for dramatic cost reductions relative to their present-day counterparts.

The problem of cost competitiveness, which is fundamental to solar development, presents a wide field of research opportunities for improving design, fabrication, and operation of solar systems. Inexpensive and durable structural materials are needed to reduce the cost of existing collectors. An especially important requirement is the development of highly transparent and nondegrading plastics for protection of collector faces and for encapsulating photovoltaic devices. New automated manufacturing and processing technologies are also needed to maximize volume-related cost reductions for solar hardware and materials.

An improved understanding of basic physical and chemical phenomena is necessary on many topics, including absorbtivity and emissivity characteristics of collector surfaces, solid state processes for increasing photovoltaic conversion efficiency, photochemical conversion processes, and biochemistry and genetics relevant to production of energy from biomass.

In addition, there are research opportunities in novel systems development and design engineering disciplines for operations in ocean, space, and terrestrial environments under extreme conditions; in systems operation and control technology for integration of large numbers of dispersed devices into large power networks or centralized electric and thermal power facilities; in biochemical and thermochemical conversion processes for biomass development; in improved meterological and environmental data on solar availability and wind characteristics; and in solar storage, or load leveling equipment.

Much of the research that has been undertaken to support the development of solar energy is quite recent, and overall support has accelerated rapidly. For example, solar R&D in 1974 was $15 million while the level proposed by the President for fiscal year 1980 is over $600 million. This reflects the realization that collecting and converting energy from the sun will provide mankind with an ultimate ceiling on the cost of energy resources. The trick is to accomplish this in a manner that can substitute for conventional energy forms.

GEOTHERMAL ENERGY

Geothermal energy is the heat in the Earth's crust which, in some areas, is close enough to the surface to be tapped as a source of power. In addition to generating electricity, geothermal energy may be used for industrial and residential space heating and cooling, industrial process heat, and agriculture. Exploitation of this resource requires research in a number of key areas, including confirmation of geothermal resources through novel techniques for locating and delineating reservoirs, their geologic structure, and their flow patterns; research into direct heat applications for geothermal fluids in industry, agriculture, and space-conditioning; design, operation, licensing, and maintenance of experimental facilities to produce electric power and to test systems and components; development of new drilling technology and well-stimulation techniques to make geothermal reservoirs more productive; and research in resource exploration, reservoir mechanics, geochemical engineering, materials science, and energy conversion systems to improve geothermal system performance. There are three general types of geothermal energy that may prove useful before the year 2000. They are vapor and liquid dominated hydrothermal energy, geopressurized hydrothermal energy, and hot dry rock.

Vapor and Liquid Dominated Hydrothermal Energy

Vapor and liquid dominated geothermal energy refers to hot water and steam that has been produced by contact with relatively shallow masses of hot rock. The liquid or vapor is trapped below ground in fractured rock or porous sediment. Vapor reservoirs are relatively rare, but Larderello, in Italy, has been operating since 1904, and the Geysers, in California, has been operating since 1960. Hot water dominated reservoirs are about 20 times as numerous as ones dominated by steam. Here the liquid is flashed to steam or passed through a heat exchanger to vaporize a hydrocarbon fluid that is used to drive a turbine-generator. The spent geothermal fluid is injected back into the ground.

Geopressurized Hydrothermal Energy

Geopressurized hydrothermal resources are hot water reservoirs containing methane and trapped under high pressure. They are located mostly along the Gulf Coast of the United States. This resource offers three types of energy: thermal, kinetic, and dissolved methane. While the total resource is believed to be large, the economics of recovery are unknown.

Hot Dry Rock

Hot dry rock, as the name implies, is rock that has a high heat content but little or no water; the introduction of water is required to supply usable power. This resource is also estimated to be very large, but significant use of it is probably several decades away.

FOSSIL ENERGY

Fossil fuels are now, and will remain, the base of our energy supply through the end of the century. There is an overriding need to develop means to improve fossil fuel extraction and utilization to postpone the time in which petroleum shortages severely inhibit national objectives. With many extraction techniques, one-half of coal and two-thirds of oil resources are left underground. Generation of electricity is less than 40 percent thermally efficient, and environmental control of coal combustion and oil shale retorting is very costly, partly due to the lack of the application of sophisticated techniques.

Modern scientific tools are revealing the chemical structure of fossil fuels and the fundamental mechanisms involved in their extraction, conversion, and combustion. Instruments not previously available include nuclear magnetic resonance, neutron activation, chromatography, microwave, and laser applications, as well as techniques of isotopic tracers, mathematical modeling, thermodynamics, and, especially, recent advances in catalytic chemistry. An important feature of our knowledge is the basis it provides for innovative ideas for new or much improved technology in such areas as those described below.

Conversion of Coal to Clean Fuels

Major research needs include innovative process concepts involving advanced catalysts for liquefying and gasifying coal, advanced synthesis gas conversion, and upgrading of coal liquids and flash hydropyrolysis; fundamental studies to develop new coal chemistry and a better understanding of coal conversion processes, the structure of coal, and the role of mineral matter in coal in the conversion process; and essential studies on mechanisms, kinetics, and the effects of reaction constituents on the catalyst as a function of time and temperature. Such research could lead to the development of combination catalysts that would enable steam-carbon gasification, water-gas shift, and methanation to take place simultaneously in the same reactor at reduced overall cost.

Also needed are supporting research and development activities that provide back-up for advanced liquefaction processes and improvements in liquefaction technology in general. In addition, indirect liquefaction techniques (liquid fuels via syngas from coal) provide an alternate route for furnishing transportation fuels that can ease the pinches caused by the depletion of petroleum supplies.

Direct Combustion of Coal

The ultimate goal of the DOE coal strategy is to decrease the nation's dependence on imported oil through a range of cost-effective uses of coal. Close to 90 percent of the coal consumed in this country is and will continue to be burned directly, so a primary objective in achieving the goal is to assure that existing and new facilities burning coal can do so while meeting applicable environmental standards. Unless new technologies are available for meeting these standards, many facilities that might use coal will turn to other fuels. In the long run, other fuels may simply not be available.

Scientific investigations to solve these problems include the elucidation of the chemical reactions controlling NO_x and SO_x formation, control of effluents from combustion of highly aromatic synthetic fuels, chemical desulfurization of coal, mathematical predictive models for coal devolatilization, combustion and ash fouling parameters, stability of coal/oil and water mixtures, and fluid bed engineering. More basic knowledge of fuel cell catalysis and high temperature phenomena in magnetohydrodynamics (MHD) can provide the basis for important technical advances.

Enhanced Oil and Gas Recovery

The fundamental relationship between structure and performance of micellar surfactants and the physical chemistry of oil displacement must be established. Measurement of the quantitative basis for the induced fracture patterns in tight formations and consequent gas transport is essential.

Oil Shale

Major tasks include determining the physical chemistry of in situ oil shale conversion and providing mathematical models for process engineering.

Coal Liquefaction and Gasification

Identifying the basis for catalytic breaking of coal molecules on a molecular mechanism basis will lead to new ideas for selective conversion under mild conditions. Engineering and chemical studies will yield discoveries on how to accelerate the steam-carbon reaction at relatively low temperature, employing base catalysis and transition metal complexes.

Electricity and Power Generation

Efficient generation of power entails the establishment of fundamental electron transport phenomena in MHD and fuel cell systems and the determination of the fundamental chemistry of fossil fuel combustion.

FISSION ENERGY

The nuclear fission strategy of DOE is based on three fundamental objectives: to refine and extend the use of light water reactors (LWR) as a major source of electricity by dealing with such problems as siting and nuclear waste management; to develop, assess, and analyze a range of alternate reactor concepts that would be major contributors to future energy supplies; and to develop international agreements, assurances, and safeguards so that a world nuclear power economy will not lead to the further spread of nuclear weapons. In this perspective the central question becomes, "How long, and with what technical improvements in fuel cycle and operating efficiency, will available uranium resources allow reliance on the LWR?" This defines the period in which alternate technologies (such as breeder reactors) must be developed in order for nuclear energy to continue to be a major energy source. Most of the research is carried out in government and industrial laboratories.

LIGHT WATER REACTORS

Research is directed at developing and demonstrating technology for increasing the efficiency of uranium utilization in light water reactors operating on the once-through fuel cycle, and developing methods for reducing the exposure of nuclear workers to radiation fields. Issues to be addressed include the development of low enriched uranium oxide fuels capable of burn-up of 45 to 50 thousand megawatt days per tonne, and development of chemical cleaning techniques to reduce radiation fields in the reactor plant.

An alternate means of continuing with nuclear power short of a commitment to breeders is the development of advanced isotope separation technology. This involves two laser methods and a plasma concept. Key technical issues to be addressed are laser development, materials development, and an increased understanding of the basic interactions involved in the various processes.

BREEDER REACTOR SYSTEMS

Breeder reactor systems offer the prospect of extending the nation's uranium resources base to a practically inexhaustible source of energy and providing one of the major choices for future alternate energy systems. The breeder programs are liquid metal fast breeder reactors, water cooled breeders, gas cooled fast breeder reactors, and fuel cycle research and development. Research is being directed toward maintaining and improving the technical bases of the breeder systems while providing technical support to the nuclear nonproliferation studies.

Liquid Metal Fast Breeder Reactors

This breeder utilizes the excellent heat transfer properties of sodium which permit operation at high temperatures without resorting to high pressure. The program emphasizes alternative fuels and fuel cycles that might offer strategic nonproliferation advantages. A strong base-technology effort stresses the development of engineering information applicable to large plants. Program technology objectives are to ensure that future plants will be safe, reliable, economic, and licensable. Key technical issues to be addressed include development of advanced and alternate fuels, safety testing, plant design, and scale-up of such major components and systems as pumps and steam generators.

Water Cooled Breeders

The water cooled breeder program is developing technology which will significantly increase the amount of useful energy that can be produced from nuclear fuel reserves using water cooled reactors. The central effort is to confirm that breeding can be achieved in existing and future light water reactors using the thorium/uranium-233 fuel system.

Gas Cooled Fast Breeder Reactors

The objective of this program is to develop a gas cooled fast breeder reactor. The key technical issues at this point are conceptual design and supporting areas of fuels, physics, thermal hydraulics, and safety.

NUCLEAR WASTE MANAGEMENT

To ensure that the nation has a nuclear energy option available, it is particularly important that a sound and credible means be found for disposing of nuclear wastes. The DOE program is designed to protect the public health and safety and the environment from radioactive materials that result from national defense programs, research and development, and commercial activities. The need for an adequate nuclear waste disposal program is the largest single constraint on the development of nuclear power.

During 1978, significant steps were taken toward the development of a national nuclear waste management policy that will guide R&D and policy development over the next five years. Under Presidential directive, the Interagency Review Group on Nuclear Waste Management, representing all Federal agencies having responsibilities in nuclear waste management, conducted a broad-based review of the subject with full participation of the public and state and local governments. Several key issues of both a technical and an institutional nature have been identified; they will require particular attention and will greatly influence the nuclear waste management effort in years to come.

A key finding is that a social consensus, supported by perceptions of scientific evidence, is essential to the establishment of long-term programs for permanent waste isolation. Consistent with this view, the Interagency Group has endorsed several measures to enhance the role of the public, as well as state and local governments, in reviewing and shaping waste management policy.

Although the Interagency Group stated that it appeared to be possible to isolate radioactive wastes safely for thousands of years, it acknowledged that it is nevertheless necessary to resolve several specific problems before proceeding with extensive programs for permanent isolation. The technically conservative approach that was advocated would involve the careful characterization of candidate disposal sites and the development of standards for the establishment of intermediate-scale and, eventually, full-scale mined repositories. At the same time, research and development will be intensified to expand knowledge of the behavior of wastes in emplacement media and to ascertain that all dominant sources of risk, or pathways of radioactive wastes to the human environment, are identified and accommodated.

Research needs include the development of conceptual designs for the conversion of wastes into possible forms suitable for acceptance at a long-term geologic repository; evaluation of alternative forms of processed wastes to study compatibility with emplacement media, leachability, and other properties vital to isolation from the biosphere; the development of handling, storage, and monitoring techniques to minimize reliance on long-term maintenance and surveillance; geotechnical studies of specific phenomena in, and physical properties of, candidate media, including hydrology, thermal effects, radionuclide migration, and structural characteristics; transportation R&D; and facility decontamination and decommissioning.

FUSION ENERGY

Fusion provides a major alternative future source of energy. Although the technological and economic viability of fusion has not yet been demonstrated, major engineering experiments in the late 1980s should establish fusion as an energy source capable of commercialization in the 1990 to 2000 time frame.

Two fundamentally different approaches are being followed toward development of fusion energy: magnetic confinement and inertial confinement. In magnetic confinement, the hot ionized gases (plasmas) that will undergo fusion reactions are contained in carefully shaped magnetic fields. Diverse methods of heating are used to raise the plasma temperature to that required for substantial fusion reactions. In inertial confinement, the fuel that will undergo fusion reactions is contained within a small pellet. A large amount of energy is deposited into the pellet by an appropriate device, such as a laser beam, raising the pellet temperature high enough to cause the fusion reaction. Within each of these conceptual approaches there are many different schemes for achieving a controlled thermonuclear reaction.

In the near future, the fusion program will investigate many different schemes since it would be premature to narrow the field of options at this time. To complement the investigations, vigorous research in plasma physics and fusion technology and engineering is needed.

The near-term objectives of fusion energy research emphasize the demonstration of scientific feasibility, establishment of a sound engineering development base, maintenance and development of a strong scientific and technological base, encouragement of research in alternative concepts and scrutiny of their positions relative to established major projects, and reduction of large experimental device costs through international cooperation and cost sharing.

Although at present net energy is not being produced by fusion, the technical requirements for sustained fusion reactions are that the fusion fuel must be heated and maintained at a temperature near its ignition point (approximately 100,000,000 K for the D-T reaction), and the plasma must be confined for a time and must have a density sufficiently high to ensure that the ions are close enough for reactions to occur. (The density is the number of ions per unit volume of plasma.) To achieve the requisite plasma parameters, a considerable wealth of technologies has already been started to support the scientific studies. For magnetic fusion, these technologies include high-intensity, large-volume magnets (both normal and superconducting), high-intensity, high-energy neutral beam injectors, radio frequency heating systems, sophisticated laser diagnostics, and data acquisition systems. Some of these are unique to fusion development, while others are adaptations of existing technologies. For inertial fusion, the technologies include high-power, well-focused particle beams, low-cost pellets, accurate pellet injection and tracking systems, short-pulsed power supplies, and advanced pellet diagnostics.

TECHNICAL ISSUES THAT NEED INVESTIGATION

Plasma Behavior

The analytic and computational capability to predict plasma behavior under a wide variety of conditions needs continuing development. Of special interest to magnetic fusion are magnetic field changes, long pulses, impurities, cross-sectional area, and shape. Of interest to laser fusion are radiation-plasma interaction, pellet implosion phenomena, and general pellet configuration effects.

Scaling Laws

The ability to scale up the present generation of experiments to larger experiments and prototype reactors must continue to be developed and validated.

Prototype Development

The various plasma confinement concepts require more experimental and theoretical development so that we can evaluate their feasibility as prototypes upon which to build a more realistic, commercial-size reactor design.

Technology Development

Some of the technological obstacles to magnetic fusion include the development and construction of large and powerful superconducting magnets, the creation and maintenance of very high vacuums, the heating of fusion fuel to ignition temperatures, and the refueling of burning plasmas. Technology requirements for inertial fusion include development of short-pulsed power supplies, development of short-pulsed high-power photon and particle beams, development of beam transport systems, and development of pellet manufacturing, injection, and tracking systems.

Engineering Issues

Practical fusion power will require facilities operating at high reliability and availability. Considerable attention and ingenuity are necessary to develop designs with sufficient access, maintainability, simplicity, fabricability, etc., to reach the reliability and availability goals.

Materials Behavior

Structural materials that will have long lifetimes in a fusion reactor environment of high temperature and large flux of energetic neutrons must be developed. Some extrapolation can be made from measurements in fission reactors, but new dedicated testing facilities are required.

ENVIRONMENTAL ISSUES

Although fusion energy is much more benign than fission energy, environmental aspects of its development and commercialization must continue to be examined.

Radioactivity Control

Methods for managing the large inventories of radioactive tritium required in large experimental devices and prototype reactors will have to be demonstrated. Low activation materials capable of functioning in a fusion reactor environment must be developed. Radioactive material management and potential accident hazards depend on the degree to which this is achieved. Nonelectric applications, such as breeding of fissile material, may present additional problems which will be examined at an appropriate time.

Safety Analysis and Engineering

Safety analyses of early conceptual fusion reactor designs are being carried out to identify potential hazards so that appropriately engineered safety features can be designed. Lithium, which is used to breed tritium fuel, must be carefully managed because it contains a large amount of potential energy in liquid metal form.

Human Health and Ecological Hazards

The occupational health impacts of routine maintenance of fusion reactor facilities must continue to be investigated. The ecological and human health effects of some of the possible neutron-activated products of potential reactor materials have not yet received extensive study. There must also be an assessment of the adequacy of current health and safety standards for workers associated with current and planned fusion research and development.

ENVIRONMENT

Another responsibility of DOE is to assure compliance with environmental, health, and safety plans, regulations, and procedures. To this end, the Department conducts a comprehensive program of research and development on the health, safety, and environmental effects of energy technologies and programs. Independent reviews and assessments of DOE-controlled activities are initiated by DOE to assure that there are no undue risks to the safety and health of the public and employees, and that adequate protection is provided to property and the environment.

Each individual technology program office in DOE is responsible for the environmental compliance and acceptability of the technologies it is developing for commercialization. In addition, a central environmental office conducts research in support of, and in cooperation with, the technology program offices to determine the biomedical and ecological effects associated with the development and use of fossil and nuclear fuels, renewable energy resources, and conservation measures.

Source terms under study include trace and heavy metals, polycyclic aromatic hydrocarbons, particulates, noxious gases, organic fluids, radionuclides, and complex mixtures associated with fugitive emissions, residuals, byproducts, and products associated with all stages of the fuel cycle. With a near-term potential for new fossil energy processes to enter both demonstration and commercialization phases, expanded new chemical and biological characterizations, and health and environmental effects, research will be conducted in support of the following technologies: in situ and surface coal gasification, in situ and surface oil shale retorting, coal liquefaction and enhanced oil recovery, and large-scale deployment of diesel powered light-duty vehicles. Specific research areas are described below.

Pollution Studies

Studies on the characterization, measurement, and monitoring of pollutants are designed to

(1) Determine the chemical and physical mechanisms of pollutant transformation in the atmosphere;

(2) Investigate the underlying processes that govern pollutant interaction in biological systems;

(3) Develop empirical models for predicting biological effects of energy-related pollutants;

(4) Develop advanced radiation, chemical, and particulate measurement technology;

(5) Develop improved health protection methodology; and

(6) Characterize and quantitatively measure the effluents and emissions from selected technology processes.

Environmental Studies

Studies of the environment aim to

(1) Define and quantify the characteristics, transport, and transformations of energy-related pollutants in the atmosphere, on land, and in water;

(2) Identify major pathways and rates of transfer of pollutants through the biotic compartments of the environment;

(3) Determine the biological effects of pollutants on ecosystems and the rates and mechanisms by which populations and ecosystems recover or accommodate to such stresses;

(4) Determine the effects of massive physical disturbances and disruptions on all environmental compartments;

(5) Define and implement management strategies for land, atmospheric, and water resources that are subject to impact during energy development;

(6) Develop and evaluate countermeasures to eliminate pollutant pathways to man and to lessen or combat the impact on the environment; and

(7) Guide fundamental research in the geophysical media, structure of ecosystems, and responses of biotic communities to stress, and develop methods for early detection of changes in environmental systems.

Health Studies

Health studies are oriented to identify and characterize biologically the active forms of hazardous agents that may present health risks associated with development of energy technologies by

(1) Coupled chemical-biological identification of potentially toxic, mutagenic, carcinogenic, and teratogenic agents;

(2) Definition of effective doses of pollutants by measuring exposure, distribution, and fate in the body, and retention and excretion;

(3) Correlation of the effective pollutant dose with the magnitude of effect observed in cells, tissues, and organs of several species of animals; and

(4) Continuing development of experimental approaches to overcome the obstacles to extrapolation of estimates obtained in experimental animals to human risk assessment.

It is important to develop and apply useful and sensitive indicators and short-term screening systems for detection of biological damage by the

(1) Use of microbial, mammalian, and human cell culture systems to identify agents that damage the genetic material and lead to mutagenesis and carcinogenesis;

(2) Development of applications of tests for recognition of agents that impair the function of cells, organs, and physiological systems;

(3) Measurement of potency of carcinogens, mutagens, and toxic agents in model biological test systems;

(4) Use of molecular systems to define site and mechanism of action of compounds identified as potentially hazardous; and

(5) Identification of repair mechanisms and correlation of their sensitivity to, and effectiveness for repair of, damage produced by toxic agents.

Further studies will be designed to determine the qualitative and quantitative effects of pollutants on humans who have been exposed in the occupational and general environment. Specifically, the aims are to

(1) Expand epidemiological studies to determine long-term effects of low-level radiation in individuals exposed accidentally, occupationally, or as a result of weapon explosions;

(2) Expand studies on the occupational and other effects associated with petrochemical refineries, recovery of oil along the outer continental shelf, and exposure to chemicals used in enhanced oil and gas recovery operations; and

(3) Expand health and environmental effects assessment efforts to include oil shale, coal liquefaction, fluidized bed combustion, energy storage, fossil fuel combustion, diesel emissions, and geothermal energy.

Carbon Dioxide Studies

Research and assessment studies on carbon dioxide effects will try to

(1) Predict future atmospheric carbon dioxide concentrations;

(2) Predict the climatic effects of increasing the levels of atmospheric carbon dioxide;

(3) Predict the cascade of environmental effects resulting from climate changes;

(4) Predict social, political, and economic consequences of global environmental change; and

(5) Investigate strategies and technological "fixes" for reducing the consequences of carbon dioxide impacts.

THE BASIC ENERGY SCIENCES, HIGH ENERGY PHYSICS, AND NUCLEAR PHYSICS PROGRAMS

Three major basic research outlay programs are administered by DOE: Basic Energy Sciences, High Energy Physics, and Nuclear Physics.

Basic Energy Sciences

Basic Energy Sciences (BES) is DOE's principal program for sustaining basic research in the disciplines underlying energy supply and conservation. The central role for the program as a whole is to provide resources that will sustain and enhance the national energy research and development enterprise. Primary users of the results from the current research will be the scientists and engineers working in that enterprise one and two decades from now, in either federally sponsored efforts or efforts funded by the private sector. Research under the BES program is seen as a key part of the Federal role in assuring that the nation's future energy needs are met as economically as possible and with minimum adverse environmental impact. The budget reflects the great importance attached to basic research in energy-related fields. With the help of scientists from throughout the country, research opportunities of long-term importance for dealing with the nation's energy problems are identified. In this way, the sustaining research in the BES program complements the programs built around specific energy technologies.

The scope of the BES program is being extended significantly in fiscal year 1980. The new division of Biological Energy Research, for example, will broaden its coverage as it seeks to advance scientific understanding in the disciplines underlying solar biomass technologies, as well as other biological approaches to energy conversion and conservation. There are similar needs in applied mathematics and geosciences; the latter is of crucial concern in location of shock-sensitive facilities, exploitation of geothermal heat and fuel resources, and the safe immobilization of nuclear wastes. We are undertaking new activities in basic research involving engineering and new exploratory energy concepts. At the same time, further emphasis is needed, and is planned, on research in combustion, catalysis, and corrosion of materials. Research under this program is carried out at the DOE national laboratories and, increasingly, at universities throughout the nation.

High Energy and Nuclear Physics

DOE's role in High Energy and Nuclear Physics differs in a fundamental way from its role in other basic research efforts. For compelling historical and pragmatic reasons, DOE serves as the principal executive agent for research in these fields. The work can be viewed as maintaining a special trust since the Department's responsibilities in these fields stand somewhat apart from the main body of DOE energy responsibilities. The principal mission of these programs is the conduct of basic research.

Interactions with other DOE programs are important, but they determine neither the magnitude nor the directions of the principal efforts in High Energy and Nuclear Physics. These programs seek deeper understanding of the ultimate structure of matter and physical events. High Energy Physics deals with elementary particles—their creation, their transformations, and the forces and other relationships among them. Nuclear Physics deals with collections of elementary particles as they appear in the many different nuclei of atoms.

The experiments in both of these disciplines are centered around accelerators. A number of major new accelerator-based facilities are under construction. At the same time, a review of the national effort in these fields has led to conclusions consistent with maintaining DOE support at essentially a constant level of effort over the next several years. The research is recognized to be of great national importance. DOE seeks to maintain world leadership for the United States. Bringing the new facilities into effective operation within a firmly constrained level of effort will present major challenges to the communities of scientists involved in these efforts. The prospects remain excellent, however, for a new synthesis of the basic laws of physics which, if history is a valid guide, may enhance profoundly the ability of scientists to deal with a wide range of problems and needs, including those related to energy.

Department of Health, Education, and Welfare Alcohol, Drug Abuse, and Mental Health Administration

The Federal efforts to reduce, or eliminate, where possible, health problems caused to the people of the United States by the abuse of alcohol and drugs, and to improve the mental health of the nation, are managed by the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA). Biological, psychological, sociological, and epidemiological research is conducted to carry out this mission. In addition to basic research, ADAMHA is concerned with the application of principles, methods, and techniques and the development of personnel for research, treatment, and prevention in its areas of responsibility.

ADAMHA's programs are carried out by the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the National Institute of Mental Health. The broad headings in this paper correspond to the three Institutes; within each section, the current and near-future areas of research confronting each Institute are described.

NATIONAL INSTITUTE ON ALCOHOL ABUSE AND ALCOHOLISM

The long-range goal of the research program of the National Institute on Alcohol Abuse and Alcoholism (NIAAA) is the development of knowledge about the causes and consequences of harmful alcohol use in order to reduce the incidence and prevalence of alcohol abuse and alcoholism and to reduce the morbidity and mortality associated with these conditions. With this goal in mind, the research program supports and conducts research in six major areas: etiology, pathogenesis, early identification, treatment, and prevention of alcohol-related problems, and the development of basic techniques and information relevant to these concerns. Within these six areas, the proposed research programs over the next five years will include efforts to increase knowledge of the following:

Medical and Psychosocial Antecedents

Funds in this area will allow for research on such issues as the genetic bases for alcoholism, the identification of physiochemical markers that could be used for diagnostic and treatment purposes, and the determination of familial or social (including occupational) factors that influence the development of alcohol problems. Particularly useful will be research to determine those factors that may function as generic causal agents as well as those that may relate to particular population groups (e.g., women, youth, minorities, the aged, and specific occupational groups).

Mechanisms of Alcohol Effects

Research is needed on such questions as the natural history of alcoholism, the processes of habit formation, the development of dependence and tolerance, and the mechanisms by which alcohol may adversely impact on metabolism, immunologic responses, endocrine functions, neuropsychiatric processes, and nutrition.

Relation to Major Medical Disorders

Preliminary evidence has indicated a relationship between alcohol use and both cancer and heart disease, and th findings should be refined. We need to investigate possible mechanisms by which alcohol might function as a cocarcinogen and as a cofactor in the development of heart disorders, as well as the processes by which alcohol might inhibit or advance radiologic or pharmacologic treatments for these disorders. Research is also needed to investigate the relation of alcohol use to such disorders as the fetal alcohol syndrome (FAS), cirrhosis, malnutrition, immunologic disorders, depression, and Wernicke-Korsakoff's syndrome.

Diagnostic Techniques

Objective tests that can be used by clinical personnel for the early identification and assessment of problems of alcohol abuse in their patients need to be developed. To be maximally useful, these tests should be readily and inexpensively available and include functional diagnostic criteria as well as clinical laboratory criteria.

Treatment Assessment

In this area we should develop methodologies by which assessments of new and existing treatment techniques can be made. Research to determine the appropriate criteria and techniques by which to define and study the treatment process, diagnostic classes of patients, and outcome criteria is a necessary precursor to a systematic and coordinated collaborative program of treatment assessment research. Such research must also focus on the differing needs and problems of diverse population groups (e.g., workers, women, and youth).

Techniques To Reduce Alcohol-Related Accidents and Violence

NIAAA will support research on the development of basic techniques and principles for use in prevention programs. Such research will be especially useful as it relates to particular population groups and to particular alcohol-related problems (driving accidents, occupational accidents, and familial violence). Research in this area will include the development of amethystic (sobering) agents that can be used as tools to prevent driving and other accidents and as a therapeutic technique. Research will focus also on the exploration of economic and regulatory techniques and incentives (e.g., at the workplace), as well as familial and peer pressure, as mechanisms in prevention programs.

Diagnostic Procedures for Fetal Disorders

Procedures to assist in the early identification of fetuses that have been adversely affected by parental alcohol use will be developed. Such techniques will aid in the early treatment of affected neonates, reduce the number of babies born with FAS, and aid as well in the detection and treatment of women with alcohol problems.

Primate Model for Impact on the Fetus

The development of a primate model that will allow for research on the impact of alcohol on the fetus is important. For clinical and ethical reasons, such research cannot be easily conducted in the human. Issues to be investigated include such concerns as the determination of a dose-response relationship, pathogenesis of FAS, and procedures for preventing pathological perinatal outcomes.

Alcohol Beverage Control (ABC) Laws

Support in this area will allow for research on the impact of ABC laws and their use as a preventive technique. Research is needed particularly on the relation between such laws and the incidence and prevalence of youthful problem drinkers, the relationship to the rate of alcohol-related driving accidents, and the efficacy of economic disincentives, media restrictions, and legal age restrictions in reducing alcohol consumption. Research is needed to assess differences among states as well as within states that change their ABC laws.

Identification of High-Risk Drinking Practices

NIAAA will support epidemiologic and other research on such concerns as the establishment of baseline data on drinking practices, particularly of special populations such as women of childbearing age, working women, and teenagers; the determination of socioeconomic and occupational parameters relevant to alcohol consumption and alcohol-related disorders; the determination of the medical consequences of alcohol use and the relation to other major medical illnesses; and the relationship of specific fetal defects to specific maternal consumption patterns.

Networks for Dissemination of Research Information to Clinicians

The appropriate and timely dissemination of research findings to health care practitioners is a particularly important area. Such information must be in a form that is useful to clinicians and must be easily accessible to them. NIAAA will support the development of mechanisms to identify and assess clinically relevant research findings as well as effective techniques and channels for transmitting useful and usable information to health care providers.

NATIONAL INSTITUTE ON DRUG ABUSE

The National Institute on Drug Abuse (NIDA) has primary Federal responsibility for the initiation, development, and execution of a comprehensive research program in the field of drug abuse. This program, which consists of basic as well as applied and developmental research, is concerned with the social, behavioral, and biomedical aspects of drug abuse, its prevention, and treatment. The goal of the Institute in supporting research is to develop a substantive data base which will lead to a better understanding of the dynamics involved in drug abuse and its treatment. With this knowledge, it will become substantially more possible to contain and control this serious public health problem.

Over the next five years NIDA anticipates substantial research efforts in such areas as endogenous substances and their receptors, development and testing of new treatment drugs and modalities, and the behavioral correlates and commonalities of substance abuse, including tobacco use.

A more specific discussion of NIDA's expectations with regard to research directions follows.

New Drugs in Treatment

During the past few years the most important treatment program has been the development of L-alpha-acetyl-methadol (LAAM), an alternative to methadone in the treatment of heroin addiction. This drug is presently undergoing its final phases of clinical testing. We anticipate further clinical testing in 1980 to answer particular questions raised by the Food and Drug Administration concerning the planned submission of a New Drug Application for LAAM. Large-scale clinical testing of LAAM in women should also begin in 1980. The final carcinogenicity testing of LAAM will also be concluded in 1980.

There has also been an extensive program for the development of naltrexone, a narcotic antagonist in the treatment of heroin addiction. In 1980, the large-scale clinical tests of naltrexone should be continued. Pending the results of toxicity studies, it is anticipated that the clinical testing of a long-term depot preparation for naltrexone will also be initiated.

NIDA is beginning to develop a new drug that may have significant potential as a treatment agent. This drug, buprenorphine, is a partial agonist of the morphine type with a long duration of action. In practical therapeutic terms, these characteristics of buprenorphine indicate that this compound has the potential for producing a psychic effect that should be acceptable to addicts in treatment. It will require a dose schedule less frequent than methadone's and possibly as infrequent as with LAAM. Buprenorphine appears to be significantly less toxic than other treatment drugs and may even be nontoxic.

Further, buprenorphine appears to be capable of blocking the toxic, euphorogenic, and physical dependence-producing effects of self-administered heroin or other narcotics. It will produce cross-tolerance, as methadone does, but will also act as a competitive antagonist similar to naltrexone.

Finally, buprenorphine appears to produce little if any physical dependence, and maintenance therapy could therefore easily be terminated. In 1980, the results of initial limited scale Phase II clinical testing of buprenorphine will be completed and a decision made on whether to proceed with the continued development of the drug.

Psychological Aspects of Treatment

NIDA also plans to develop a comprehensive and practical approach to the treatment of the heroin-dependent person and drug abusers that is based on theoretically sound knowledge of the psychiatric state of dependent persons and on the psychodynamics of drug abuse and psychological dependence. Research has shown that many addicts suffer from depression and that treatment with antidepressants is of considerable value in treating drug addiction in these patients. These studies will be continued and expanded during the next five years. NIDA-supported research will seek to assess the psychiatric status of addicts in and out of treatment and will also attempt to assess the impact of various types of psychotherapies on treatment outcomes.

Treatment Personnel

There is also a need for a study of the context of treatment and of the characteristics of treaters. Insufficient research is being performed on types of treaters and their effects on the treatment process. NIDA plans to develop a methodology to measure the treatment process, clinical milieu, and various characteristics of therapy. This is an area in which cooperation with the alcohol research field would be especially important because it takes a particular type of dedicated professional, in both fields, to stay with treatment of what must seem almost incorrigible diseases. A model for such a study was done years ago in a study of psychiatrists in public and private institutions. (The "burnout syndrome" among treaters should also be explored.) This kind of program will have to be done on a broad basis to capture the varieties of existing programs and their staffs.

Treatment Evaluation

The Treatment Outcome Prospective Study (TOPS) has been developed to critically examine client characteristics and performance during pretreatment, intreatment, and posttreatment periods. The study is a large-scale, long-term (5 to 7 years) effort to better understand the natural history of clients entering federally funded drug abuse treatment programs. TOPS provides for baseline demographic and performance measures to be collected at the time of admission to the participating treatment programs. These programs represent the more common treatment modalities (methadone detoxification, methadone maintenance, drug-free outpatient, and therapeutic communities) in six geographically disparate cities. The field test (Phase I) is underway in two cities—Chicago and Des Moines.

Pain

NIDA has taken the lead role in sponsoring research dealing with endogenous mechanisms of pain relief, including studies on endorphin and narcotic analgesics (including heroin).

The knowledge generated from studies on pain control mechanisms has important implications not only for the future of pain research, but also for studies on positive states of well-being. NIDA's effort in pain control provides a starting point for developing approaches for the study, in collaboration with NIH, of the quality of affect and states of well-being. Such studies should have direct relevance to efforts to improve the quality of life for the chronically ill, the terminally ill, and the elderly. They should also help focus research directed toward health maintenance.

Cigarette Smoking

Cigarette smoking is the nation's most widespread, costly, and dangerous habit. It involves one-third of the national population and leads to 300,000 deaths per year. NIDA has recently initiated a study of cigarette smoking behavior that will continue to further our understanding of the etiology and basic mechanisms of nicotine dependence and withdrawal and to increase our effectiveness in the treatment of this public health problem. Research will be conducted to identify biomedical, psychological, and social factors that may predispose many, but not all, individuals to become dependent on cigarettes. The basic physiological and psychological effects of nicotine dependence and withdrawal will be studied.

Endogenous Substances and the Addictive Process

The identification in the brain of opiate receptors and opiate-like polypeptides (endorphins and enkephalins) has led to suggestions that these receptors may be sites at which opiate drugs induce analgesia and euphoria, and that the polypeptides themselves may be transmitters or modulators in brain systems for the regulation of pain and pleasure. Investigators are expanding on the original research hoping to answer some fundamental questions not only about the mechanisms of tolerance, dependence, and drug abuse but also in the wider context of mental function. The tremendous amount of research activity that followed the discovery of opioid peptides has resulted in increased basic knowledge, new conceptual advances, and even some practical applications.

Our knowledge base is rapidly expanding as increasingly sophisticated methodologies—radioimmunoassay (RIA), immunohistochemistry, radioautography, etc.—are being used to study the neural pathways that contain endogenous opiate-like substances. Additional biochemical and physiological evidence supports the hypothesis that the enkephalins, and perhaps the endorphins, have neurotransmitter-like roles in certain pathways in the brain, and B-endorphins may have a hormonal function in the pituitary. The physiological actions of opiates ought, in general, to be explicable ultimately by their effects on opioid peptide/opiate receptor systems.

The conceptual advances, largely catalyzed by the discovery of endogenous opiates, have led to some new ways of looking at the roles of various chemical substances in the communication processes among neurons. The endogenous opiates did not fit into the current concept of a chemical neurotransmitter, and it was realized that a new concept of neuromodulations was needed. This, in turn, is leading to a new perspective on the various forms of mental illness and to a greater understanding of brain function in general. The next few years will undoubtedly see continuously improved diagnosis of mental illness and the development of new, effective treatments for some mental disorders. Further, it seems reasonable to postulate that other, nonopiate receptors will be found and linked with other endogenous substances, shedding new light on many of the psychological processes and behavioral phenomena of daily life.

Physical Predisposition to Drug Abuse

A body of information is beginning to form which clearly implicates certain biomedical factors as predisposing individuals to drug and substance abuse. Based on these findings, some of which are indicated below, NIDA is now able to seriously consider mounting a major extramural research effort concerned with these physiological factors.

For example, recent studies have indicated that individuals from normal college student populations who have low activity levels of an enzyme (monoamine oxidase, MAO) that affects brain activity, score high on personality measures such as the sensation-seeking scale (SSS). Moreover, high scores on this scale have been associated with increased use of marihuana, hashish, amphetamines, and LSD. Individual preferences for stimulants or polydrug abuse have also been associated with high sensation-seeking scales. During the past year, NIDA researchers have shown that activity levels of MAO in male marihuana smokers were significantly lower than in a comparable group of nonusers and that the level of current marihuana use was inversely correlated with MAO activity. It would appear that these enzyme activity levels may provide a reliable index of an individual's proclivity to abuse drugs if given access to the drugs.

Other examples of biomedical factors predictive for drug abuse include the possible role of low endorphins, depressive states, and psychopathy in the etiology of narcotic addiction. A number of additional biomedical factors should be studied for their role in drug abuse, including increased levels of endogenous antagonists, which may be the functional equivalent of low endorphin, and perinatal exposure to drugs. Once a person is exposed to or addicted to narcotics, there appears to be some type of biological memory, and readdiction may occur with greater ease; in addition, there appear to be a naturally occurring benzodiazapine-like receptor and a presumably naturally occurring benzodiazapine-like chemical. The abuse of or need for tranquilizers may be caused by a deficiency in this system.

Neuroscientific Basis of Drug Abuse

There is a need for further expansion of studies concerned with the basic neural mechanisms underlying drug abuse. Of particular interest is the neurobiology of euphoria and dysphoria, of the reinforcing properties of drugs, and of compulsive and habitual behaviors.

Disorders of Self-Control

A recent emphasis in the research program derives from awareness that the etiology of much drug abuse relates to general patterns of excessive and obsessive behavior. These patterns of habitual behaviors are extraordinarily difficult to eradicate. These types of behavior, which may also be referred to as disorders of self-control, include the use of licit and illicit drugs, alcohol, and tobacco; general dietary excess; and, possibly, certain aberrant behaviors such as gambling and child abuse. Taken together, these behavior patterns can conservatively be considered as leading contributors to the nation's overall health problems and constitute the group that would appear to be the most amenable to prevention and alleviation through understanding the mechanisms of self-control. NIDA will pursue this area of research during the next five years.

Gateway Drug Use

This initiative would seek to identify more fully the conditions of acquisition of gateway drug use. It would be primarily focused on individuals immediately at risk of experimentation with gateway drugs, those 10 to 13 years of age, for it is with this group that we feel progress can be made in interfering with the variables leading to drug abuse. The plans include investigating the role of drug-using and smoking in this age group, and the impact of parental smoking and drug-taking.

Natural History of Drug Abuse

Another area of renewed interest involves the natural history of drug abuse. We plan to increase our understanding of the various stages of initiation, practice, and final outcomes of drug abuse. Such information should prove helpful in evaluating treatment programs and in developing prevention programs. For example, such research might show that there are periods of increased or decreased use in an addict's career. Knowledge about the variables preceding the changes in behavior could be invaluable.

Longitudinal Studies of Marihuana

There has been increasing concern on theoretical and clinical grounds that marihuana use may have more serious deleterious effects on an individual when used during late childhood and early adolescence than it does on the young adult, who has been the subject of most marihuana research. Such adverse consequences might be the result of a special vulnerability arising from such factors as endocrinological changes, rapid growth, and immature ego development which are characteristic of this age group.

The primary purpose of this initiative is to focus on the possible special hazards of marihuana use for those from age 11 to 15 (approximately). Such studies will concentrate on the biological, social, and behavioral consequences of use in this age group. Planned areas of research involve studies of genetic, endocrinological, immunological, and developmental effects and studies of interpersonal relations, school performance, and psychodynamics.

Prediction Studies

Evidence shows that equations can be formulated to predict which adolescents will or will not get into moderate or heavy drug use. Primary focus has been on elucidating early or antecedent predictive variables. The next step for prediction studies is to delineate those elements that predict dysfunctional drug use among persons already using drugs. A related and significant area for further inquiry is to establish sets of variables that predict nonuse of drugs, i.e., those variables that insulate individuals from the need to engage in drug-using behaviors. The latter inquiry is often referred to as a search to identify "invulnerables" and the variables associated with their invulnerability.

Family, Socialization, and Group Factors Involved in Drug Abuse

NIDA will continue to place a certain amount of emphasis on adolescents, a group perceived to be particularly at risk for drug experimentation. Future investigations will include a continued interest in studies of family factors implicated in the initiation, maintenance, cessation, and prevention of drug abuse, and the role of peer groups in these same behaviors. NIDA will seek methodologically strong research investigating the role of numerous family-related variables in children's future drug abuse behavior and studies investigating the relationship between drug abuse and differing family units and situations. Less traditional and emerging family situations, e.g., addict mates, families in which the father has been awarded custody of the children in divorce proceedings, or families in which both parents are regularly employed on a full-time basis, will be given particular attention. Epidemiological studies will be initiated to determine the incidence and prevalence of drug use and those variables that are common to families involved in drug treatment situations.

Research on Specific Drugs

NIDA also will support research on certain drugs. Studies of PCP will seek to identify a blocking agent to be used to treat overdose victims; improve medical management of overdose cases and find faster ways to eliminate the drug from the body; improve understanding of the motivations of users in taking a drug so well known to be dangerous, so that its use can be decreased; and find improved and more widespread analytical methods for detection in bodily fluids.

Studies of marihuana will include further exploration of marihuana (THC) as an antiemetic in cancer chemotherapy (in collaboration with the National Cancer Institute); further studies of effects of chronic use (hormonal, motivational, pulmonary, carcinogenic); longitudinal studies, especially in youth; and systematic comparison of the effects of marihuana and alcohol (being done in collaboration with NIAAA).

Drug Use and Criminality

Further study is necessary to elucidate the connections among drug use, a criminal lifestyle, and economic factors. The traditional simplistic explanations have been shown to be based on erroneous assumptions.

Nutritional Factors in Drug Abuse

It is conceivable that nutritional factors, either in terms of deficiency or excess, may play a role in influencing "craving," the choice or the timing of drug consumption. We hope to see new initiatives in this area.

Drug Abuse and Reproduction

The NIDA program to study the long-term effects of drugs (both abuse and treatment compounds) on reproduction, genetics, growth, and development will be continued.

Drug Abuse in the Aged

Older persons may not use street drugs, but their use of prescription medication, particularly sedatives and tranquilizers, is significant. The consequences of such use should be studied in the context of aging: effect on immunologic defenses, development of cancer, enzymatic reactions, etc.

NATIONAL INSTITUTE OF MENTAL HEALTH

The National Institute of Mental Health (NIMH) conducts, as part of its program, a wide-ranging and extensive research program in mental health, from the study of neurotransmitters on the micro level to the study of service-delivery systems on the macro level. The following highlights the major areas of need and promise in mental health research.

Disorders of Childhood and Adolescence

The study of psychopathology in children and adolescents has not kept pace with the more recent advances in the field of adult psychopathology nor with the identified incidence and prevalence of these disorders in the population. There is a need for research on the development of a more coherent diagnostic classification of emotional disorders in children and adolescents, on epidemiological information, on biobehavioral etiology, on the effects of various treatment approaches, on the prognosis over time for different groups, and on normal growth and development.

Disorders of Aging

Not only is there a relative paucity of research on aging, but the need for research-based information will increase as the composition of the American population shifts toward rising percentages of older persons. In the area of mental illness of the elderly, additional epidemiologic studies are needed, as well as exploration of promising leads to the etiology of senile dementia (e.g., slow virus, genetic factors, autoimmune response, and neurotransmitters). Considerable research is needed into the etiology, clinical course, differential assessment, and diagnosis of senile dementia and other disorders, notably depression, in the aged. Basic studies in a wide array of fields are also needed. While there have been some significant developments in treatment research, many therapies have not been systematically assessed for efficacy with the elderly. However, systematic findings could be anticipated soon because of the general methodological advances in treatment research. Research on services for elderly persons with a variety of mental health and social difficulties should be emphasized.

Epidemiology and Mental Health Services Research

The need for better data about the extent of mental disorders, how and where they are treated, and at what cost, is great. This kind of research is inextricably bound to research on prevention and on treatment efficacy and therapeutics. Moreover, it is basic to the resolution of issues about underserved or inappropriately served persons and population groups known to have special needs, such as chronic mental patients, children, adolescents, the aging, women, and racial and ethnic minorities. Epidemiological research should emphasize the definition and classification of disorders, the incidence and prevalence of specific mental disorders in the population at large and among particular subpopulations, longitudinal and cross-sectional studies focusing on risk factors associated with the incidence of specific disorders, and studies of the natural history of disorders, both treated and untreated. Health services research should include studies of health systems, health programming, and health behavior, as well as clinical research.

Treatment Research

Two main areas warrant attention. The first is the pressing need to develop an adequate knowledge base about the efficacy of psychotherapy and psychosocial treatments for mental health problems. A cumulative base of research data is needed to address the many unanswered questions about different types of therapy; the relative importance of therapist factors such as experience, training, etc.; and the nature and stability of changes produced via psychotherapy and other psychosocial treatments. The second area is drug therapy—we need information about precise dosage levels, their effects on subpopulations, their adverse side effects, and how to determine the minimal effective dosage levels for individual patients. NIMH should also expand efforts to stimulate research in the broader area of therapeutics: how drugs are actually used by practicing physicians, how physicians are trained in the use of psychoactive substances, how patients comply with drug treatment regimens, whether patients can recognize potentially adverse side effects early, and other questions about long-term maintenance in drug treatment.

Basic Research

NIMH's investment in basic research (about 35 to 40 percent of its research dollar) should be maintained with the additional funds recommended by the President's Commission on Mental Health (PCMH). Many research areas require attention, among them neurotransmitters, genetics, cognitive and emotional functioning of children and the aged, family structure and dynamics, brain chemistry, personality structure, and socioenvironmental factors. Coordination of overall program emphasis with the National Institute of Child Health and Human Development, the National Institute of Neurological and Communicative Disorders and Stroke, the National Institute on Aging, and other National Institutes of Health components will continue to be necessary.

Clinical Research and Mental Illness and Behavior Disorders

Studies of possible causative factors—biological, psychological, and sociocultural—are needed for further understanding of schizophrenia, depression, the neuroses, and behavior disorders. Studies aimed at better diagnostic procedures—especially culturally sensitive diagnostic and assessment instruments—are of paramount importance in many service and treatment issues. Since significantly more women than men are depressed, further research on the social, psychological, and biological precursors and concomitant features of depression among women is warranted. Increased resources for studying childhood and adolescent psychopathology are clearly necessary, for research on the etiology and treatment of childhood and adolescent disorders has been hampered by a lack of consistent approaches to the description and diagnosis of the patient population.

Prevention Research

The Research Task Panel of the PCMH noted that current research on the prevention of mental disorders is largely confined to identifying groups at risk and is ready to move on to the question of determining the importance of specific risk factors. There are many possible research approaches to prevention, and the implementation of a prevention research program will require difficult conceptualization, direction, and management tempered by a recognition of the limits of research in a relatively undeveloped and complex area.

Public Attitudes Toward Mental Illness

In its chapter on "Improving Public Understanding," the PCMH included a recommendation dealing with research on public attitudes toward mentally ill people and toward mental health services and facilities. It called for designing research instruments and methodologies to measure attitudes and behaviors, for developing public educational programs based on the research results, and for assessing the effectiveness of public education programs in creating a climate of community acceptance and support for mentally disabled persons.

Department of Health, Education, and Welfare Food and Drug Administration

The U.S. Food and Drug Administration (FDA) is responsible for ensuring the safety and efficacy of U.S. foods, drugs, cosmetics, vaccines, medical devices, and diagnostic and radiation products. The wide variety of exposures subject to FDA regulation and their direct effect on the public health make it particularly difficult to anticipate all scientific problems of national importance to be encountered by the Agency in the next five years. Often the time between the first announcement of a potential health problem in the research literature and widespread public demand for prompt, but scientifically responsible, regulatory action can be measured in weeks. Despite such difficulties we can already anticipate, on the basis of experience, several important areas in which emerging scientific problems and opportunities are likely to affect FDA regulation in the years ahead. This report briefly describes 26 such problems, divided for convenience into seven broad areas: problems and new directions in bioassay methodology; problems and new directions in toxicology; safety and efficacy of new classes of substances; new demands on analytic chemical methods; radiation, medical device, and diagnostic products; risk-benefit and decisionmaking methodology; and other problems and opportunities.

Of course, these areas are interrelated rather than exclusive, and they are best regarded as part of a continuum of emerging problems, needs, and opportunities. The topics chosen are exemplary, not exhaustive, and other related examples could be adduced. Many of these problems are shared by other agencies involved in screening materials for potential hazards including the Environmental Protection Agency, Occupational Safety and Health Administration, and the Consumer Product Safety Commission. No attempt has been made to organize this outlook as a five-year science plan, since that is not its purpose; thus, neither the length nor the order of individual topics is intended to indicate their current or potential priority.

PROBLEMS AND NEW DIRECTIONS IN BIOASSAY METHODOLOGY

There are some 3.5 million known chemicals, of which roughly 25,000 are produced in bulk in the United States. The rapid expansion of technology, the competiveness of the marketplace, and the rising expectations and demands of the public lead to the introduction of large numbers of new products annually. The protection of the public from the potential hazards of these substances is a major regulatory goal. Since ethical considerations preclude human experimentation and since human latency periods can be 20 to 30 years in the case of cancer, in vivo animal tests are currently the prime scientific means of studying, evaluating, and regulating potentially toxic exposures. It is becoming increasingly difficult to test all such products in a timely and appropriate matter. Crucial problems exist both in the animal tests themselves and in their replacement or supplementation with more rapid in vitro assays.

Availability of Model Animals

Because of their similarities to man, primates are often used—despite their expense and long lifetimes—as human model systems. Their popularity has led to a severe and increasing shortage; the United States used 34,000 primates in 1977 alone. The natural supply of nearly all nonhuman primates is rapidly "drying up," either through actual diminution of populations or by being declared "endangered species." For example, in 1972 India exported 50,000 rhesus monkeys (all destinations); in 1974, only 20,000; and since April 1978, has banned all further shipments of primates. In 1977, India provided 12,000 of the 14,000 rhesus monkeys used in the United States. Efforts to expand existing breeding colonies and implement the proposed National Primate Plan will eventually, but not immediately, help alleviate the problem; furthermore, breeding colony primates cost more than twice as much as imported ones captured in the wild.

The adverse effects on biomedical research are obvious, but there are important regulatory consequences as well. For example, the FDA currently requires both the manufacturer and its own Bureau of Biologics to use 60 rhesus monkeys to assay each lot of polio vaccine. Should current procedures be relaxed to allow joint testing or fewer animals? What would be the consequences on public health? Although the situation is expected to ease in the early 1980s, the development and validation of alternative animal tests (at the least in smaller, more easily bred primates) or in vitro methods is of high priority.

Continuing public objection to the laboratory use of certain other animals, such as dogs and cats, will provide a steady—if less acute—stimulus for their replacement as well.

Development of New Animal Models and Extrapolation to Man

The need for statistically significant results requires the use of 20 to 100 animals in each in vivo bioassay test group at several different dose levels. Sex, species, and route of administration effects can further increase the number of animals required. Thus, experimentation logistics usually limit routine bioassays to rodent (mouse, rat, hamster, and rabbit) models. The relevance of such tests to predicting human reactions and the development of more predictive models are major continuing problems. Each new adverse effect (endpoint) added to present regulatory concerns raises this issue anew. There is little reason to assume a priori that an animal model for chemically induced human cancer will be equally relevant for assessing immunotoxicity, neurotoxicity, asbestosis, bioequivalence, and microwave susceptibility. The Bureau of Drugs, for example, is currently developing a miniature pig model for pharmacokinetics and hopes to develop a pygmy goat model for mother/fetus effects. As the number of recognizable forms of toxicity increases, the problem of developing appropriate models and extrapolation procedures will become increasingly acute. Interspecies comparisons (e.g., computerized correlation analysis) for given stimuli and endpoints can help elucidate both the underlying mechanisms of toxicity and the likelihood of their relevance to man. Indeed, considerably more basic information about human and animal metabolism, biological cycles, etc., will be required to identify the optimal models for a given application.

Known species variations are sufficient to preclude diluting positive findings of toxicity in one species (e.g., mice) by negative observations in another (rats). However, genetically characterized strains of the same species can differ dramatically in metabolism (HSFR/N versus HSFS/N mice), spontaneous tumor rates (C3H/HeN versus C3Hf/HeN mice), etc. Indeed, animals are often specifically bred for such characteristics. The general problem of appropriately combining data from different strains of the same species is a particularly vexing one, especially since local availability encourages investigators in different countries to utilize different strains.

Finally, "extrapolation to man" (in the singular) is somewhat misleading. The heterogeneity of the human population to be protected results in certain subpopulations being at higher risk than others. Depending on circumstances, special animal models for such subpopulations may be required. Relevant "groups-at-risk" may depend upon age (e.g., geriatric/pediatric), genetic (diabetes- or cancer-predisposed), nutritional (vegetarian, malnourished), experiential (smoking, pregnant), and ethical (lack of informed consent) factors. In utero exposures are already widely studied, although the choice of appropriate doses remains a problem. A demand for increased attention to other groups can be anticipated (consider, for example, the increasing number and self-awareness of the elderly). Research on the biochemical basis for the differences between selected subpopulations could also improve the diagnosis and treatment (including more appropriate drug dosages) of such groups.

In Vitro Testing and Structure-Activity Predictions

The sheer number of chemicals in the environment clearly exceeds our ability to effectively test them for potential adverse effects. There are more than 2,000 chemicals in cigarette smoke alone. Animal tests can easily cost more than three years of effort and several $100,000 each. Clearly, comparatively rapid and reliable in vitro tests (involving biochemical or microbiological events highly correlated with in vivo toxicity) and theoretical procedures are needed to handle this ever-increasing burden. At first such procedures would be used as a prescreen to identify materials particularly deserving further in vivo study. Later they could be used as supporting evidence of hazard, especially in restricted, well-studied areas. Although a long-term possibility, the validation of such procedures to the point where they can routinely replace the more costly and time-consuming animal studies is unlikely to be reached during the five-year period considered in this report.

Much current interest focuses on rapid, in vitro carcinogen screening. There are bacterial mutagenicity tests (e.g., the "Ames Test" using *Salmonella typhimurium*), DNA repair tests (in liver tissue culture), cell transformation tests (in mammalian tissue culture), etc. The use of human cell cultures provides direct contact with human materials (human cell surface receptors and markers, enzymes, etc.), but lacks many of the complex interactions (metabolic/immune/hormonal) of whole, living animals. For some materials, such as potentially nongenotoxic carcinogens (promoters, hormones, metal ions, asbestos), the development of in vitro tests may be particularly difficult, if not impossible. The further development, validation, and application of such tests will probably be a top priority for FDA and its sister agencies over the next several years.

Use of the large amount of toxicological data compiled from previous assays and increased knowledge of some of the physical and chemical factors associated with toxicity (e.g., the liver activation of some materials to "classical" electrophilic carcinogens) may allow limited predictions of toxicity based on structural considerations alone. Computerized structure-activity, pattern recognition, and correlation methods can help organize our "experience" and perhaps function as a "positive"-enriching prescreen during the next five years, although their full impact on the "bioassay bottleneck" may not be felt until later (see II FDA Science Symposium, *Structural Correlates of Carcinogenesis and Mutagenesis*, HEW-FDA-78-1046).

Placebo Effect

The conditions of human exposure to potential toxicants differ from those found in current animal bioassays in many ways: the simultaneous exposure to many uncontrolled agents (synergy), the lack of population homogeneity, chronic exposure to low doses, and so on. One of the more mystifying, and at times tragic, is the placebo effect, i.e., changes in the state of a test subject due to psychosomatic effects, rather than pharmaceutical effects related to treatment. Telling a patient that he is being treated with a new and powerful drug can itself affect his progress; a fact well known to both reputable physicians and quacks. Judgments of the efficacy of drugs, especially analgesics and psychoactive agents, are strongly affected by the placebo phenomenon, and in clinical trials care must be taken that the control (placebo) and test treatments are as nearly identical as possible. The investigators themselves must not be told which is which since subtle differences in investigator behavior can be subconsciously registered by the patient. So subtle are the clues involved that even in such double-blind experiments, a statistically significant number of subjects may correctly report whether they were using the drug or placebo.

The placebo effect can be a major public health problem in the case of diseases, such as cancer, with frequent spontaneous remission. One need only consider the tragic controversy surrounding laetrile. Obviously, many taking this placebo will experience spontaneous remission and attribute it to the drug. Others, hearing of its power will feel improved, at least until their developing cancer and deteriorating condition cannot be denied—and often not even then. Since conventional cancer therapy depends upon early treatment, the results of delay are unnecessary suffering and mortality. Far more must be learned about the placebo effect, the subtle sensory clues involved, the mechanisms of its psychosomatic action (including their potential legitimate uses), how to avoid it in clinical experiments, and how to educate the general public of its dangers. The question of animal placebo effects (due to the prior associations with specific colors, odors, procedures, etc.) is largely unexplored.

Recently discovered morphine-like enkephalins in the central nervous system (CNS) could possibly mediate environmentally induced changes in the perception of pain. One could monitor the levels of enkephalin-like substances in human cerebrospinal fluid by using sensitive immunochemical or radioactive binding assays, although the rapid metabolism of the low molecular weight (MW) enkephalins is a problem. Preliminary studies in Sweden, however, suggest that emotional states and pain perception may correlate with the presence of high MW substances (enkephalin precursors?) that can bind to morphine-selective receptors. Sensitive analytical methodology will be needed to study placebo phenomena on the chemical level.

PROBLEMS AND NEW DIRECTIONS IN TOXICOLOGY

Recent advances in analytical methodology can reveal the persistence of potentially hazardous substances in foods, body fluids, and tissue at ever lower concentrations, but the toxicological significance of such low-level exposures (or even the appropriate methods to assess such effects) remains a major problem. Similarly, as we become capable of detecting increasingly complex (e.g., behavioral, immunological) and precise (e.g., changes in the natural levels of structures of specific proteins or hormones) endpoints of biological significance, new demands for protecting the public from the adverse sequelae of such effects concomitantly arise. In vitro and in vivo tests, screening programs, standard regulations, and monitoring programs must be devised and evaluated. The amount of basic information required to generate such protection in a "short" time (i.e., 5 to 10 years from discovery) can be staggering. Another major problem is the development of new toxicological testing procedures for new exposures (inhalation, percutaneous absorption) to older materials with known endpoints. Finally, lack of information on toxicological mechanisms and pathways severely complicates the problem of safety assessment. Often one is not even sure what metabolites or decomposition products to test, whereas it is clearly unfeasible (especially given the current bioassay burden) to test them all.

New Toxicological Endpoints: Immunotoxicity, Neurotoxicity, Behavioral Toxicity

Recent reports that a wide variety of drugs, food additives, and contaminants may affect the immune system have led to FDA concern in the rapidly developing field of immunotoxicity (see IV FDA Science Symposium, *Inadvertent Modification of the Immune Response*, in press). Materials as diverse as gallic acid (a food antioxidant metabolite), carrageenan (a food thickener), aflatoxin (a food contaminant), SO_2, NO_2, ozone, lead, cadmium, and some pesticides are known to affect humoral and/or cellular immune functions in vivo and in vitro. Several such agents are already under study by the Bureau of Foods. The next five years should see considerable progress in clarifying the extent and severity of the problem, in developing appropriate in vivo and in vitro assay systems, in elucidating the actual health significance of a given observed test result, and in developing regulatory programs for such materials if necessary. There is currently considerable dispute in each of those areas. The FDA National Center for Toxicological Research (NCTR) is developing methodology for assessing hypersensitivity, immunosuppression, tumor immunity, autoimmunity, and immunocompetence during pregnancy and lactation.

Immunotoxicological research is also expected to provide several useful opportunities. First, in some cases, the immunologically detectable effects are so sensitive that they precede all other toxicity markers, i.e., they define the lowest level of biological response. As usual, the development of such sensitive methods is not without thorny regulatory implications. Second, immunological assay methods may result in new diagnostic tests. In some cases, toxic (or natural, disease-induced) lesions may expose normally sequestered haptens (e.g., the appearance of migration inhibiting factor, MIF, after myocardial lesions) which might be useful for post facto diagnosis, or even in detecting potential problems before the development of overt clinical symptoms. A battery of tests for known products of lymphocyte activation is being carried out in drug-treated laboratory animals by the Bureau of Drugs. At present, comparatively little is known about the potential immunotoxicity of veterinary drugs, feed additives, and other exposures on food-producing animals, an area of potential—if unknown—future problems.

There is also increasing information on chemicals that can attack the nervous system to adversely affect behavior, development, and neurological function. The complexity and subtlety of such effects demands a multidisciplinary approach including behavioral analysis, neurophysiology, neuropathology, and neurochemistry as well as classical toxicological and statistical considerations. One major problem is the development of sufficiently simple and reliable models, perhaps based on neurochemical or physiological indicators, that can predict such effects in man and that are suitable for routine screening.

The FDA and the Environmental Protection Agency have established a joint program in neurotoxicity that should make valuable progress in this area during the next five years. This program will systematically collect information on such neurotoxicants as heavy metals, pesticides, nonionizing radiation, and oxidants; determine the sensitivity and cost effectiveness of existing test methods; develop new short-term tests based on our increasing knowledge of neurotoxicant mechanisms and validate such tests using current neurobiological and physiological techniques; and compare health effects in several species (including human epidemiological studies) to evaluate interspecies extrapolation.

Research in behavioral teratology is also expected to grow, both because of its own importance and as a potential sensitive indicator of developmental toxicity. Latent effects following such exposures in developing young animals and humans may be more likely than in adult exposures. Again, improved methodology is needed for screening purposes.

New Routes of Exposure: Inhalation and Percutaneous Absorption

Each route of exposure to a toxicant presents novel features. Ingested materials may be metabolized in the gut, intravenously injected materials reach the blood stream unaltered, implanted materials may present high local concentrations and reach otherwise inaccessible tissues, inhaled particulates depend on their aerodynamic and mechanical properties to reach their targets, and topical agents depend on their percutaneous absorption. Only recently has attention focused on the last two routes, which are comparatively less developed and present special intrinsic problems. For example, the accurate determination of dosage is a major problem in assaying the toxicity of topically applied products such as hair-dyes, cosmetics, and ointments. Animal models are needed to predict the extent and rate of penetration of the human skin barrier by such agents and their eventual concentrations in underlying tissues and in the blood.

Assaying the inhalation toxicity of particulate matter provides another example of dosage problems. Respirable asbestos fibers represent a major carcinogenic exposure, which depends critically on fiber length and diameter. However, the dose introduced into a sample chamber may bear little resemblance to the actual ambient material in the chamber, the respired material, the lung-retained material, or the material transported to and lodged in various body tissues in either concentration, size-distribution, or aspect ratio. Careful monitoring of these parameters by physical (air sampling followed by electron-microscopic analysis) and biological (occasional sacrifice, autopsy, and analysis of mineral fibers lodged in tissues) means is required. There are remarkably few good health effects studies using well-characterized samples and dose-exposures. Indeed, the methodology for the collection, preparation, and electron microscopic analysis of asbestos fibers is in a state of rapid, but as yet incomplete, development (see I FDA Science Symposium, *Electron Microscopy of Microfibers*, HEW-FDA-77-1033). We expect appropriate regulatory methods to become available within the next five years and that substantial progress will be made in automated size, aspect ratio, and chemical (using energy dispersive x-ray analysis) measurements.

"Realistic" Exposures: Chronic, Low-Dose Toxicity and Synergy

Appropriate dosing is a major problem in toxicity bioassays, and a linear dose response is not always encountered. Very low doses may not sufficiently activate certain body defense or clearance mechanisms or metabolic pathways, whereas overly large doses may saturate them, bringing mechanisms and pathways (with their associated products) into play that would not be activated under more "normal" exposures. Thus, one cannot always assume qualitative equivalence between high and low dose responses. Cost considerations often focus routine screening on relatively acute or high-dose studies, but the next five years should see increased demand for more complete chronic, low dose measurements. Since, to maintain statistical significance, large numbers of animals are needed to detect toxicity at low dose levels, improved models and statistical procedures are needed for extrapolation to low dosages.

The recent NCTR experiments on the dose-response of chronic exposure to 2-acetylaminofluorine (2-AAF), a potent carcinogen, was an important first step in this direction. A linear response was found for liver tumors, but not for bladder tumors, which display an initial dip in dose-response. The experiment required 26,000 animals, and its lowest low-dosage group corresponded to only 1 percent incidence. Even so, the doses used far exceeded typical human low-dose exposures; a 1 percent total incidence would correspond to 2 million cancer fatalities in a total population of 200 million. Since few manufacturers could mount tests of this size, the problem of the appropriateness of high dose experiments will remain a controversial topic during the next five years.

Methods to identify and evaluate the effects of cocarcinogens and promoters are also sorely needed. Many noncarcinogens, i.e., non-"initiators," can greatly influence carcinogenicity by acting as "potentiators," "activators" and "modifiers." A major Agency problem will be to more clearly define the relative roles of such agents in carcinogenesis and their regulatory implications.

Another reality factor is that humans are rarely exposed to a single, pure agent. Our environment contains thousands of compounds that interact chemically with each other and can have synergistic effects on biological processes. In particular, methods to identify and evaluate the effects of cocarcinogens and promoters are sorely needed. Although more rapid, automated in vitro tests could help, clearly "brute force" methods cannot succeed. To study the synergistic effects of just 500 agents would require 124,750 pair-wise tests, 20,708,500 triplet tests, etc. Interactions of interest have to be predicted beforehand from theoretical considerations and from new information generated on the behavior of selected representative classes of materials. The materials to be considered will also strongly depend on the outcome of metabolic and fate studies.

Fate Considerations: Metabolism, Transformation, and Transport

We are becoming increasingly aware that the metabolism of chemical agents in the body or in the external environment (e.g., soil bacteria) and their transport (e.g., to more sensitive targets) and transformation by light, oxygen, and other chemicals in the environment can greatly modify their biological effects. For example, the procarcinogens vinyl chloride, benzo(a)pyrene, benzo(a)anthracene, and aflatoxin B_1 apparently must be converted to active, alkylating epoxide forms (see II FDA Science Symposium, ibid.). This severely limits the indiscriminate application of in vitro activity assays, since until the appropriate activation systems are included, only the relatively inactive parent compound will be tested. Far more needs to be known about the chemical structures and interactions of the animal, plant, and bacterial metabolites of ingested and inhaled chemicals and industrial wastes, and about their transformation, bioaccumulation, and transport in the environment (the varying kinds of industrial phosphate esters in edible fish, the binding of amines and halogenated hydrocarbons to humic acids in the soil, the bioaccumulation of heavy metals in shellfish, for example). The need to test increasing numbers of recognized potentially hazardous metabolites places impossible burdens on in vivo bioassay procedures. In vitro procedures, although more rapid, require foreknowledge (or fortuitous inclusion) of the relevant activation mechanisms to be effective. The collection of more basic knowledge on the metabolism of compounds of interest is needed to improve screening procedures, and automation and computerization should be pursued where possible.

Short-lived metabolites, those produced in small amounts (although perhaps highly potent and/or concentrated locally), those already present in the body (endogenous materials), those highly complex in structure, and those highly variable in concentration (depending on precise circumstances), present special problems. New scientific and regulatory approaches are sorely needed. Better analytical methods may also be needed, as with radiopharmaceuticals such as those containing technetium 99, whose chemical transformations are poorly understood. Similarly, better methods are needed to collect and identify metabolites in blood, saliva, urine, and other clinical materials.

SAFETY AND EFFICACY OF NEW SUBSTANCES

The FDA is responsible for assessing the safety and efficacy of new drugs, vaccines, medical devices, and diagnostic products. Although it is obviously not possible to predict all of the new products to be encountered in the next five years, a few typical anticipated examples are given below. In addition, there are classes of potential products for which efficacy itself must be better defined. For example, how would one define and test potential geriatric products for retarding aging? How can we objectively evaluate psychoactive agents for efficacy against psychological disorders such as schizophrenia? Relevant biological parameters and statistical procedures must be further developed.

Vaccines

Recent progress in fundamental infectious disease research suggests numerous opportunities for the development of new viral and bacterial vaccines and for the improvement of current ones. Improved test procedures and animal models are anticipated for Types A, B, Non-A, and Non-B hepatitis. Improved, inactivated influenza vaccines (both live and attenuated) and vaccines against varicella, herpes, and cytomegalovirus are also expected. Pneumococcus and meningococcus vaccines can and should be improved and extended to additional strains. Pertussis vaccines can be improved. Particularly exciting are the potentials for subunit vaccines of extremely high specificity, purity, and degree of chemical characterization. (Most current vaccines are created from complex, partially characterized materials.) Carrying this process farther, one can attempt to develop and characterize synthetic subunit vaccines in which only the precise molecular component(s) responsible for the desired immune effect are present.

Antiviral Materials

There is growing optimism that drugs effective against a variety of viral diseases will be discovered in the next five years. Adenosine arabinoside (Ara-A, Vidarabine) has already been approved by the Agency for use against herpes keratinitis and encephalitis. Assay methods for testing and certifying such agents must be developed, since current in vitro assay methods for antibacterial and antifungal drugs are not directly applicable to viruses. The outlook is also good for significant progress in developing broad-spectrum agents such as interferon and interferon-inducing substances.

Polypeptides

Many polypeptide hormones (insulin, oxytocin, ACTH) are already in clinical use, and new biologically active polypeptides (enkephalins, thymosins) are constantly being developed. Some are small enough to be synthesized directly. The enkephalins are endogenous polypeptides in the central nervous system, which can bind to morphine-receptors (or rather vice versa) and which affect perception of pain. Considerable progress is expected in developing similar substances into non-addictive analygesics.

The thymus gland controls the development and function of the immune system by secreting polypeptide factors called thymosins, which help regulate the formation of and differentiation of subpopulations of T-lymphocytes. Several thymosins have already been sequenced, and synthetic thymosin-alpha-1 is now undergoing clinical trial as an adjunct to cancer chemotherapy. Some thymosins may induce helper and suppressor T-cell activity. T-cell maturation may be accompanied by increases in membrane fluidity which may facilitate antigen or mitogen response. Fluorescence and electron spin resonance (ESR) techniques are being developed to study these changes, and the activation of membrane-bound enzymes is being developed as a rapid indicator of thymocyte activation by antigens. We anticipate increasing interest in, and development of, thymosins, thymosin analogs, and other chemicals affecting immune function.

Molecular Biological Opportunities and Risks

The last decade has seen a revolution in our understanding of the molecular basis of genetic control and our ability to deliberately manipulate such processes and the genetic materials DNA and RNA themselves. In particular, recombinant DNA techniques offer the potential to produce (molecularly "clone") large amounts of hitherto scarce genetic and polypeptide materials by inserting appropriate DNA sequences into easily replicated hosts such as bacteria. Such techniques should greatly accelerate progress in such areas in immunology, developmental biology, and cancer research, although fields as diverse as agriculture (e.g., addition of nitrogen-fixation ability to food crops) and drug certification (e.g., the development of improved mutants for microbial assays of antibiotics) may also be affected. Certainly, the FDA can expect to receive requests to certify drugs, biologicals, and perhaps foods created using recombinant DNA or plasmid-transfer technology. Already, mammalian insulin, somatostatin, and growth hormone have been successfully produced using *E. coli* bacteria as hosts.

Agency decisions on the safety (for potential hazards, see 41 FR-27902, 41 FR-38425, 42 FR-49596, 43 FR-33042) and efficacy (e.g., will human insulin produced by bacteria have the same effects and side effects as natural bovine or porcine insulin?) of such products will require considerably expanded expertise and facilities in this rapidly developing area. The Agency has already issued an "Intent to Propose Regulations" (44 FR-60134, December 22, 1978) that would "require that any firm seeking approval of a product requiring the use of recombinant DNA methods in its development or manufacture demonstrate the firm's compliance with the requirements of the National Institutes of Health (NIH) Guidelines" in effect at that time. The NIH Guidelines would also "be incorporated in good manufacturing practice (GMP) regulations." We anticipate, however, that both manufacturer and consumer advocates will point out significant differences between laboratory and manufacturing needs and environments.

NEW DEMANDS ON ANALYTICAL CHEMICAL METHODS

Some agents (e.g., the dioxins) are known to produce toxic effects even at extremely low exposure levels—parts per billion (ppb) or trillion (ppt). The development of sufficiently sensitive methods often taxes the state of the art in chemical analysis. Conversely, in the case of contaminants impermissible at any detectable level, the development of new, more sensitive assays raises complex regulatory issues. Some selected examples of these problems are mentioned below.

Characterizing and Monitoring Contaminants

Rapid, inexpensive, preferably automated, multiple-residue assays are needed for the determination of veterinary drug residues and their metabolites in food-producing animals before the carcass enters the food chain. More information is required on environmental factors, tissue binding, and health effects. Similar methods are needed to detect pesticide residues, industrial contaminants, mycotoxins, etc., in foods, especially at low levels (e.g., chlorinated dioxins and dibenzofurans in milk and fish, aflatoxin in grain, paralytic shellfish poison). In some cases (fish-borne ciguartera poisoning), the active toxins have yet to be isolated and structurally identified. In others (decomposition of fish and shellfish products), organoleptic methods are still in use, due to the lack of adequate analytic technology.

Endotoxins are examples of microbial contamination problems; although sterilization procedures such as heating and ultraviolet radiation kill bacteria, they do not remove the dead microbes or associated products. Regulatory techniques for monitoring viral contamination of the food supply are virtually nonexistent, while those for cholera and many other microorganisms are marginal.

Characterization of Regulated Products

The Bureau of Drugs and the Executive Director for Regional Operations (EDRO) are currently reviewing and improving current compendial (USP and National Formulary) monographs through the application of modern analytical techniques to determine purity and quality. The 70 U.S. antibiotic standards are also being subjected to full physical and chemical characterization, and one preliminary monograph has already been published (see "Amphotericin B," in *Analytical Profiles of Drug Substances*, Vol. 6, K. Florey, Ed., Academic Press, 1977). The characterization of the 27 classes of over-the-counter (OTC) drugs is another goal, although the complexity of their formulations may require chromatography, mass spectrometry, computer analysis, and other sophisticated techniques. The development and characterization of standard materials for clinical chemistry are another recent challenge; our Agency will be cooperating in this area with the National Committee on Clinical Laboratory Standards (NCCLS) and other national bodies. We also need to develop analytical methods for certifiable color additives and their metabolites in foods, and to prepare (and characterize) such compounds in sufficiently high purity to serve as standards and reference compounds.

Applying New Analytic Technologies

Recent advances in analytic technology are being increasingly applied to regulatory problems. Computerized mass spectrometry (including field desorption), proton nuclear magnetic resonance (NMR), ^{13}C-NMR, and Fourier transform infrared spectroscopy (FT-IR) are available and in increasing demand. Laser Raman spectroscopy (LRS), visible (e.g., for food colors) and ultraviolet resonant LRS, electron spin resonance (ESR), and x-ray structural determination facilities are available elsewhere. Increasingly sensitive thin-layer (TLC), high-pressure liquid (HPLC), gas (GC), and gas-liquid (GLC) chromatographic methods for a wide variety of materials are becoming available. Sensitive immunological methods, such as radioimmunoassay (RIA), specific antibody fluorescent tagging, and enzyme-linked immunosorbent assay (ELISA) techniques, will also be increasingly developed and utilized for regulatory purposes over the next five years. The identification of microscopic particulate contaminants (such as asbestos microfibers) presents special challenges that may be met by energy dispersive x-ray analysis (EDXA), selected area electron diffraction (SAED), and/or micro-LRS techniques.

Some recent Agency successes include the structural determination of the fungal toxins, satratoxins G and H, and xanthoviridicatin D using ^{13}C-NMR, and aflatoxin and vitamin analyses using HPLC followed by mass spectroscopic confirmation of the isolated fractions. Advanced new instrumentation for specific analyses (e.g., a nitrosamine analyzer using GC followed by oxidation of trapped fractions to light-emitting excited NO_2 gas) should also become increasingly available. Furthermore, increased automation, computerization, and advanced data collection (e.g., optical multichannel analyzers) should greatly speed current analytic procedures.

Such rapid progress also creates regulatory problems. Should regulations be changed to keep pace with current capabilities (should insect fragment standards based on old image-analysis methods be replaced by newer ones based on quantitative criteria such as RIA)? If so, how? What happens when increased sensitivity of methods suddenly changes the lowest detectable level of a contaminant? How can one unambiguously define a "no detectable residue" level, i.e., "zero"? (The corresponding toxicological question is: how can one unambiguously define a "no detectable effect" level?) These complex issues are already being faced in the case of potentially carcinogenic residues in the edible products of food-producing animals (42 FR-10412), and more such problems can be anticipated.

RADIATION SAFETY

The next five years should bring increased pressure to reduce unnecessary exposure to clinical x-rays and other sources of ionizing radiation and to better understand and protect the public from the adverse health effects of nonionizing radiation (e.g., microwaves, ultrasound, ultraviolet and visible radiation).

Reducing Unnecessary X-ray Exposure

There are both human and technical factors in unnecessary clinical exposure to x-rays. We need to better understand the factors that cause physicians to order (and patients to accept) diagnostic x-rays, to perform clinical tests to validate the efficacy and safety of alternative patient referral criteria, and to study the effects of credentialing on radiation technologist performance. We must further examine the efficacy and cost of various imaging modalities (conventional x-ray; CAT scanning; nuclear medicine; positron-, single-photon- and ultrasonic-imaging) in diagnosing specific disease states. The next five years should also see improvements in the use of new rare-earth phosphors, better scatter rejection methods (e.g., replacing grids with scanning devices), optimally filtered x-ray beams, and improved image processing and display. New directions include the development of positron imaging and single photon tomographic procedures, which allow the use of short-lived isotopes (lower patient dose) and provide quicker response time, depth-sensitive imaging, and high resolution images. The health effects of other clinical procedures, for instance, the biological consequences of the neutrons produced incidentally by high energy accelerator beams used in radiation therapy, must also be elucidated.

Health Effects of Nonionizing Radiation

Ultrasound imaging has developed from a laboratory curiosity to a widely accepted diagnostic tool. Pulse-echo ultrasound techniques are used to visualize soft tissue structures, such as the human heart or fetus. Doppler ultrasound equipment is used to detect fetal movement in utero. Even acousto-optical holography is being developed for biomedical applications. The rapid growth is resulting in increased public exposure to ultrasound, and potential hazards—especially to fetuses or sensitive internal tissues—must be more thoroughly explored (see FDA/BRH Workshop, *Interaction of Ultrasound and Biological Tissues*, HEW-FDA-73-8008). Many early studies used intensities too high and endpoints too gross to be of use in evaluating diagnostic exposures. The effects of continuous versus pulsed exposures, temporal peak versus temporal average intensities, and the stage of the patient's development (e.g., fetal, postpartem) must be carefully considered. Reduced fetal weight and delayed neuromuscular development have been observed in rats and mice in recent experiments. New equipment and regulatory methods must be designed to monitor acoustic field patterns and intensity levels from pulsed ultrasonic sources, and imaging techniques should be improved to minimize patient exposure.

More information is needed on the health effects of microwave exposure, including elucidating the underlying mechanisms, kinetics, and critical thermal levels in various tissues (see FDA/BRH Symposium, *Biological Effects and Measurement of Radiofrequency/Microwaves*, HEW-FDA-77-8026). Better theoretical models are needed to predict the internal fields ("dose") in man from external near-field measurements.

The combination of visible, ultraviolet (UV) or infrared (IR) radiation and certain chemical agents can have synergistic effects in susceptible individuals. Such effects account for both photochemical therapy (e.g., for psoriasis) and phototoxicity. We must learn more about the frequency dependence ("action spectrum") of radiation-induced viral activation, carcinogenesis, retinal aging, modulation of error-prone repair, etc. (see FDA/BRH Symposium, *Biological Effects and Measurement of Light Sources*, HEW-FDA-77-8002). We also need to know more about chemical photosensitizers and their prevalence in the environment. Many conventional light sources, such as fluorescent lamps, provide chronic, low-level exposure to ultraviolet radiation which might adversely affect some photosensitive individuals. The increasing use of laser products emphasizes the continued need for their intelligent regulation.

RISK-BENEFIT AND DECISIONMAKING METHODOLOGY

Making regulatory decisions under conditions of partial information (uncertainty) is a difficult, if necessary, task. Sophisticated statistical methods and objective scientific procedures are needed to define, analyze, and solve such problems.

Development of Statistical Definitions and Risk-Benefit Procedures

The regulatory use of malleable phrases like "proof of safety," "proof of lack of safety" and "lack of proof of safety" can lead to needless confusion and wrangling on what, in fact, constitutes adequate proof. Such terms must gradually be redefined in terms of acceptable levels of certainty (size of the α and β regions for Type I and Type II errors) and the requisite sensitivity of tests (the minimal detectable difference for a given α, β) for a variety of alternatives to the null hypothesis. The designing of appropriate statistical methods for decreased latency in the absence of increased incidence (crossing hazard functions), extrapolation to low "realistic" doses, interstrain and interspecies comparison, etc., remain important, controversial problems.

Risk analysis raises a wide variety of technical (e.g., accounting for metabolites, synergy), practical (one cannot test everything, what is "zero"), and ethical (informed consent is inapplicable to saccharin-consuming minors) problems. The prospect that current regulations may be expanded to include benefit considerations raises additional issues (see III FDA Science Symposium, *Risk/Benefit Decisions and the Public Health*, in press). Who should be polled as to the perceived benefits: those at risk, those who "know," or those who regulate? And what mathematical techniques are available for expressing results not readily reduced to a single metric? Finally, a retrospective study should be done on the effectiveness of previous risk-benefit analyses. Which kinds of studies and techniques have been of use? When can such official-looking studies actually gloss over uncertainties and produce an unwarranted, unrealistic confidence in "scientific" predictions based on inadequate data?

Baseline, Consumption, and Incidence Surveys

To correctly assess the impact of an agent on public health, one must know more about its ambient, endogenous, and consumption levels; about the baseline incidence of its reported effects in specific subpopulations; and about predicted changes in such parameters. For example, nationwide background levels of cadmium in specific agricultural products, the corresponding levels for sludge-treated or filled cropland, and the consumption patterns of such products in specific subpopulations (e.g., infants versus adults) would be useful. The occurrence and significance of estrogens in foods (e.g., plants) is another example. The concept of total diet exposure is becoming increasingly important in developing new regulations, and total diet studies for pesticides, industrial contaminants, and nutrients can be anticipated.

OTHER PROBLEMS AND OPPORTUNITIES

Although this report cannot include all relevant items, the following additional topics deserve mention.

Bioavailability and Bioequivalence

Given the increasing emphasis on generic drug substitution to offset rising health-care costs, the Agency is faced with assuring that the substitutes are equivalent in their therapeutic potency and pharmacokinetics. Differences in tablet dissolving rates, vehicle effects, isomer ratios, etc., can affect the time course and levels of dosage in target tissues. Improved physical-chemical, pharmacokinetic, and clinical models, and analytical methods are needed. Potential causes of such differences should be elucidated. For example, the vigorous grinding of drug powders can convert them from a crystalline to an amorphous state as measured by powder x-ray diffraction; the effects of such processing-induced changes on bioequivalence should be further studied.

Nutrition

The relationship of nutrition and nutritional requirements to human development and disease must and will be further elucidated. The positive and negative effects of such over-the-counter "remedies" as megavitamins, liquid protein supplements, and herbal preparations must be dispassionately examined. Practical indicators of nutritional status, "normal" dietary patterns for specific subpopulations, the effects of synergy between nutrients, and interactions between nutrients and environmental contaminants should be studied. The development of modern, reliable methodology is a necessity.

Biocompatibility and Nondestructive Evaluation of Medical Devices

The rapid development and expanded use of such implanted medical devices as heart valves, pacemakers, and artificial knee joints have created the need for additional information on the biocompatibility, chronic toxicity, and carcinogenicity of materials used for implants. More must be learned about the leaching and local concentration of potentially hazardous substances. Improved in vivo and in vitro test methods must and will be developed. Localized stresses and failure modes, and noninvasive (or at least nondestructive) methods to detect and evaluate potential defects in implanted or external life-supporting systems also require further study.

Diagnostic Products

The past few years have seen a dramatic rise in sophisticated clinical and diagnostic products. Improved validation methods, surrogate materials (to replace human specimens), and clinical reference standards require further development.

Control Technology

Improved field surveillance methods are needed to detect food decomposition and bacterial, viral, or insect infestation. For example, current organoleptic evaluation of fish may be inadequate, since modern processing techniques can remove or obscure the usual signs of spoilage (odor or discoloration). New thermal and irradiation sterilization of foods and packaging materials must be evaluated, and migration of potentially harmful substances from food packaging explored both theoretically (e.g., diffusion kinetics) and experimentally (radioactive tracer studies). Improved procedures are needed for characterizing natural variations in food quality and nutrition, and for analyzing consumer reactions to labeling.

Membranes and Membrane Receptors

The mechanisms of action of many important pharmaceuticals, toxicants, and immunology products are poorly understood, although their effects on membrane fluidity, permeability, and/or surface receptors are believed to be involved. For example, monoclonal antibodies to placental insulin receptors could be used to purify insulin receptor preparations, to study receptor internalization (postulated to mediate the growth effects of insulin), to search for postulated insulin receptors in the cell nucleus, etc. Insulin may also act directly on ion transport mechanisms—another hypothesis to be tested. These studies could lead to improved in vitro methods for the evaluation of animal insulin, recombinant DNA produced human insulin, and new insulin analogs.

Science Planning and Emergency Response

The increased complexity of government research needs, the interrelationships between research tasks, and the need for highly specialized facilities create a pressing need for more effective planning, cooperation, and coordination within the FDA and between the FDA and the other regulatory and health research agencies. Furthermore, research directions must be thoroughly coordinated with regulatory needs. Two recent initiatives, the formation of the Interagency Regulatory Liaison Group (IRLG) (see 42 FR 54856, October 11, 1977) and the National Toxicology Program (NTP) (HEW Publication, "NTP Establishment Document," November 15, 1978), are respectively designed to serve the national needs for coordinated regulatory actions and cooperative research efforts for the furtherance of the public health.

The need to coordinate regulatory needs and research directions is made clear by an example drawn from the case of asbestos. If an asbestos standard is to be based on current state-of-the-art detection capabilities ("no detectable asbestos content"), a methods research program will be required. If, instead, the standard is to be based on ambient conditions ("no more than X times the naturally occurring background"), surveillance methods and a nationwide incidence research program are needed. If, in fact, the standard is to be based on adverse health effects—the most reasonable, but more difficult, approach—a research program is needed to examine the effects of fiber length, diameter, and chemical composition on health. A simple calculation will demonstrate the sheer impracticality of obtaining a detailed map of the response space using in vivo studies alone, so in vitro correlates of in vivo response must be studied and developed into validated methods for extrapolating between in vivo determined points. Better procedures are also needed to scientifically exploit emergency situations, such as the polybrominated biphenyl (PBB) contamination of Michigan. Not only can rapid scientific response help alleviate the problem, but it can provide invaluable information in its early stages to improve our country's ability to minimize the adverse effects on future occurrences or of other human exposures.

Department of Health, Education, and Welfare National Institute of Education

The fundamental goal of the National Institute of Education (NIE), derived from its legislative charter, is to aid through the support of scientific inquiry into the educational process, the attainment of equality of opportunity to receive an education of high quality. In pursuit of its fundamental goal, NIE is concerned with human intellectual functions and learning and how they vary among individuals, ways that individuals can be helped to learn in school and other settings, means of testing and assessing teaching and learning, effective organization and institutional practices in education, educational finance, and the utilization of results of research and development to improve education.

Recent progress in three areas of science and technology is making strong contributions to research and development programs in education and is expected to have important effects. *Cognitive science* is providing new understanding of human intellectual processes, which allows better design of instruction and testing and better fitting of them to cultural background. *Information-handling technology* is in a period of revolutionary increase in capability and lowering cost, which will make feasible greatly increased individualization of instruction and testing. *Organizational science* is giving us new understanding of processes occurring in organizations that will facilitate improvement of schools and colleges through better flow of useful innovations and reduction of dysfunctional practices. These three areas are discussed in detail in the following paragraphs.

COGNITIVE SCIENCE

Cognitive science—the science of human mental process—is a rapidly growing interdisciplinary movement including psychologists, computer scientists, philosophers, anthropologists, linguists, and educators. It is characterized by a concern with understanding the mechanisms by which human beings carry out complex intellectual activities, including learning, comprehending, problem-solving, and creation of new ideas, and how these activities are influenced by the individual's previous experiences.

Glaser (1978) has pointed out, "At the present time, modern cognitive psychology is the dominant theoretical force in psychological science as opposed to the first half of this century when behavioristic, antimentalistic stimulus-response theories of learning were in the ascendance." He goes on to say that "the internal cognitive world of complex human behavior involved in thinking, problem-solving, acquiring understanding of various domains of knowledge, and personal expectations and self-development . . . which have been only peripheral aspects of . . . behaviorist psychology, now are the starting points for cognitive psychologists." Newell and Simon (1972) describe this change as follows:

> For several decades psychology . . . focused on learning, lower organisms, and tasks that are simple from an adult human viewpoint. Within the last

dozen years a general change in scientific outlook has occurred. One can date the change roughly from 1956: in psychology, by the appearance of Bruner, Goodnow, and Austin's *A Study of Thinking* and George Miller's "The magical number seven"; in linguistics, by Noam Chomsky's "Three models of language"; and in computer science, by our own paper on the Logic Theory Machine.

This shift in emphasis is very important for education. In the past, many teachers and persons in the disciplines concerned with the substance of education were unable to take seriously behaviorist psychology's seemingly simplistic view of what is to be learned and how it is to be learned. The new cognitive psychology accords much more closely with common views of teaching and learning, while at the same time providing a scientific basis for improving our understanding of these processes and thus for improving education.

The following quotation from *Linguistic Communication* (Miller, 1973), a report prepared for the National Institute of Education, expresses well the ways in which this improvement can be achieved in reading and mathematics education.

> There is no intrinsic difficulty (although there may be many practical and technical problems) in testing or measuring whether people can comprehend oral or written messages. Comprehension is tested by determining whether people, after receiving a message, can answer questions about its contents or implications, or can carry out its instructions. But there is a long leap from measuring comprehension of a message to discovering the thought processes a person must go through in order to achieve that comprehension.
>
> Standard forms of school instruction in reading and language arts do much to facilitate the growth of language skills in elementary and secondary school children as they mature. These instructional methods are based on a wealth of practical experience in teaching, but they are probably most effective in handling average, normal children, because it is such children to whom most of our experience relates. Standard instructional programs are certainly not based on any extensive analysis of the processes underlying the skills, because such analysis still remains largely to be carried out.
>
> With a better theoretical grasp of the reading comprehension process to supplement conventional wisdom, we could devise more powerful diagnostic tools to identify with accuracy the weak and strong points of individual students—the processes in which they are competent, and those in which they need additional help and training. With better diagnosis, our deeper understanding of process can also lead us to more effective teaching and learning techniques than those we now possess.
>
> During the past decade there has been a great burst of new research activity in psychology and related disciplines aimed at understanding complex human thought processes, including processes for comprehending language and problem-solving processes. One reason for this burst of activity, and the consequent progress that has been made, is the availability of new tools of research—cameras for detecting and recording eye movements, and digital computers, to mention just two examples. These tools have provided us with new ways to identify successive steps in human thought processes, and to formulate models of those processes that will permit their operations and pathologies to be explored and tested. Studies of successive foci of attention in reading tasks, using eye-movement cameras and thinking-aloud protocols, should continue to expand on our knowledge of these processes.
>
> How language conveys information from one person to another is no longer an unpenetrated mystery. Enough is known about how language carries meanings to permit us to write programs for computers that accept instructions in natural language, and other programs that solve the problems posed by those instructions and carry out the tasks they set. These programs "understand" language by the same criteria we use to assess human understanding—they can answer questions on the basis of what they have read, they can carry out instructions, they can solve simple problems. Moreover, these programs are open to analysis, so we can study exactly what processes enable them to do these things, and compare their processes with the behavior of humans to whom the same tasks are presented.
>
> Computer simulation of human cognitive processes is one of several promising techniques for modeling the methods used by people to perform complex tasks. It is well adapted to diagnosis of individual differences in performance and to tracing the differences in process that underlie them. For example, in learning algebra, problems posed as stories—so called algebra word problems—are a matter of special difficulty for many students. Use of a computer program to model the processes of interpreting such problems, and comparison of the model with students' behaviors, reveals that there are (at least) two strikingly different styles for tackling these tasks. One style is based on a literal, grammatical translation of the problem statement; the other on conceiving a real-world situation that corresponds to the problem statement. Most students use one or the other of these processes almost exclusively and appear largely unaware of the one they don't use. The ablest students, however, tend to use both processes in reinforcement of each other. This kind of process model can now be developed further as a tool to diagnose individual stu-

dents' difficulties with word problems, and as a framework for developing any additional skills that students are found to lack.

CENTER FOR THE STUDY OF READING

Applications of cognitive science to education are being carried out by NIE through its Program on Teaching and Learning and by other agencies. NIE now supports a Center for the Study of Reading (CSR) created jointly by the University of Illinois at Urbana and Bolt Beranek and Newman in Cambridge, Massachusetts. NIE is also supporting a number of individual cognitive science research projects in reading and mathematics learning. In addition, NIE and the National Science Foundation (NSF) have a joint program supporting projects in which persons from cognitive psychology collaborate with persons from one of the scientific disciplines in studying learning and teaching of that discipline.

The Center for the Study of Reading is giving major attention to such areas as the following:

The Role of Background Knowledge in Reading Comprehension

Comprehension is now seen as a fitting of what is being read to the content and conceptual structure of the reader's existing knowledge rather than an isolated assimilation of new information.

Prior Language Experiences

The Center is attempting to understand cultural differences in children's experience with language prior to the middle grades. Because earlier research on surface (e.g., phonological) differences has revealed little about cultural differences in reading comprehension, the Center is examining higher level aspects of language—structural properties of conversations, conventions of usage, and expectations about the social interactions that appear in stories. It seeks to identify specific features of educationally advantaged homes—such as reading to preschool children—that may contribute to later success in reading comprehension.

Reading Strategies and Tactics

Teaching children to monitor their own comprehension and to take action when they realize they do not understand, a characteristic of skilled readers, is proving to be a powerful instructional technique. The Center continues to identify specific strategies children can use to aid in directing their attention to the most important content, recovering from outside events that disrupt their concentration, and questioning their own understanding. This work envisions direct instruction in high level comprehension processes and study skills. It is believed that these advanced skills may be the ones that account for major cultural differences in student achievement.

EDUCATIONAL TESTING

A 1978 NIE conference on research on testing pointed out that the essential form and structure of educational testing has changed little in over 50 years despite the marked changes that have occurred in psychological knowledge and technology in that half century. They recommended an extensive national research and development program designed to increase the appropriateness of testing and its value in the instructional process. As a first step, NIE has called for research proposals in several areas described below.

Use of Tests as Teaching Instruments

Attention will be placed on two specific topics: application of cognitive models of the knowledge structure of various subject matters and of learning and problem-solving to construction of tests that identify processes underlying test answers, analyze errors, and provide information about what students know and don't know; and strategies for integrating testing information with instructional decisions.

Multiple Forms of Assessment

Research will focus on several needs; among them are procedures for assessing skills as they are displayed in various performance situations, detailed analyses of specific skills requirements for school tasks, multiple forms of testing that capitalize on advances in computer technology and research on cognitive processes, and techniques for presenting the results of multiple assessments in meaningful ways.

Cultural and Linguistic Factors in Testing

Many cultural factors that affect test interpretation and performance will be studied, including the linguistic demands of tests used in education, effects of differing vocabulary knowledge or knowledge of subject matter, and different ways in which language knowledge affects test performance, including culturally specific interpretations of instructions and items.

COMMUNICATION IN THE CLASSROOM

Understanding the social context of cognition—the communication systems of classrooms and other learning environments—is necessary if new advances in cognitive science and technology are to be applied to the classroom. In the past 10 years, a new field of study has developed, drawing on the perspectives of linguists, anthropologists, and sociologists. This area, variously called sociolinguistics or the ethnography of communication, has greatly increased our knowledge about the regularities of human communication in both verbal and nonverbal modes. Technological advances in audio and video tape equipment have made possible the development of systematic procedures for identifying recurrent patterns of communication in particular settings. In their application to educational research, ethnographic methods broaden the parameters used to assess the competence of teachers and students by examining them in a range of naturalistic activities and cultural contexts. One of the major recommendations of an NIE Conference on Studies in Teaching (1975) was for research on teaching as a linguistic process, which would help children and teachers understand each other more fully:

> The study of linguistic phenomena in school settings should seek to answer educational questions. We are interested in linguistic forms only insofar as through them we can gain insight into the social events of the classroom and thereby into the understandings which students achieve. Our interest is in the social contexts of cognition: speech unites the cognitive and the social. The actual (as opposed to the intended) curriculum consists in the meanings enacted or realized by a particular teacher and class. In order to learn, students must use what they already know so as to give meaning to what the teacher presents to them. Speech makes available to reflection the processes by which they related new knowledge to old. But this possibility depends on the social relationships, the communication system, which the teacher sets up.

Responding to this recommendation, the NIE Program in Teaching and Learning solicited research on three questions: What is the nature of communication in the classroom? How do students acquire the rules of classroom discourse? What are the effects of inadequate learning of classroom discourse rules? Awards were made for a number of investigations into communicative patterns among teachers and children in culturally diverse elementary school classrooms, bilingual programs, and homes. While these research projects do not deal with implementation of research findings, they have important implications for the training of teachers, improvement of instruction, assessment of linguistic competence in students who speak little English, and evaluation of new programs in bilingual and bicultural education.

The interest in studying teaching and learning in a more holistic manner and in more naturalistic contexts is also leading to educational research in out-of-school as well as classroom settings, in adults as well as in children.

INFORMATION-HANDLING TECHNOLOGY

THE TECHNOLOGY

We have entered a revolutionary age in the technology of information-handling. The microelectronic revolution, extensively documented in the September 1977 issue of *Scientific American*, is making possible a roughly tenfold decrease each five years in the cost of the integrated circuits that are at the heart of contemporary digital computers. We have available for individual use, at relatively low prices, sophisticated calculators and computers and means of generating images on video tubes. As supplements to the now universally available TV receiver, we have video tape recorders and players and will soon have the even cheaper and, in educationally important ways, more flexible videodisc player. Following are some examples of currently available devices useful in education.

Computers and Calculators

Handheld calculators providing the four arithmetic functions plus square root now sell for about $10, and the cost of calculators providing trigonometic and logarithmic functions has declined to the same level. Calculators able to handle 400-step programs are now at the $100 level. Handheld instructional devices providing drill and games in arithmetic sell for $15-25 and in spelling, with a simulated human voice, for $65. Personal computers, which include a TV tube display, keyboard, processor, a sizable memory, and weigh less than 50 pounds, are now available for $600. These computers are comparable in capacity to computers costing hundreds of thousands of dollars a decade ago, and they are only the first of many that will become available at decreasing prices or with much greater capability at the same price. Incorporation of microcomputers into TV sets could make substantial computer power available in the home to each child.

Video Storage, Retrieval, and Display Devices

A videodisc system coming onto the market for $700 will provide, in conjunction with a regular TV set, 54,000 separate frames of full-color picture image, or alphanumeric or computer information, on one side of a disc similar in size and cost to an LP record. The 54,000 frames correspond to 30 minutes of video at 30 frames per second. The system also provides rapid random access (dialed in more expensive models) to each of the frames, which are individually numbered and can be viewed individually for any length of time desired and may contain roughly one-quarter of a page of easily legible text (or a full page with special high-resolution video tubes), and easy replay, at regular or slower speed, of any portion of a motion picture. Combinations of videodisc players and microcomputers, referred to as intelligent videodisc systems, allow controlled sequencing of video frames (which may be based on student responses) as well as the possibility of transferring computer programs from the disc to the computer. While no such systems are available commercially at this time, NSF has let two contracts for development of prototypes designed to be useful in education. The disc can store 10 billion bits of information (*Encyclopaedia Britannica* contains 2 billion bits; a human chromosome has a capacity of about 20 billion bits; and the human brain perhaps 10,000 billion). This immense storage capacity could probably include on a single disc all the computer courseware ever published. There are now available, for less than $1,000, home color TV tape recorder-players. Tape systems have the advantage of allowing recording and playback of video program materials that have not been prerecorded, but tape is currently more expensive per minute of playing time than a disc, and single-frame display is more difficult.

Other Technology

Technology in several other new forms can be applicable to education. For instance, TV games that sell for as low as $40 to $50 are a familiar form of video graphics. Computer-controlled graphics with 15 colors are available for $300 to $400. A speech generator that strings together common phonemes in response to computer commands is available for under $400, and a device that can be taught to recognize a number of spoken words starts at $300. Music generators now range from ones that string together single notes into a melody to those that permit control of attack, decay, and timbre. Word-processing devices that allow preparation of hard copy of written text after editing and margin justification are available in various forms; among them is an accessory to home computers. Existing and other foreseeable and barely foreseeable devices will make possible great improvement and cost reduction in existing educational applications of technology, as well as major qualitative changes in the use of technology in instruction.

APPLICATIONS

New technologies substantially improve our ability to keep students actively engaged in learning for sustained periods, a prime requisite for effective education; allow teaching to be adjusted to the rate and style appropriate to the individual student and to the particular difficulties the student encounters; reach teachers with high-quality inservice training coupled to approaches they are using; and test students in ways that are fairer and provide more precise information on the sources and cures of difficulties the student is experiencing. In addition, their cost is reasonable.

In 1978, the subcommittee of the House Committee on Science and Technology issued a report, *Computers and the Learning Society*, based on extensive hearings. The report found wide agreement for the ideas that properly implemented uses of computer technology in schools can make significant improvements in teacher effectiveness and student learning and that the microcomputer revolution very likely will make personal computers commonplace in the home in the near future and dramatically reduce the costs of general purpose computer systems, thereby introducing unique educational opportunities. The subcommittee made several recommendations for Federal action responding to these opportunities. They recommended that the Federal Government be the primary source of funding in the computer-based education field, taking leadership initiative in development and implementation of a research strategy that will satisfy the requirements for compatibility between broad national goals and those of the country's highly autonomous school systems. They recommended that federally sponsored research be directed toward increasing basic understanding of appropriate hardware, software, courseware, and instructional theories, and toward identifying and testing organizational and managerial techniques needed for effective use of technologies in education.

The National Institute of Education and the National Science Foundation are carrying out a number of activities consistent with these recommendations. NIE is supporting a study of the effects over time on elementary school students of varying amounts of computer supplements to regular instruction in arithmetic, language arts, and reading. The NIE-sponsored conference on research on testing, referred to earlier in connection with cognitive science, laid great emphasis on the potential of new low-cost information-handling technology for creation of better educational tests, and NIE's research grants program in testing can provide support for some such work. The technology makes feasible, for example, tests that simulate real situations in which students would be using what they learned, tests that follow students through more complicated tasks, and tests that diagnose difficulties students are having and suggest remedies. As noted, NSF is supporting development of prototype intelli[illegible] videodisc systems for educational use, and it is supporting other projects developing applications of technology to science education. NSF and NIE are also cooperating in a new joint program of research and development in the use of technology in the teaching of mathematics.

The activities described here are not as yet at a level sufficient to realize the great potential of technology for education. Private enterprise is already producing a variety of technological products related to education, but, as in many other areas, individual firms are not generally likely to be able to justify support of the costly research and development required for full exploitation of the possibilities. As pointed out by the congressional report, *Computers and the Learning Society*, and an earlier Commission on Instructional Technology (1970), the Federal Government seems the only feasible source of the necessary funds.

The following are examples of the application of new technologies to basic skills in reading and computation.

Drill and Practice

Currently available full computer-based drill and practice programs in elementary reading, language usage, and computation were designed to be economically feasible with the costly technology of the 1960s. At current or projected prices of computer capability, considerably more sophisticated programs could be produced for delivery at the same cost. Alternatively, the same program should be deliverable more conveniently at considerably lower cost (e.g., it seems feasible to build an entire computational drill and practice program into a handheld calculator costing perhaps $50 or less).

Testing

Much standard testing is now carried out in the multiple-choice format to facilitate either hand scoring or machine scoring based on half-century-old technology. Current and prospective technology makes possible computer-aided diagnostic and performance testing of far greater sophistication at costs comparable to tests based on the old technology.

Instructional Programs

The combination of microcomputer and videodisc (with video and audio) makes possible sophisticated instructional programs at low cost, including instruction of students who cannot yet read instructions. These programs are generally designed to be useful to the teacher as a supplement to regular classroom education. They can be advanced examples of existing types of computer-assisted instruction (CAI) programs, but new low-cost computing power and new knowledge of human learning processes make feasible qualitatively different sorts of interactive instruction in which the computer is "knowledgeable" about the subject matter being taught, and about tutorial strategies suitable for dealing with the learning behavior of individual students.

It is easy to imagine computer-videodisc programs teaching reading comprehension strategies in ways that have not been possible in the past. Early work is underway on programs that allow children to learn mathematics (or reading and writing) in the same way they learn to understand and speak their native tongue—by "immersing" them in a world where mathematics is "spoken."

ORGANIZATIONAL SCIENCE

Modern organization science has shifted attention to what actually happens in organizations away from almost exclusive concern with what ought to happen, thereby adding science to what was previously often more managerial engineering than managerial science. The importance of this line of thinking was emphasized recently in the award of the Nobel Prize in Economics to Herbert A. Simon whose classic *Administrative Behavior* laid the foundations for this new approach. The following discussion, from the standpoint of organizational science, of organizational problems in education and bureaucratic pathologies in educational organizations is taken from an unpublished summary by James G. March of a series of conferences which he arranged with the support of NIE.

ORGANIZATIONAL PROBLEMS IN EDUCATION

Not all education occurs in organizations, and the educational institutions of the future may be different from the schools with which we are familiar; however, much of education is conducted in relatively large bureaucratic organizations. Those organizations are the focus for considerable contemporary complaint.

Administration

Educational administration has been variously characterized as too weak, too strong, too professional, and too incompetent. Unfavorable comparisons have been made between the 20th century gains in productivity in most industries and the rising costs without apparent gains in education. Educational organizations have been described as excessively resistant to change and easily corrupted by change. School systems have been enlarged in the name of efficiency and equity; they have been reduced in size in the name of community control and efficiency. A system of professional certification and organizational careers has been constructed in pursuit of quality and is now challenged on the same basis. Parents, students, ethnic groups, unions, legislatures, and state and Federal agencies are challenging various aspects of school organization, though rarely in a coordinated or consistent way.

Contemporary distress over the schools is not directed exclusively, or even primarily, at educational organization. We have become catholic in our distastes. Criticisms have also been directed at the curriculum, the teaching, the technology, the educational materials, the physical conditions, the composition of the student body, and the school lunches. Nevertheless, all of these are, in some degree, a product of the organization and administration of education.

Changes in Demography

Concerns about educational organization have been intensified by demography. Changes in the birth rate have produced changes in the demand for education because of both the number of children attending schools and the number of taxpaying parents willing to support schools. Changes in the ethnic and socioeconomic mix of students in public schools have produced changes in the character of the demand for education: different things are wanted, and their achievement apparently requires different techniques, if indeed they can be achieved. Changes in the professional training and career aspirations of teachers, other professional staff, and school administrators seem to have reduced turnover in educational occupations, restricted entry into them, and increased the costs of staffing. These changes seem systematically to be raising the cost of education while reducing the price society is willing, or able, to pay.

Efficiency and the Role of Operations Research

One implication is that education may be inefficiently organized or managed. As a result, significant effort has been directed toward bringing into educational management the modern techniques of management science, operations research, and information-handling. The effort has made notable contributions to the way educational institutions manage records, account for expenditures, coordinate utilization of space and equipment, and forecast flows of students and student activities.

Despite such successes, operations analysis in education has been subject to criticism among both management scientists and educational administrators. Some of the criticisms are technical; some are visceral. But there seems to be a moderate consensus on two important impediments to making gains in efficiency through applications of standard management science techniques to education. First, a very large fraction of total educational expenditures is for teaching personnel and related professionals organized around classrooms. Using most management science techniques requires considerably more precise information than we now have about the details of teaching activities and their susceptibility to change. Second, educational organizations appear to have less well-defined goals and less well-understood technologies than most organizations that have been sites for major operations analysis successes.

In addition, many educational practitioners and students or organizations caution against a too narrow definition of organizational "efficiency" in education. There is some risk that we will attend to those things we understand and can measure, and ignore those things, however important, that we have greater difficulty translating into an explicit criterion function or an explicit production function. Without denying the numerous good opportunities in education for improving organizational practice through existing tools of management science, we need other kinds of administrative and organizational creativity that attend more closely to some fundamental properties of educational organizations and to our understanding of them.

Similar observations can be made about other attempts to improve the effectiveness of educational organizations. In recent years we have attempted to improve education through various forms of organizational change. For example, community participation, vouchers, collective bargaining, and consolidation. Advocates of these and other changes have been able to cite observations suggesting that the changes have been useful, but there have been disappointments in every case also. Our experience with organizational reform commends a certain caution in our confidence that we know how to predict the consequences of changes in organizational structure, practice, or personnel.

There are many things that can be done. Educational organizations are not immune from simple waste, ordinary mismanagement, and bureaucratic foolishness. We do not require research to know that something should be done in such cases; nor do we require research to know what to do. But our history of modifying organizations in education (as well as other public bureaucracies) suggests that some of the problems are resilient to standard analysis even when changes are fervently believed in and vigorously pursued. This suggests the possibility that we might want to invest a modest fraction of our educational investment to developing a better understanding of the organizations that spend it.

BUREAUCRATIC PATHOLOGIES AND EDUCATIONAL ORGANIZATIONS

If bureaucracy were not an efficient instrument for organizing and implementing action, we could easily solve bureaucratic problems by eliminating bureaucracy. Our difficulties arise not from the unrelieved inefficiency of large-scale organizations but from the fact that those features of bureaucracy that seem to be most central to efficiency also seem, under certain circumstances, to produce large, perverse effects.

We call these difficulties pathologies of bureaucracy to suggest the view that health and disease may be manifestations of broader processes that produce both. In public bureaucracies we appear to face two major kinds of pathologies. In some cases, a feature of organization that is generally functional under some circumstances seems to expand uncontrollably, feeding upon itself until it becomes an essentially grotesque caricature of efficiency. The classic example is the proliferation of formal rules. In other cases, a feature of organization that is generally functional is applied to a situation that differs in some unrecognized vital way from the standard situation, so that essentially sensible activities lead to disaster. A perfect example is the application of classic notions of rational planning to domains in which objectives are unknown or in conflict.

Five Types of Pathologies

Bureaucracies in the public sector appear to be subject to a long list of such pathologies. Many of them can be summarized under five classifications:

(1) The pathologies of procedures—efforts of organizations to administer themselves through regulations and rules produce not only reliability but also rigidity, arbitrariness, and evasiveness.

(2) The pathologies of hierarchy—the use of hierarchical organization, administrative careers, and managerial authority as cornerstones of organizational action has natural complications.

(3) The pathologies of specialization—the division of labor and the development of competence can help create problems of communication, coordination, and control.

(4) The pathologies of rationality—our attempts to make decisions intelligently, to plan, and to coordinate actions with objectives may cause complications when goals are ambiguous or in conflict, futures uncertain, and alternatives ill-defined.

(5) The pathologies of renewal—the ways in which organizations renew themselves through problem-solving, change in personnel or policy, evaluation and accountability, and learning can present unanticipated difficulties.

Procedures, hierarchy, specialization, rationality, and renewal are fundamental to modern bureaucratic organization. Although there are periodic efforts to design organizations without those properties, they generally have appeared to be so basic to the efficient operation of organized action that reform more commonly seeks to introduce more, rather than less, of them. By labeling certain consequences of such standard bureaucratic forms as pathological, we mean to suggest that they are potentially present whenever we organize a bureaucracy, that the processes changing effective bureaucratic actions into pathologies are powerful, and that they can be understood and ameliorated. Although they are inherent in the forms of organization we use, they are not simply costs that must be borne.

Educational Bureaucracy's Special Vulnerabilities

These pathologies of bureaucracy are familiar to anyone who has studied or lived in a large-scale organization. They can be seen in business firms and armies. But educational institutions have special properties as organizations that make them particularly vulnerable to some manifestations of the phenomena. None of the properties are unique to education, but they are conspicuous there. Consider the following four characteristics:

First, educational organizations are social institutions. They exist in, are affected by, and affect an environment of parents, taxpayers, professional groups, politicians, and governmental agencies. Teachers and administrators perform a variety of social roles simultaneously. Education is a social value and an ideological battleground. Schools change (or fail to change) in response to various signals, pulls, and pushes from the environment.

Second, educational organizations are hierarchies. They use a conventional bureaucratic form of organization with relatively standard ideas of authority, administration, and control. Personal prestige and perquisites are associated with position in the hierarchy; there is a well-defined managerial group; meetings are held, decisions made, and programs implemented. Individuals are presumed to rise in the administrative ranks on the basis of relatively impersonal criteria of merit. Leaders are fired if the organization does poorly.

Third, educational organizations are professional. The major work force (teachers) is professionally trained and identifies with a profession. They (and others) expect their behavior to be governed by professional rules of conduct; they (and others) expect that teachers will have considerable control over their own activities as long as they are related to professional work; and they (and others) expect to resolve questions of proper practice by reference to outside professional authority rather than internal bureaucratic authority.

Fourth, educational organizations make decisions under conditions of ambiguity. Educational technology is poorly understood; asserted educational objectives tend to be either vague, contradictory, or not widely shared; participants in educational organizations include individuals and groups who move in and out of activity in the organization sporadically.

These four features of educational organization are fundamental and pervasive. The same features can, of course, be found in other organizations, and research directed at exploring their consequences will be relevant to other hierarchical, professional organizations operating under ambiguity in a social context. In particular, many of the same things could be said about other public bureaucracies. But basic research bearing on such factors is particularly necessary for understanding and improving educational practice.*

THE NIE PROGRAM

NIE has instituted a continuing research grant program in organizational processes in education designed to support research along lines such as those discussed above. Results of the research will be made directly available to educators and others in easily usable form and will also be taken into account in the design of other research and development programs, dissemination and improvement of practice programs, and advice given by NIE on design of other educational support programs. A recent example is the following amendment to ESEA Title I legislation (Education Amendments of 1978—Public Law 95-561, Sec. 124):

> (i) TEACHER AND SCHOOL BOARD PARTICIPATION
>
> A local educational agency may receive funds under this title only if teachers in schools participating in programs assisted under this title, and school boards or comparable authority responsible to the public with jurisdiction over the schools, have been involved in planning for these programs and will be involved in the evaluation thereof.

This provision was based on research which showed that innovations imposed from above in educational organizations have much smaller likelihood of success than those developed in the way called for in this statute.

* End of summary by James G. March.

REFERENCES

Bruner, J.S.; Goodnow, J.J.; and Austin, G.A. *A Study of Thinking*. New York: Wiley, 1956.

Chomsky, A.N. "Three models for the description of language." *IRE Transactions on Information Theory*, IT-2(3) (1956) pp. 113-124.

Commission on Instructional Technology. *To Improve Learning*. New York: R.R. Bowker Company, 1970.

Glaser, R. "The contributions of B.F. Skinner to education." In Suppes, P., ed. *Impact of Research in Education: Some Case Studies*. Washington, D.C.: National Academy of Education, 1978.

Miller, George A. "The magical number seven, plus or minus two." *Psychological Review* 63:81-97, 1956.

Miller, George A., ed. *Linguistic Communication: Perspectives for Research*. Report of the Study Group on Linguistic Communication to the National Institute of Education. International Reading Association, 1973.

National Institute of Education. *Teaching as a Linguistic Process in a Cultural Setting*. Report of the National Conference on Studies in Teaching, Vol. 5. Washington, D.C.: National Institute of Education, 1975.

Newell, A. and Simon, H.A. "The Logic Theory Machine: a complex information processing system." *IRE Transactions on Information Theory* IT-2 (3) (1956) pp. 61–79.

Newell, A., and Simon, H.A. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1972.

Simon, H.A. *Administrative Behavior*. 3rd Edition. Riverside, N.J.: Free Press, 1976.

U.S. House of Representatives. *Computers and the Learning Society*. Report prepared by the Subcommittee on Domestic and International Scientific Planning, Analysis and Cooperation of the Committee on Science and Technology. Washington, D.C.: U.S. Government Printing Office, 1978.

Department of Health, Education, and Welfare National Institutes of Health

The years ahead present unusual opportunities for progress in biomedical science. Investigations in the basic areas of genetics, immunology, virology, molecular and cell biology, and neurobiology are moving forward rapidly and promise to open penetrating new insights into the nature of life, the functioning of biological systems, the character of disease, and the conditions necessary for health. The new recombinant DNA technology, for example, is expected to provide a powerful experimental approach to a fuller understanding of diseases that are recognized to have genetic components, including diabetes, cancer, and cystic fibrosis; this technology may yield new therapies for previously untreated diseases. We are also moving toward a better understanding of the interaction of the chemical, physical, and biological phenomena that form the basis of nutrition, therapeutics, toxicology, and environmental health. The prospect for further research advances in addressing the major chronic and metabolic diseases seems promising.

However, an increasing divergence between opportunities and resource capabilities seems probable for the foreseeable future. While the vitality of the research enterprise remains high, certain disquieting trends have become perceptible: over the past several years, national support for health research has leveled off; the quality of scientific plant and equipment has deteriorated; research opportunities for young university scientists have been diminishing; and the proportion of M.D.'s electing to train for careers in biomedical research has fallen sharply.

To help cope with these and other critical problems, the Secretary of Health, Education, and Welfare (HEW) has proposed the development of a five-year strategy to guide the allocation of the Federal health research dollar. Such a plan may serve to level out the annual appropriations fluctuations and thereby facilitate more rational health research planning. As a first step in this process, a National Conference on Health Research Principles was convened in Washington, D.C., on October 3-4, 1978, to receive public testimony from witnesses suggesting health research principles and to develop draft panel reports organized around several substantive foci. The planning principles adopted at that conference* are intended to serve as a structure for the development of a strategy that could contribute to the resolution of pressing national problems in the health sciences.

A number of critical issues of national significance have emerged and will require the special attention of the National Institutes of Health (NIH) in the next five years. These concern the need to sustain the vigor of basic research; address concerns associated with the development of the new recombinant DNA technology;

* See Conference Report, *National Conference on Health Research Principles*, U.S. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, January 1979.

press for further advances against the nation's major health problems; develop appropriate measures to confront critical environmental factors affecting human health, particularly in the areas of environmental toxicology and nutrition; strengthen the capacity for technology and information transfer in the health field; and assure essential supporting resources—funds, facilities, and manpower—for the nation's biomedical research effort. Each of these concerns is discussed below.

SUSTAINING THE VIGOR OF BASIC RESEARCH

Increasing fiscal constraints and pressures for a stronger orientation of health research toward immediate applications will undoubtedly complicate support for basic research and raise persistent problems of balance and priority. Significantly, the most frequent assertion of commentators at the National Conference on Health Research Principles was that support of the science base (basic research and supporting resources)* should be stabilized to the extent possible; indeed, conference participants agreed as a governing principle that a national commitment to fundamental research is essential in meeting the full range of public health expectations.

Also stressed by the conferees were the following: (1) that the decision to fund fundamental research in a given area should be based on an assessment of related ongoing research and new scientific opportunity (here, the peer review system must be regarded as an efficient and essential instrument in the conduct of research grant programs); (2) that investigator-initiated research proposals must continue to be emphasized in the conduct of fundamental research and must be restored to their previous prominence as a mechanism for allocation of funds; and (3) that recognition must be given to a broadened concept of health factors and a research base developed to investigate them, with particular emphasis on studies in the behavioral sciences and population sciences, including epidemiology and biostatistics.

In the debate on the allocation of resources, it has been necessary to emphasize the importance of basic research in contributing to human health and well-being. Indeed, that contribution has been a gratifying one. The elucidation in recent years of immune processes at cellular and molecular levels, for example, has led to improved treatment of diseases of allergic origin, progress in preventing graft rejection in patients, and greater understanding of genetic susceptibility to disease. The recognition of the role played by "slow" or "latent" viruses in disorders of the central nervous system is another important achievement. For his work in this area, Dr. Carleton Gajdusek of the National Institute of Neurological and Communicative Disorders and Stroke was awarded the 1976 Nobel Prize in medicine. And the striking advances made in the past 10 years in understanding the structure and mechanisms of enzyme action have offered promise that enzyme replacement therapy may lead to effective treatment for certain genetic disorders. These and many other illustrations of the contributions of basic research to human health and well-being have been described in detail in the National Science Board's Tenth Annual Report (1978), *Basic Research in the Mission Agencies.**

It is encouraging that we have reached the point in biomedical science when investments are likely to be very productive. Investigations in several fields—genetics, immunology, virology,** and cell biology—are moving forward rapidly; these areas are basic to our understanding of virtually all disease processes. The neurosciences, with their enormous potential for contributing to physical and mental well-being, are also considered very promising. (Significant progress has been made, for example, in elucidating some of the mechanisms involved in such diseases as multiple sclerosis, myasthenia gravis, and parkinsonism; the development of new chemical compounds may well transform the pharmacological treatment of neurological disorders.)

The advances of the past and the promises of the future do not, however, signify that the central tasks of basic biomedical research are nearing completion; the fundamental science bases needed to solve the problems of cancer, heart disease, diabetes, arthritis, hereditary diseases, and mental disorders are exceedingly complex and will require more time. Investments, however, can be made in ways that take advantage of available opportunities and avoid premature emphasis in areas where the necessary knowledge infrastructure has yet to be adequately developed. Inevitably, difficult choices must be made between costly development of half-way technologies, on one hand, and more basic research—with a longer latent period before payoff and higher risk—on the other.

* The "Science Base" concept is now being tested by the National Institutes of Health as part of a broader system of allocation of monetary resources designated "*SATT*" (Science Base, Applications Research, Technology Transfer, and Training.)

* See, also, *NIH Research Advances* (published annually by the National Institutes of Health) and Appendix A of the *Report of the President's Biomedical Research Panel* for additional examples of recent research achievement.

** Some concern has been expressed that the emphasis, interest, and opportunities in cancer virology may be blunted by the growing demands for funding of other promising areas, such as environmental toxicology and nutrition; the net effect may be a decrease in support for research in virology and virus diseases.

RECOMBINANT DNA RESEARCH: ISSUES GENERATED BY A NEW TECHNOLOGY

The techniques used to produce recombined molecules of DNA, the complex chemical that codes genetic information for all living cells, hold great promise for significantly advancing our understanding of fundamental biological processes. These techniques also offer hope of important progress in cloning modified cells for the large-scale production of biological compounds for the treatment and control of disease. For example, the possibility of producing insulin through recombinant DNA techniques has now been demonstrated, and commercial production of this hormone appears probable within the next five years. Applications of recombinant DNA techniques in agriculture are equally exciting, and studies are now underway to examine the possibility of increasing food production through the development of new disease-resistant plant strains, and a decrease in the fertilizer requirements of crops by directly transferring to plants the microbial enzyme systems that perform nitrogen fixation.

From the pioneering days of this research, however, many of the nation's leading scientists expressed concern that the insertion of foreign genes into microorganisms could carry the potential for harm by yielding new disease-producing organisms. Although no harm has resulted from recombinant DNA research to date, there has been widespread uncertainty about the degree of risk involved. On January 2, 1979, revised NIH Guidelines governing recombinant DNA research in this country went into effect. These guidelines should permit the promise of DNA research to be realized without presenting any significant risk to public health or the environment. They provide a flexible, open system that can accommodate new scientific information that may warrant change, either to relax or to increase safety requirements. Concurrent with their release, steps have been taken to require compliance by private companies engaged in recombinant DNA research, primarily through use of the regulatory authority of the Food and Drug Administration, and through appropriate action by the Environmental Protection Agency. Additionally, public participation in the process of administering the guidelines is assured through representation on the Recombinant DNA Advisory Committee and Institutional Biosafety Committees.

It is essential that the mechanisms established for regulating this research function effectively; consequently, over the next few years, NIH will carefully review their operation. As directed by the Secretary of HEW, NIH will also formulate a plan for carrying out a balanced program of risk-assessment experiments to better judge whether the guidelines—and the actions taken under them—afford appropriate protection for health and the environment.

ADDRESSING THE NATION'S MAJOR HEALTH PROBLEMS

PRINCIPAL CAUSES OF MORBIDITY IN THE UNITED STATES

Cancer, diseases of the heart and lung, and certain chronic and disabling diseases such as arthritis, diabetes, and digestive diseases constitute the principal causes of morbidity in the United States and result in enormous treatment costs to those afflicted and to the taxpayer. Infections remain the single most common cause of morbidity and cause significant premature mortality.

The pervasiveness and impact of these problems may be briefly noted. More than 100 clinically different cancers are recognized, each having a distinct pathologic appearance and requiring different kinds of treatment. Cardiovascular diseases cause over half of all deaths in this country and account for more than one-third of all potential years-of-life lost as a result of illness. The social and economic burden of diseases of the lungs and respiratory system is equally severe; while these cause fewer deaths than cardiovascular diseases, many kill earlier in life and result in exceptionally high treatment costs. Arthritis, diabetes, and digestive and kidney diseases are the first cause of all physician visits and of all short-term hospital stays.

To confront these critical problems, health research and programmatic efforts will focus on the areas briefly highlighted below.

Cancer

Promising research has generated rising expectations, particularly for cancer prevention. The promise and the expectation stem from recognition of the causative role of environmental factors. In recent years, new knowledge of the cancer process, new experimental models, and new research methods have made it possible to explore the actions of multicausal factors, the relation between mutagenicity and carcinogenicity, the potential reversibility of various cellular abnormalities, and the mechanisms of action of agents influencing successive steps in the initiation and promotion of neoplastic changes. New approaches to epidemiologic research and improved carcinogenicity test bioassay methods also promise added insight into the questions of cause, risk, and prevention.

In the area of cancer control, a program to study, evaluate, and demonstrate new methods of notifying high-risk populations with a view toward influencing health-related behavior will be implemented in fiscal year 1980. Elements of this program include activities on smoking, nutrition, alcohol, radiation, and work place exposures. A new effort will also begin in fiscal year 1980 to establish and operate clinical facilities for the evaluation of high-energy neutron therapy.

Early detection still holds promise for improved survival, but the issue has proven to be more complicated than suspected. More research in validation of the effectiveness of early detection techniques is needed, including the use of clinical trials, before early detection methods are put into wide practices.

Gains in the treatment of leukemias, lymphomas, and certain childhood cancers have been dramatic. To extend these advances, clinical trials of adjuvant chemotherapy applied to some of the cancers more common in adults are currently in progress. Preliminary results in the treatment of breast cancer are considered promising.

Diseases of the Heart and Lung

There is a challenge to reduce not only the morbidity and mortality of cardiovascular and pulmonary diseases but also the staggering economic costs. To sustain the momentum of the dramatic decline in deaths of the 1970s, the National Heart, Lung, and Blood Institute (NHLBI) will continue in the 1980s to pursue a national program that encompasses all activities in the biomedical continuum from the acquisition of new knowledge and the validation of existing information, to the translation of proven findings into health practice.

Prevention and effective treatment of overt disease depend to a large extent on being able to detect early, often asymptomatic, stages. Recent advances in fundamental research have produced a number of non-invasive methods for early diagnosis. The Institute is developing and evaluating instrumentation to image atherosclerotic lesions using ultrasonic methods. The prototype instruments are able to image cross-sections of arteries to show the presence and structure of small early lesions. An acoustic pulse response technique, similar to sonar, is also being tested as a way of detecting and characterizing small airway obstructions in lung diseases such as asthma, bronchitis, and emphysema. These and other new methods not only show promise as safer, easier, and less costly methods of diagnosis, but they can be repeated without hazard over time to evaluate a patient's therapeutic progress.

Clinical trials test the efficacy and safety of preventive and therapeutic regimens with the potential to save hundreds of thousands of lives and billions of dollars each year. The Aspirin-Myocardial Infarction Study (AMIS) will document the role and efficacy of the use of aspirin in the treatment of coronary artery disease, and the Hypertension Detection and Follow-up Program (HDFP) will provide information on the value of treating mild hypertension. Two new NHLBI trials began their patient recruitment phase this year. The Beta-Blocker Heart Attack Trial (BHAT) will test the efficacy of propranolol, a drug with anti-arrhythmic properties, in the prevention of recurrent heart attacks. The Multicenter Investigation of Limitation of Infarct Size (MILIS) Trial is designed to assess the ability of therapeutic interventions to reduce permanent heart muscle damage at the time of a heart attack.

The results of all NHLBI trials are to be disseminated to the health care practitioner as soon as the data are analyzed so that translation to regular health care practice can take place expeditiously. Such translation has been successful in the National High Blood Pressure Education Program which has contributed to an increase in national awareness, an increase in patient visits for high blood pressure, and a decrease in stroke deaths. From experience with this program, the Institute is moving more visibly into nutrition education. Much of the new program will focus on the potential value of nutrition education at the point of food purchase. Successful pilot projects have already been completed in cafeterias and with vending machines. The "Food for Health" project is the forerunner of programs designed to help consumers make informed choices in other food selection locales.

Arthritis, Diabetes, Digestive Diseases, and Other Chronic and Disabling Diseases

It is anticipated that the special centers program of the National Institute of Arthritis, Metabolism, and Digestive Diseases (NIAMDD)—the Multipurpose Arthritis Centers (MAC) and the Diabetes Research and Training Centers (DRTC)—will continue to expand at a modest rate over the next five years from the current levels of 24 MACs and 8 DRTCs, creating the potential for significant impact on these extremely important public health problems.

In the area of digestive diseases, cognizance is taken of the recent report of the National Commission on Digestive Diseases which shows that these disorders account for 10 percent of the total cost of illness and 15 percent of general hospital admissions, for an economic loss of $17 billion yearly in medical care and another $35 billion in lost work and wages. Clearly, a carefully formulated expansion into this area to take advantage of available research opportunities appears appropriate.

Infectious Diseases

Paths of future research have been charted in the six-volume report of the Task Force in Virology created by the National Institute of Allergy and Infectious Diseases (NIAID). Vaccines against epidemic respiratory and bacterial infections are needed and are being sought, a hepatitis B vaccine is nearing clinical trials, and the way seems open to development of additional safe and effective antivirals. Equally important, there is growing evidence of the role of infection and immune reactions as the underlying cause of many chronic and degenerative diseases such as some forms of arthritis, kidney disease, and diabetes.

OTHER PRIORITY AREAS

The health problems discussed above have been selected largely because of the magnitude of their impact in human and socioeconomic costs. There are, of course, other important problem areas, such as aging, developmental medicine, disorders of the visual systems, oral and dental diseases, neurological and communicative disorders, and prevention. Two of these areas—aging and developmental medicine—are briefly reviewed below.

Aging

Unprecedented growth in the older population has clearly resulted in major economic, social, and health care problems. It is anticipated that epidemiological and demographic studies will further our understanding of the health status of the elderly (nutritional status, characteristics of the institutionalized aged, incidence of senile dementia, circumstances and events of the last days of life). Special research efforts—such as the program on organic brain diseases at the newly established Laboratory of Neurosciences at the National Institute on Aging—will address promising scientific opportunities. An immediate pressing need is to strengthen academic capabilities in geriatric medicine.

Developmental Medicine

Research in reproductive and developmental biology offers a unique opportunity to approach the health problems that interfere with the ability to plan fertility and with processes that affect normal fetal and infant development. Of importance are studies on the impact of different life styles on health, particularly in the causation of chronic diseases, and investigations into environmental influences that deter normal growth and development.

SPECIAL INITIATIVES IN ENVIRONMENTAL TOXICOLOGY AND NUTRITION RESEARCH

Greater cognizance of environmental and nutritional factors as causes of disease has inspired initiatives whose impact should be felt increasingly in the next five years. These are the HEW-wide National Toxicology Program and the special NIH initiative in human nutrition, both briefly considered below.

National Toxicology Program

Few problems in public health are more important than the prevention and mitigation of our population's exposure to toxic substances. These are known to contribute to cancer, chronic degenerative diseases, reproductive disorders and birth defects, immunologic diseases, and other health problems. In recognition of the serious nature of this issue, the Secretary of HEW, on November 15, 1978, established an experimental National Toxicology Program to strengthen the Federal Government's activities in the testing of chemicals of public health concern, as well as in the development and validation of new and better integrated test methods. At present, the program comprises the relevant activities of the Food and Drug Administration, National Cancer Institute, Center for Disease Control/National Institute for Occupational Safety and Health, and National Institute of Environmental Health Sciences. It will be planned and carried out as a coordinated whole under the direction of the Director, National Institute of Environmental Health Sciences.

Initial efforts will address current toxicology testing capacity and how that capacity is being utilized, the amount of test capacity that will be available in the coming year, plans for test development, and validation of test systems which take into account research opportunities and needs of the field. It is intended that the program will broaden toxicological characterization of those chemicals being tested, increase the rate of chemical testing within the limits of available resources, and develop and validate a series of protocols more appropriate for regulatory needs.

Serious problems with the information transfer process in toxicology and related fields were recognized more than 10 years ago. The National Library of Medicine's Toxicology Information Program has tried to address this problem by developing modern computerized information services such as TOXLINE and CHEMLINE. Through this program, the library will provide information support to the new National Toxicology Program.

Human Nutrition

Nutrition policy is now at an important juncture due largely to such factors as the continued development of new (and often conflicting) knowledge about the relationship between nutrition and health; sustained political pressures for a comprehensive national nutrition policy embracing social, economic, agricultural, energy, and health considerations; and mounting evidence that governmental policies and regulations and federally funded programs concerned with nutrition lack essential coordination. Public interest in nutrition and the health consequences of the American diet has also increased in recent years, as have the demands for more visible and deliberate governmental action to counterbalance, if not eliminate, potentially negative influences (e.g., television advertising) on diet and health.

At NIH, the biomedical and behavioral nutrition research and training program is coordinated through the Nutrition Coordinating Committee (NCC). NCC consists of representatives from the 11 institutes and 2 divisions that support research in nutrition, plus liaison members from other agencies of the U.S. Public Health Service, HEW, and the Office of Science and Technology Policy.

Nutrition research has passed through two stages and is now entering a third stage. The first stage saw the discovery of vitamins and the development of many of the basic nutritional requirements. The second stage reduced nutrition into subcellular and molecular terms within the areas of biochemistry and physiology. The third stage, dealing with the effects of diet on a wide variety of chronic diseases, will require further research and large-scale clinical trials.

Nutrition research at NIH will concentrate on several critical areas: clinical nutrition throughout the life cycle, including effects of infant nutrition on subsequent physiological, immunological, and mental development; the role of nutrition in disease development, particularly the elucidation of mechanisms by which dietary deficiencies, imbalances, and excess lead to the development of physical and mental disease; the role of nutrition in the prevention of disease, with emphasis on all forms of malnutrition; and the role of nutrition in the treatment of disease and development of nutritional therapies for specific diseases, such as cancer, gastrointestinal disorders, and atherosclerosis. Other components of the nutrition program are devoted to technology transfer, nutrition education, training, and coordination.

Additionally, clinical nutrition units (CNUs) will be created to serve as focal points for clinical nutrition research at medical schools and other institutions conducting biomedical and health-related research.

All these initiatives will strengthen the Federal Government's capacity to respond, over the next five years, to the nutritional needs of the population and to develop effective policy and programs related to nutrition.

TECHNOLOGY TRANSFER

"Consensus Development"

Biomedical research and technological innovation have been responsible for profound improvements in health and medical care. However, a number of technologies have been widely incorporated into medical practice and have persisted although they are only marginally useful or even harmful; still other, well-validated innovations have been slow in being utilized in clinical medicine.

Certain health technologies, such as immunizations, have produced tremendous savings by preventing illness; others may increase the need for personnel and simply add to the "intensity" of care, without contributing significantly to outcomes. For example, the rapid dissemination and utilization of some new technologies, such as CAT scanners and coronary bypass surgery, have resulted in increases in cost without adequate evaluation of their efficacy or cost-effectiveness. Under what circumstances, then, is a given technology justifiably utilized?

To address this issue, the NIH, in October 1977, established an Office for Medical Applications of Research (OMAR), which was charged with coordinating and facilitating NIH technology assessment activities. The process utilized for assessments is "consensus development"—an agreement reached by a majority of medical experts, representatives of the public, and others on the safety and efficacy of technologies. Consensus panels assess the scientific information underlying the clinical significance of new findings, adequacy of studies for efficacy and safety, need for additional studies, and state of readiness for use of the technology in medicine. Twelve consensus meetings were held in 1977-78 on a diversity of topics: breast cancer screening, asbestos exposure, dental implants, standards for laboratory use of toxic substances, burn therapy, surgical treatment of morbid obesity, and other subjects. More than 25 consensus conferences were planned for 1979.

The consensus development activities at the National Institutes of Health have been underway for over one year, and the impact of the earliest conferences has already been felt. The breast cancer screening conference in September 1977 is an example. As a result of that conference the screening program for breast cancer is ongoing; the prescribed Informed Consent Record now includes the dose of radiation received; the routine use of thermography was terminated; women 40-49 years old are being screened routinely if they have a familial or personal history of breast cancer; women 35-39 years old are being routinely screened only if they have such a personal history; and since April 1977, screening for women under 50 has been reduced 65 percent to a level commensurate with the number of women in the risk categories.

As the number of assessments increases and as the dissemination of information expands, immense benefits to health care, its quality and financing, could result.

Information Transfer for Practitioners

Providing information on new research findings in forms useful to health care practitioners is considered vitally important. A newly defined program within the Lister Hill National Center for Biomedical Communications (the research and development component of the National Library of Medicine) is addressing this problem by developing, demonstrating, and evaluating a prototype computerized information transfer system responsive to practitioners' needs.

As a comprehensive bank of information, this system is to provide immediate, current, and substantive answers to questions posed by practitioners. The answers are the consensus of a group of experts; they are accompanied by supporting data and citations to primary publications for more detailed study if desired.

The disease "viral hepatitis" has been selected to serve as the initial test model, and a data base suitable for automated search and retrieval techniques has been constructed. At the same time that expanded demonstration and evaluation of the Hepatitis Data Base is continuing, other applications of the present methodology will be explored.

The computerized data base presently contains computer-generated graphics, which greatly enhance the text material. Other research activities of the Lister Hill Center are exploring telephone access to data base information contained on audio tapes, using remote input from the telephone touchtone key pad for automated selection and tape control, and using advanced storage and retrieval technologies such as the videodisc to provide a broad range of digital and audiovisual information.

MANAGEMENT OF SCIENCE RESOURCES

Essential to the resolution of the health problems described above are the resources—funds, facilities, manpower—that undergird the nation's biomedical research effort. Significantly, all of the panel reports of the recent National Conference on Health Research Principles (discussed earlier) contain concrete suggestions for protecting and enhancing research capability to assure future health gains. As noted previously, the general—almost unanimous—conclusion was that the science base, that combination of basic and applied biomedical and behavioral research and its supporting resources on which future progress in health care depends, must be protected from erosion, that it should be restored to previous levels of vigor, and that Federal support of this base must be stabilized over the long term. This is in consonance with the national goals enunciated in P.L. 94-282, Section 102(c)(3):

> Federal promotion of science and technology should emphasize quality of research, recognize the singular importance of stability in scientific and technological institutions, and for urgent tasks, seek to assure timeliness of results. With particular reference to Federal support for basic research, funds should be allocated to encourage education in needed disciplines, to provide a base of scientific knowledge from which future essential technological development can be launched, and to add to the cultural heritage of the Nation.

As the Secretary, HEW, noted in his address before the Annual Meeting of the American Federation for Clinical Research in April 1978, the building of our current capacity for biomedical research has involved a 25-year investment of the American people, and to jeopardize this investment so carefully assembled would not be in the public interest. Clearly, the leveling off of Federal funds for research presents difficult new problems which will require imaginative solutions if the high level of excellence built over the years is to be sustained.

Two of the more urgent problems which will require special attention in the next five years are the deterioration of scientific facilities and equipment in the national research enterprise and the decline in research opportunities for young scientists. These are briefly considered below.

Physical Facilities

In medical schools and research institutions throughout the country there exist research facilities that require renovation, rehabilitation, or replacement. In addition, major equipment in some instances is becoming obsolete.* It seems reasonable to suppose that the degree of Federal involvement in health research places on the Federal Government the primary responsibility for the development of a strategy to assure the adequacy of physical facilities for the nation's health research enterprise. To conserve resources, emphasis must be placed on sharing of equipment, both within and among institutions. Methods must also be developed to stimulate greater ingenuity in the design and construction of laboratories and buildings. Greater flexibility should be a hallmark of future construction. Given the characteristics of changing requirements of research, high costs, and long life expectancy, the dollar threshold of instrument cost at which sharing would occur must be carefully weighed. Other considerations, such as energy conservation and lowest possible maintenance costs, should also be recognized.

* For a discussion of the status of academic research facilities, see B.L.R. Smith and J.J. Karlesky, *The State of Academic Science: The Universities in the Nation's Research Effort*, Change Magazine Press, New York, 1977, pp. 162-178.

The Young Investigator

Research is often held to be an activity of the youthful mind. Increasingly, the threat of inflation has deterred older investigators from retiring while the paucity of other competing job opportunities diminishes defection from science. If simultaneously the opportunities for the youthful investigator are decreased, then inevitably the average age of the population concerned with research will move upward.

The suggestion has been made that it would be desirable to develop far closer ties than presently exist between the university and medical school campuses of this country and the National Institutes of Health. A far higher degree of personnel mobility between these sites might permit scientists to maximize their contributions to the system at various stages of their careers and at the same time to derive maximum satisfaction to themselves. The alternate argument has been made, however, that academia's value to society would be greater the more independent of the Federal establishment it is able to be. Greater thought must be devoted to defining a better partnership between government and academia; their current relationship has been characterized as fragile and not well balanced.

Some critics have suggested that a modification of tenure system may be necessary to offset the aging of established faculty and provide more positions for younger scientists, and it is clear that tenure is being regarded anew by academic institutions. However, at this time such restructuring seems highly problematical.

Over the next five years careful consideration will be given to the selection of fields of activity for the Ph.D. candidate and the postdoctoral fellow in order to avoid overproduction in fields with limited job opportunities and possible underproduction in those with a higher demand. This matter has been under the continuing scrutiny of a committee of the National Research Council and NIH and, as opportunities for young scientists decline, will require especially urgent attention. Currently, the awarding units of NIH are initiating program planning papers to define mission priority areas. These are slated for discussion with the units' various advisory bodies.

RESEARCH CONSTRAINTS

Cognizance is accorded the changed condition of a world of increased resource constraints. Biomedical research cannot be immune from such constraints. However, the consideration of the appropriate level of research activity must include both an awareness of the unique Federal role in this field and the potential to achieve important national health goals with modest investment. It is primarily through appropriate research in the causes, prevention, and treatment of disease that current medical practice can be improved and that cost containment can be achieved. Some potential exists for funding through a program of partnership between the Federal Government and voluntary health agencies, philanthropic foundations, state and local governments, and industry. Such joint efforts will be carefully explored, as will opportunities for international collaboration in programs where costs are great and might be shared among the nations likely to benefit.

Department of Housing and Urban Development

As its name connotes, the Department of Housing and Urban Development (HUD) is concerned with a broad spectrum of issues related to improving the country's communal environment. These concerns range from soaring housing costs and free choice in housing to revitalization of urban areas and single-family use of solar heating.

HUD's research program focuses primarily on policy-relevant projects of an applied nature. This thrust is dictated by the scale and urgency of the daily problems confronting HUD. Consequently, most of the Department's science and technology activities are classified as "applied research."

The following sections describe HUD's research in economic development, public financing and tax policy, housing services to people with special requirements such as the elderly and handicapped, and the preservation and revitalization of neighborhoods. Other topics discussed are fair housing, financial aspects of housing, management of federally assisted housing, evaluation of HUD's operating programs, and the projected expansion of the Department's data base and information sources.

URBAN ECONOMIC DEVELOPMENT, PUBLIC FINANCE, AND TAX POLICY

HUD's research in state and local economic development and public finance assists state and local governments in dealing with problems of development, private employment opportunities, revenue, and finance. Solutions to the urgent problems of housing, neighborhood decline, and loss of vitality of central cities depend directly on local economic development efforts that can attract private investment, generate employment and entrepreneurial opportunity, and sustain the local tax base. The current fiscal crises of many local governments underscore the interrelated need to improve local government finances and the efficiency with which services are provided.

During the next five years research on current and emerging national issues will be undertaken in many areas. In urban economic development, research will focus on zoning and subdivision practices, developable land supply, processing development approvals, state urban strategies, comparative development costs, state and local land use commissions, public land acquisition for private development, regional strategies, cost of development regulations, and urban infill and rehabilitation opportunities.

Areas of study in public finance will include urban impact analysis, urban capital infrastructure, urban development act grants evaluation, national development bank impacts, urban policy impacts, urban service sector studies, the economic future of central cities, public pension issues, and private sector barriers to economic development.

Topics warranting study in the area of tax policy are the development of finance incentives, the equity of state and local taxes, state and local revenue and expenditure limitations, state and local tax policies as they relate to the development of central cities and suburbs, and the incidence of state and local taxes on central cities and suburbs.

HOUSING NEEDS AND SERVICES FOR SPECIAL USERS

The focus of the HUD Special User Research program is on providing housing and housing-related services to people who have special requirements caused by such factors as age, physical condition, cultural background, or location in a rural environment. Research sponsored by this program seeks to identify impediments to independent living, to develop information about the need for the delivery of housing and special services, and to design and implement better techniques for their delivery.

While the main populations considered are the elderly, the handicapped, and rural residents, research is also sponsored on the needs of such diverse groups as occupants of single rooms, Indians, and Alaskan natives. The basic research questions addressed for all of these target groups are similar: What are optimal strategies for helping special groups obtain suitable housing? What service package is essential to the success of special housing programs? Are there special management or administrative issues that must be addressed? What funding mechanisms can be used to provide the housing and services? What changes of law or regulations are necessary?

The emphasis of this program will be on the collection of data on the housing situations of both the elderly and the handicapped and on means to help alleviate their housing problems through the provision of specialized housing or through services delivered to housing located in the community. One of the primary issues to be addressed in this research is how to help the elderly and handicapped remain independent in their current housing situations.

NEIGHBORHOOD REINVESTMENT AND REVITALIZATION

Despite considerable research, many issues related to the stability and conservation of older inner-city neighborhoods remain unresolved. Incomplete understanding of the dynamic elements of neighborhood change seriously hamper public efforts to preserve the housing stock and to maintain an acceptable quality of life for inner-city residents. Cities are experiencing continued decline and disinvestment in some of their older areas, while simultaneously other areas are showing signs of fairly intense revitalization. Although there are many extant theories of neighborhood decline, there is little empirical or theoretical knowledge of the forces that lead to the natural reclamation and revitalization of blighted neighborhoods.

Neighborhood Dynamics

A full understanding of neighborhood dynamics requires that the neighborhood be examined as one component of a more complex system. General research into intracity and intrametropolitan population mobility is an important component of this work. Studies of the factors that lead to population movement are needed to improve our ability to predict the effects of exogenous changes in the social and economic setting on decisions about household location. Research into neighborhood change should examine interneighborhood relationships in a comprehensive framework. We need to know not only why some neighborhoods decline while others revitalize, but also the impact of revitalization of particular areas on the balance of a metropolitan area.

Understanding Neighborhood Change and Preservation

An effective understanding of neighborhood change and preservation also requires interdisciplinary study. Neighborhoods are complex social units which relate to each other through the metropolitan political and economic systems. Rarely, however, do studies of neighborhood change include complementary analyses of political, economic, and social dynamics. Considerable effort should be placed on the merging of these disciplines in future work on neighborhood change. Much of the current work in neighborhoods relies on economic analyses of housing market dynamics to provide measures of stability and revitalization, with little recognition that a critical component of neighborhood health is the social interaction. These may not be revealed through economic indicators. A failure to realize their importance or to include them in studies of neighborhood change restricts our ability to set realistic goals for neighborhood preservation efforts and to understand fully the best programmatic approaches to achieving neighborhood health.

Neighborhood Preservation

Technical issues related to the processes of achieving neighborhood preservation require exploration. Housing rehabilitation is a key element. Among the relevant issues are the costs of rehabilitation, the kinds of structures amenable to rehabilitation, and the impact of local building codes and other regulations on cost and feasibility. As important as the state of individual houses is the larger physical environment of the neighborhood, which affects the quality of life of the residents and the long-term viability of the area. Required are studies of the relationships between specific environmental components of the neighborhood and the needs of residents, as well as analyses of the community design process and methods, and the policies and regulations that shape the environment.

ECONOMIC AND RACIAL FREEDOM OF CHOICE IN HOUSING

HUD's research in the area of fair housing and equal opportunity took a major step forward with the field audit and evaluation of Title VIII of the Civil Rights Act of 1968. The research has been extended to Hispanics and discrimination against women and minorities in mortgage credit markets. Within the five-year span, research on freedom of choice in housing will be extended to examine search behavior of minority home-seekers and further examine real estate and rental agent behavior.

Fair Housing Evaluation

HUD's fair housing evaluation study provided the first detailed documentation of the extent of discrimination in American housing markets. It identifies the form that discrimination against blacks is most likely to take and indicates under what conditions discrimination is most likely to occur. The evaluation will continue through fiscal year 1979 and will assess the effectiveness of HUD's conciliation activities.

The fair housing audit and evaluation will be repeated, beginning fiscal year 1980, to determine what changes have occurred between 1977 and 1981 in the level, form and incidence of discrimination against blacks in housing. The time for the reevaluation has been chosen for best use of census data. As part of the reevaluation, the effectiveness of HUD's conciliation efforts will also be reassessed.

The fair housing audit is providing important information on the magnitude and nature of the discrimination problem faced by blacks in the housing market; similar information is needed on the problems experienced by Hispanics. It will also be worthwhile to investigate discrimination in those portions of the housing market not covered in the original audit study, for example, homes for sale by builders and condominiums.

Programs for Women and Minorities

The new Women and Mortgage Credit project, developed in 1978, is designed to facilitate the access of women to home ownership by directing an educational campaign at two populations: the lending industry and women consumers.

At the heart of the lender education campaign is the preparation and dissemination of statistical and empirical data on women's work patterns, income stability and growth, economic independence, etc., to document women's credit-worthiness.

For potential borrowers, the project will undertake to educate women in sample target cities about the world of finance, credit, and mortgages, all too often regarded by women as a male world and one to be feared. In the sample cities, activist women's organizations and organizations with large numbers of women members having regional and national ties—labor unions, for example—will conduct workshops and seminars for women on housing options, the advantages and disadvantages of home ownership, alternate mortgage instruments, and how to acquire a mortgage. (When appropriate, materials written in Spanish will be made available at seminars held especially for Hispanic women.)

Along with efforts to attract and train minorities for research work in housing and urban studies, HUD has initiated research to develop a program to help women and minorities gain greater access to high-level professional and administrative positions in local government and in state and Federal agencies. It is unlikely that any significant change will occur in the documented under-representation of women and minorities in decisionmaking positions at all levels of government without the intervention of affirmative action programs designed to influence the urban management market. HUD's traditional relationships with local governments suggest a unique role for the Department in developing such programs. Such work also addresses the President's goal of increasing the number of women and minorities in public service. This program of research is at an early stage of development; further consultation and participation by representative groups of potential users are expected, which will help assure relevance and ultimate acceptance by such groups.

ALTERNATIVE HOUSING FINANCE MECHANISMS

TOPICS IN HOUSING AND MORTGAGE MARKETS

Housing Costs

One of the most troublesome problems in the housing market has been the soaring cost of new and previously occupied homes. The creation of a special task force on housing costs indicates concern over the many issues involving affordability, real investment returns, and even the measurement of real housing costs.

Housing Subsidies

Rising housing costs have significantly increased the problem of adequately housing this nation's low-income households. Such rising costs lead to increased demand for Federal subsidies; at the same time, the Federal Government is also under increased pressure to balance its budget for the good of the overall economy. To meet both goals, there is a need to ensure that the methods of providing subsidies are most efficient.

Financial Innovations

In 1978 several financial innovations evolved to increase competition among financial institutions. Share drafts, negotiable orders of withdrawal (NOW) accounts, and money market certificates of deposit for thrifts and commercial banks have increased the competition for loanable funds and deposits. In addition, new mortgage instruments such as the variable-rate mortgage, graduated-payment mortgage, and reverse annuity mortgage have been developed to provide special benefits for both borrower and lender. The effectiveness of each of these innovations in expanding homeownership opportunities as well as insulating the mortgage market from the ravages of adverse credit cycles will need to be reviewed before further extensions and innovations can be devised.

Financial Regulation

Although the Financial Institutions Act of 1976 was not passed, several of the reforms contained in that bill have been enacted, and the remaining proposals are receiving serious consideration. Chief among the proposals is one to eliminate the regulation placing ceilings on deposit rates of commercial banks (Regulation Q). The provision for money market certificates of deposit has had the effect of eliminating Regulation Q for accounts of more than $10,000. However, consumer groups, especially the Gray Panthers, have attacked the remaining interest ceilings as discriminatory. Consequently, recent regulatory changes as well as proposed changes need to be examined for their impact on the efficiency and well-being of the financial markets and their impact on lenders and consumers alike.

Lending Discrimination

Freedom of opportunity in housing for everyone requires not only that there be no discrimination by sellers but also none by lenders. This is important for individuals and for local neighborhoods and jurisdictions as well. The loss of available financing for specific locations often leads to further decay and increased lending risks. Not only should research pinpoint areas of proven redlining, but positive programs must be developed to combat the conditions that may be alarming to private lenders.

Financing Solar Homes

As energy resources become more costly, solar energy must be increasingly sought after to fill our energy needs. Studies depicting the economic feasibility of financing single-family solar heating systems could be instrumental in encouraging the demand for such energy-saving systems.

HOUSING ASSISTANCE FOR LOWER INCOME FAMILIES

The provision of opportunities for lower income households to obtain decent, safe, and sanitary housing at a reasonable cost is now and will continue to be a major part of the mission of the Department. To better carry out this mission, HUD is evaluating a number of housing assistance programs as well as amassing an extensive amount of information and experimental data concerning the effects of alternative approaches to housing assistance.

Section 8 Housing Research

Over the next five years, HUD will continue to evaluate the effectiveness and efficiency of the Section 8 Housing Assistance Program and the housing choices made by the participants in that program. The first phase of this research effort concentrated on the early experience in the "existing housing" portion of the program. A more comprehensive evaluation is now beginning and will continue over the next few years to examine all aspects of the Section 8 program, including the new construction and rehabilitation portions. Urban and rural sections of the country will be investigated separately and the results compared.

Experimental Housing Allowance Program (EHAP)

EHAP has gathered information on the responses of over 25,000 participants at 12 sites to the offer of direct cash assistance for housing. Households could receive allowances if their dwelling units met certain minimum quality and occupancy requirements. The analysis of EHAP data reveals a great deal about the impact of a number of variations in providing housing assistance.

All survey data have been collected, and the extremely large and complex data base will be a major source for housing researchers for several years. Features that make this data base of great value include longitudinal surveys of households, dwelling unit evaluations, landlord surveys, and surveys of various suppliers of housing services. The primary analyses of the data are still underway, with most final reports due during 1979. The last reports, completing the EHAP as originally designed, will be finished in 1981. Other analyses made possible by the richness of the data will also be conducted over the next several years. The information from EHAP and analyses of its data will be coordinated with results of the Section 8 research program, which is gathering similar data. This will provide a firm basis for policy decisions on how best to provide adequate housing for lower income households.

MANAGEMENT OF ASSISTED HOUSING

In recent years, federally assisted low- and moderate-income multi-family housing—both public and HUD-insured—has been faced with a number of severe operating problems. The financial viability of these projects has been threatened, and their effective management impeded. Such problems as rising operating costs, inadequate project revenue, deteriorating physical structures and surrounding neighborhoods, and ineffective management have created serious distress in these projects. These problems have led to inadequate housing services for project residents, neighborhood instability, and, in the case of HUD-insured projects, mortgage default. The research program for assisted housing management is designed to address these problems through the identification of the nature and causes of distress and through the development, demonstration, and evaluation of means to alleviate distressed conditions and improve management practices. Whereas much of the early research in this area involved the testing of specific remedies for individual housing management problems, current research activities are aimed at devising broader strategies for improving the financial health of assisted housing, providing better quality housing services, and maintaining the integrity of the physical stock. For projects that are in default, the research emphasis is on finding methods of property disposition that maximize resident and community involvement.

EVALUATION OF HUD'S OPERATING PROGRAMS

The following studies are typical of the type of research involved in the evaluation of all operating programs.

Community Development Block Grants (CDBG) Strategies Evaluation

A study is underway to determine under what conditions the most important activities funded by CDBG are effective in improving the housing and living conditions of neighborhoods, and to identify those who benefit from the activities. The study, which will also investigate the role of citizen groups in implementing neighborhood community development programs, began in fiscal year 1978 and will continue through fiscal year 1981.

Evaluation of CDBG in Small Cities

Another study to begin in fiscal year 1979, will give special attention to the extent to which multi-year commitments have contributed to the success of the CDBG program in small cities. It may include an analysis of CDBG decisionmaking processes in small cities. In addition, it may gather information on the impact and incidence of benefits of CDBG-funded activities in small communities.

Development and Implementation of Evaluation Tools

The Department has invested considerable resources in developing urban simulation models to evaluate HUD policies. Until recently all simulation work was performed under contract. The Urban Institute housing model is now operational and directly accessible to HUD's own research staff; a second model will be added by the end of fiscal year 1979.

Contractor resources will be devoted to improving the usefulness of these models while staff resources will be devoted to using the models for evaluation and policy analysis. Efforts will also be made to modify the Urban Institute's Transfer Income Model (TRIM) for use in analyzing the impact of HUD programs. The study will be conducted during fiscal year 1979.

Section 701 Evaluation

The evaluation of the comprehensive planning assistance grants program (Section 701, P.L. 83-560, Housing Act of 1954) will expand on the CDBG study in analyzing various aspects of the local planning process. In particular, it will examine specific planning and citizen involvement techniques. Efforts will be made to identify deficiencies in local data sources and planning techniques. Emphasis will be on the extent to which Section 701 has improved local planning and whether the improvements have affected the local community development process.

Evaluation of HUD's Environmental Quality Policies

Environmental regulations have been criticized as adversely affecting the timing and cost of housing production and as being ineffective in promoting better living conditions. It is appropriate at this time to initiate an evaluation of HUD's environmental protection policies. The study will focus on two issues. First, HUD's environmental regulations and procedures will be examined for their consistency with recent work on the theories of exhaustible resources and of congestible public goods. Second, an effort will be made to quantify the major costs and benefits associated with environmental regulations. In addressing benefits, the study will rely on the body of literature being developed and refined concerning the techniques for determining willingness to pay for environmental quality. It can be expected that the measurement of costs will be more straightforward: additional expenditures incurred by builders to meet environmental requirements and those additional holding costs incurred because of the time involved in environmental review are examples.

IMPROVING HUD'S DATA BASE AND INFORMATION SOURCES

Data Collection and Analysis

The basic housing production and marketing surveys conducted by the Bureau of the Census for HUD will continue in essentially their present form. They are the surveys of new housing completions, new one-family home sales, market absorption of multifamily rental apartments and condominiums, and new mobile home placements.

The major data collection effort is the Annual Housing Survey of the nation's housing inventory and occupants, and of the housing stock of 60 selected metropolitan areas, including most of the 50 largest metro areas. These surveys will continue unchanged, except for modest annual changes that may be made in the types of data collected to reflect current concerns and the need for a broad range of subjects to be investigated at less than annual intervals. After 1980 decennial census data become available, there will probably be some restructuring of the Annual Housing Survey sampling frame, to take advantage of the 1980 data baseline and to reflect current definitions of metropolitan areas. To the maximum extent possible, however, the year-to-year comparability of the data file will be maintained.

Work will continue on analyzing the data obtained from these surveys and from complementary sources to answer policy questions about the housing stock, including the nature of inventory changes and housing needs, changes in housing and neighborhood quality, housing cost burdens, and household migration by race and income.

Economic Analyses

Current studies of tax policies as they affect housing production and maintenance are scheduled to produce results before 1981 when many tax benefit provisions affecting housing expire. Research beyond 1980 will continue on housing market effects of tax policies, and may address other issues if changes in tax provisions require research.

Research will continue on general housing economics issues and development of new analytic techniques for analyzing housing market behavior. Availability of 1980 census results will offer new opportunities to analyze the differences among markets in patterns of mobility and housing demand.

Real estate settlement practices research will be largely completed by 1980, but concerns about transaction costs and their influence on the demand for various types of houses and the quality and condition of the housing being transferred could generate some additional research into 1983.

Department of the Interior
Bureau of Mines

As the Federal Government's primary agency for research in nonfuel minerals, the Bureau of Mines conducts scientific and technological investigations into mining, processing mineral substances, eliminating occupational health and safety hazards in the mineral industries, minimizing the adverse environmental impact of mining and mineral processing, economic development, and resource conservation.

Bureau activities cover a broad spectrum of programs for meeting the diverse needs of mineral industries, government, and the public. In the 1960s, the Bureau accelerated programs to resolve problems associated with past mining and mineral processing. In the 1970s, the energy crisis raised the national consciousness concerning our limited supplies of domestic mineral and fuel resources. Dependence on foreign sources of raw materials critical to our national economy and competition from other countries for these world supplies stimulated the need to conserve and use wisely the nation's minerals and materials. For the 1980s, the Bureau will continue to apply its mining and mineral processing expertise to research and development covering broad areas of minerals technology, health and safety technology, and environmental technology.

The following sections briefly list some of the problems requiring attention by the Bureau and identify significant technology-based activities to resolve these problems. The sections highlight research and development for the next five years in minerals environmental technology, minerals health and safety technology, and mineral resources technology. These sections reflect the reorganization of Bureau of Mines programs that was accomplished in July 1979.

MINERALS ENVIRONMENTAL TECHNOLOGY

Strong program emphasis is given to providing technology to facilitate compliance with environmental regulations and performance standards promulgated by the Clean Air and Clean Water Acts, as amended; the Safe Drinking Water Act; and the Surface Mining Control and Reclamation Act of 1977. These laws have dramatically increased the need for environmental abatement and control technology in the entire mineral cycle from initial exploration and planning, through active mining and processing, to final restoration of the land and reclamation of the processed materials. In recognition of this, the Bureau's program is structured into three categories: minerals environmental assessment, control technology—operational, and control technology—post-operational.

MINERALS ENVIRONMENTAL ASSESSMENT

The opening of a new mine or processing operation runs the gamut from the first shovel of earth removed through the last shovel of earth replaced, and requires anticipating the environmental effects before initiation of mining and processing, monitoring the actual effects during operations, and eliminating permanent adverse changes after operations have been completed.

Baseline Studies

Research is underway to evaluate the impact of mining on surface and groundwater systems. Regional data on existing hydrologic conditions are being modeled, and the effects of introducing mining operations into these areas are being analyzed. Two specific areas of current concern are the Powder River Basin in Wyoming and the Texas lignite region.

Studies with the National Academy of Sciences are taking a broad look at soil conservation from both an agricultural and a mining standpoint and at problems involving groundwater and surface coal mining. These studies will provide methods for identifying and evaluating the impacts of surface mining on soil and water resources. This will lead to suggestions and recommendations for future research and policymaking. As a result of the hydrology study, recommendations will be made for guidelines to help mine operators minimize the impact of mining on the hydrologic environment.

Environmental Planning and Analysis

Recent legislation has significantly increased the requirements for the gathering of baseline data and monitoring of possible environmental impacts, and research is being conducted to develop more efficient instruments and techniques for obtaining the necessary data. The development of efficient techniques and systems will result in complete and economical monitoring methods, accelerate the gathering of on-site data, and reduce the overall impact of exploratory investigations on a particular site.

Research is being conducted to develop or improve the planning and design of mines, including work on tools to obtain environmental data, techniques for assessing the data, and technology for measuring changes. Sufficient technology does not exist in many areas of environmental work to accurately establish the natural state of the environment as a basis for determining the impact of existing or future mining or mineral processing operations.

CONTROL TECHNOLOGY—OPERATIONAL

Each activity during the active phase of a mining or mineral processing operation places a variable demand on land, water, air, energy, and materials with the net result that the quality of life may be compromised by air, water, and noise pollution; loss of amenities or aesthetic value; socioeconomic disruptions; and ecological damage. The nature and extent of environmental impact imposed by the minerals cycle depends largely upon the size and method of operation, location, climate, geology, and topography. The major problems under study include alternative processes; noise, vibration, and fugitive dust; subsidence; and waste and water management.

Alternative Processes

Technological developments can help lessen the undesirable environmental conflicts, impacts, and occupational hazards associated with mineral processing. Research is directed toward building a better technologic base for mineral processing systems so that appropriate environmental regulations and control technologies can be implemented to prevent environmental degradation without needlessly restricting efforts by the minerals industry to meet the nation's mineral needs.

Research will be conducted to provide metallurgical engineering information for process control to enhance the recovery of secondary constituents of ores and to reduce the need for postprocess environmental controls. Techniques to permit the reuse of emission and effluent constituents will be investigated. Technology will be devised to retard the generation of undesirable emissions and effluents from metallurgical processes.

The achievement of selected research objectives for the 1979 to 1983 period will provide the expertise for particulate mineralogy analysis and mineralogical identification. New and improved methods will be devised for identifying and measuring mineralogical and metallurgical particulates present in mineral processing streams, and for reporting characteristics of fibrous nonasbestos minerals. An ion flotation technique to reduce and recover dissolved metals in water wastes so that the water can be reused in the metallurgical system will be tested. Methods will be devised to remove dissolved or suspended solids from selected metallurgical process water streams to increase the potential recycling of water or make the solutions more acceptable for discard. Field testing will be completed on flocculation thickening systems for dewatering alumina red mud wastes to increase water recycling and reduce environmental problems associated with the disposal of the solids. Laboratory technology will be provided for agglomeration of various particulate wastes to facilitate recovery of metal values or neutralization of hazardous constituents. New metallurgical engineering and process control technology will be developed for the treatment of water circuits to improve flocculation of solids and to increase reuse of process water.

Noise, Vibration, and Fugitive Dust

Research on environmental noise and dust problems includes source identification and characterization and the development of suitable methods of abatement. As part of its long-range effort, the Bureau continues working toward a complete understanding of the generation, propagation, and effects of air blast and ground vibration in order to prevent or minimize structural and environmental damage resulting from mine blasting.

Subsidence

In the area of subsidence research, the Bureau's program consists of four distinct elements: prediction, control, prevention, and damage abatement. To aid in prediction, new and existing data are being utilized to develop empirical equations to describe ground movement accompanying specific underground mining practices in various geologic and topographic situations. Complementary work is directed towards the adoption or refinement of currently available methods of analyzing rock structures. This two-pronged approach to subsidence prediction is believed to be the most effective means of providing for a basic understanding of the phenomenon of subsiding ground.

Guidelines are being sought for the methods that control or prevent subsidence, based on state-of-the-art technology. Hardware for cost-effective monitoring of ground movement, detection of subsurface instability, and delineation of underground voids is also under study. In addition, a substantial effort is being expended in studying the mechanics of subsidence-related damage to surface installations and the effects of subsidence on surface and groundwater resources.

Waste Management

Cooperative work with the Environmental Protection Agency is being undertaken for the monitoring of waste sites in order to define the best management practices for waste disposal systems for mineral operations. Economic effects of regulations governing mining, processing, and disposal operations are being evaluated, and, based on this, more cost-effective compliance techniques are being developed. Research is underway also in the evaluation of environmental effects associated with the potential development of a commercial oil shale industry.

Technology is being sought for the effective underground and surface disposal of mine wastes and tailings. Based on the results of a feasibility study performed under contract, a conceptual underground disposal system for coal mine wastes is being considered for further design and demonstration. Successful implementation of the system should result in greater resource recovery, abatement of subsidence, and a reduction in land requirements for surface disposal. Currently, studies are being undertaken to develop feasible methods of underground disposal of wastes from oil shale retorting and uranium mining. To aid in the construction of stable impoundment structures for tailings, methods of seepage monitoring and control are under study, and the consequences of liquefaction-type failures are being evaluated.

Water Management

Research is being carried out to find ways to minimize the pollution of surface and groundwater resources resulting from mine drainage, haul road sediment, erosion, seepage from tailings, and in situ leaching operations. Leachants other than ammonium carbonate will be evaluated for in situ uranium extraction, and methods for modeling and detecting leachant movement are being developed and demonstrated. The use of slime-sealed impoundment will be demonstrated for seepage control in tailings ponds, and in water-scarce areas methods of water handling, treatment, recovery, recycling, and diversion are under evaluation.

CONTROL TECHNOLOGY—POSTOPERATIONAL

Improved technology is being developed to facilitate the restoration of lands disturbed by mining and milling to a selected beneficial long-term use. Involved in this effort is the study of land use alternatives, safe mine closure methods, reclamation technology, and surface stabilization techniques.

Land Use Alternatives

Research efforts are seeking to identify methods for abandoning mines in a manner to enhance land use, develop analytical procedures for selecting land use alternatives, identify alternative methods of reclaiming waste rock to enhance final land use, and develop criteria for optimal postmining surface utilization.

Mine Closure

Final mine closure techniques for both surface and underground mining operations are being examined to identify the best methods for attaining ultimate land use, and this research also seeks information to be used in determining final land use. These activities concentrate on defining closure procedures for surface and underground mines through the design of effective mine barricades and the development of cost-effective sealing, stowing, and hydraulic isolation techniques to alleviate acid water formation.

Reclamation Technology

The biological and physical problems of reclaiming areas disturbed by mining and milling operations are under study, and cost-effective equipment systems to facilitate compliance with environmental standards will be developed and demonstrated. To facilitate revegetation of mining-disturbed lands, such methods as selective irrigation, reestablishment of original cover, isolation of impermeable strata from the subsoil, and alternative methods for regrading and reclaiming waste rock and tailings are expected to be demonstrated in the near term. Emphasis is placed on developing and demonstrating techniques that assure the long-term physical, chemical, and biological stability of reclaimed land.

The interrelationships between mined land reclamation activities and water quality, such as acid mine drainage, toxic material leaching, sediment delivery, and the quantity and quality of surface and groundwater, are also being examined. Alternative abatement techniques to inhibit and eliminate formation of acid mine water and leaching of toxic materials are being evaluated, and it is hoped that cost-effective methods of identifying toxic overburden materials and improved sediment pond abandonment techniques will be developed.

Surface Stability

Problems associated with the long-term stability of waste minerals are under analysis in order to enhance existing stabilization techniques or devise new techniques so that stability can be predicted, if not assured. The problems associated with stabilization of new surface spoil materials are being examined, and methods are being sought to reduce erosional and consolidation processes and their adverse impact on air and water quality and on surface use. Analytical laboratory techniques are being sought to define spoil material properties for predicting deformation, slaking, and stability.

MINERALS HEALTH AND SAFETY TECHNOLOGY

Research is conducted in all of the areas specified by the Federal Mine Safety and Health Amendments Act of 1977 (P.L. 95-164). The Act sets requirements and authorizes appropriation of funds for three necessary elements of a balanced attack on the problems: comprehensive enforcement of existing and new mandatory health and safety standards, expanded and upgraded health and safety education and training activities, and technical support to build a mining technology that will bring long-lasting health and safety improvements. Each of these elements reinforces the others; each works to produce improvements in both the long and the short term. Together, they offer hope for lasting solutions in health and safety throughout the minerals industries.

HEALTH-RELATED ENGINEERING RESEARCH

As technology develops to improve production, there is an accompanying need to protect and improve the health of coal and other miners. The necessary technology is being developed to reduce health-related hazards in mines and to afford adequate protection to the miners. Research is focusing on such problems as respirable coal dust, radioactive gas, noise, ventilation, and toxic mine gases and on solutions involving both personal protection and control of the mine environment through engineering advances.

Dust Control

Through the next five years, particular attention will be directed at accelerating the development of a machine-mounted dust monitor that can be used for enforcement of the dust standard in coal and metal/nonmetal mines, developing improved methods of analyzing silica and fibers because of the lower dust standards that have been proposed, emphasizing the application of existing dust collectors to mining machines, and developing high-pressure, water-powered dust collectors.

Radiation Hazards

Research is continuing on the radiation hazard, which is a major health concern in uranium mines. Long-term exposure to this hazard among uranium miners is known to cause increased probability of lung cancer. Studies indicate that uranium miners have a nonviolent death rate 40 percent higher than that of the normal population. The major thrusts of this program will be to develop and field test portable instruments for measuring the airborne radiation levels and exposures of individual miners, develop and field test a radiation warning system for active uranium mines, develop and demonstrate improved and new control technology, and establish the behavior patterns of attached and unattached radon daughters and other long-lived radioactive materials under various mine conditions.

Noise Abatement

To achieve acceptable levels of noise exposure of personnel engaged in both underground and surface mining activities, research is continuing on noise abatement and instrumentation. This research is becoming more critical because several government regulatory agencies advocate making the current standards more stringent. The principal research effort is directed at reducing noise at its source. Much of the noise abatement effort to date has been applied to developing retrofit techniques. Although this will continue to be necessary, future long-term work will be directed more toward factory integration of control measures and design of quieter machines.

Industrial Hygiene

The thrust of the research in the industrial hygiene program is directed at providing the information, analytical tools, and monitoring of warning devices necessary to provide a safe mine environment for coal and other miners. Particular emphasis is directed at developing small, extremely lightweight equipment that can be worn by miners and can accurately measure exposure to toxic gases. Direct reading instrumentation to continually monitor mine air quality to assure a healthful, safe environment is vital to the well-being of the miner. Long-range plans consider development of instrumentation which must not only be portable but must have multigas measurement capabilities. Such equipment coupled to warning devices would immediately alert the miners to a change in air quality.

Controlling Toxic Emissions From Diesel Engines

Use of diesels in metal and other mines has doubled within the past five years, and their use is expected to become more widespread. Techniques to minimize or render harmless potentially toxic emissions from this source are therefore being investigated. With the increasing awareness of the potentially toxic characteristics of diesel exhaust, the Mine Safety and Health Administration (MSHA) needs adequate instrumentation to assure proper air quality for the miner. Since air quality in mines using diesel engines is dependent upon the condition of the engines, MSHA is requesting that the Bureau undertake development of instruments designed to define the condition of operating equipment. To ensure that diesel operation underground does not pose a health hazard, the Bureau and MSHA have been working together to develop criteria for the safer use of diesels in mines. Such usage dictates the need for improved monitoring instrumentation in the workplace.

SAFETY RESEARCH

As new mining methods and techniques are developed to meet the increased demand for coal, a safer environment and improved working conditions must be assured the miners. The changing use of mechanization is exposing miners to new situations in an already hazardous environment, requiring a continual reevaluation and redesign of protective devices and procedures. This is imperative in order to maintain a stable work force and to attract new personnel into the expanding mining industry. To this end, the Bureau is conducting a balanced research program involving both long-term development of new concepts and equipment and short-term solutions to immediate problems through modification of existing techniques and equipments.

Prevention of Fires and Explosions

Although considerable progress has been made in reducing fatalities from coal mine fires and explosions, the mine environment contains a variety of potential ignition sources and countless opportunities for human error. The fire and explosion prevention program is addressing all problems related to the ignition, propagation, detection, and control of fires and explosions in coal mines. The major program goals include conducting full-scale mine fire demonstrations, characterizing the formation and ignition of methane roof layers that form as a result of different rates of release of methane gas in mining operations, evaluating the effectiveness of passive and triggered barriers in double and multiple entry systems, and installing passive and triggered barriers in working coal mines for in-mine trials.

Methane Control

Another area of research involves conducting studies on the geology and physical features of coal beds in order to develop the engineering criteria to predict the quantity of methane in a block of coal, the pressure under which it occurs, and the factors that govern its movement so as to be able to design effective systems for methane control and ventilation. The main objectives of the methane control research program through 1983 will be to demonstrate to the coal industry the practicality and economic advantages of draining methane through horizontal holes underground and removing it through a piping system, to develop technology to improve face ventilation, and to determine the applicability and effectiveness of water infusion in various coal beds.

Ground Failure

Failure of ground is the major occupational hazard causing injury and death in underground mining. Research in this area is accelerating to identify hazards before or during mining, to develop methods and equipment for controlling these hazards when they occur, and to provide inherently safe mining environments. Identification of known geologic or human induced mining hazards before extraction can lead to avoidance or control measures during mining to prevent accidents associated with these hazards. Geophysical sensing systems, such as probing radar, to detect and locate hazards will be integrated into commercialized units from which semiskilled operators can visually identify hazardous targets before mining.

Work will continue on inherently safe mining designs as all ground failures are not always associated with specific hazards, such as channel sand deposits or abandoned workings. The use of geologic and mechanical property information in a ground support index will be emphasized, and both empirical and analytical mine layout design information will be related to the selection of a support system mated to specific categories of ground stability. Imminent danger ground failure warning systems recently developed will undergo extensive and exhaustive field trials and demonstrations in fiscal years 1980 to 1983 to firmly establish failure thresholds acceptable to MSHA, mine operators, and miners. Microseismic and other failure warning indicators will be integrated into fail-safe systems to minimize both false alarms and undetected failures. Techniques for detection of hazardous roof, rib, or face utilizing simple mechanical and electronic sensing systems will be developed that can be in-mine operated. In metal, nonmetal, and coal mining, rock and coal outburst prediction systems will be finalized in fiscal year 1982. Emphasis on development and demonstration of hard rock-burst prediction equipment and preparation of guidelines for unskilled operator use will be major efforts to help the industry comply with increasingly stringent safety standards.

Industrial Hazards

The broad area of industrial hazards research encompasses human factors, engineering, electrical hazards, illumination, communications, haulage, and materials handling. Increased emphasis will be given to training during the next few years. This emphasis will be accompanied by research in two major areas of concern, namely, human error and behavioral techniques. Training research efforts will be conducted in the following areas: evaluating needs, determining efficiency and effectiveness of training methods, analyzing tasks requiring training, assessing different modes of delivery, establishing qualifications and certifications of mining personnel, retraining, and developing simulator training techniques.

Transportation

Additional research efforts for fiscal years 1980 to 1983 will be directed toward surface haulage equipment and haulage road safety. Efforts will be directed toward establishing criteria for optimized operator compartments. A major effort will be made to develop technology to facilitate compliance with the requirement for protective cabs and canopies in low coal through equipment redesign and modification. Research efforts will also be directed toward improving operator visibility, providing vehicle stability indicators, developing improved tire inspection procedures, improving operator access and egress, and developing criteria for roll-over protection structures (ROPS) on large mining equipment.

Emergency Care

Research is continuing in the development of emergency life support systems and rescue technology, which are needed to maximize the miners' chances of surviving a mine disaster. The long-range thrusts of this subprogram are to develop a substantially lighter weight and smaller volume one-hour oxygen self-rescuer, a trapped miner location and communication system for mines as deep as 3,000 feet, a life support system enabling trapped miners to survive for 14 days, and technology that will significantly reduce the overall time for miner rescue and mine recovery. Based on the location precision of electromagnetic (EM) mine locators as compared to seismic methods, development of a commercial unit using EM location techniques will be accelerated.

MINERAL RESOURCES TECHNOLOGY

Nonfuel minerals play a major role in the economy and national security. Perturbations in the nonfuel minerals economy reverberate throughout the entire nation. Nonfuel minerals research and development play a necessary and potent role in assuring the availability of essential minerals. Yet, the following trends, if not reversed, point to future disruptions in mineral supply: growing dependence on foreign suppliers to meet our basic mineral needs, increased competition for foreign supplies of minerals, lack of innovation in the domestic nonfuel minerals industry with the resulting loss of U.S. leadership in technology, and decreasing Federal involvement with nonfuel minerals research.

This program's goal is to promote a sound economy and national security through new or improved technologies for the development of domestic resources of essential minerals. It encompasses that research necessary to reduce our requirements for critical and strategic minerals through conservation, substitution, and increased recovery of minerals and metals from domestic resources. General objectives for 1979 to 1983 are to

(1) develop minerals technology to extract the maximum amount of mineral values from domestic resources and recover minerals which are bypassed or lost in current processes; (2) develop methods to recover mineral products from large, untapped, domestic resources; (3) remove technological and economic constraints on recovery and reuse of waste and scrap; (4) develop technology to replace critical/strategic minerals with substitutes or alternatives; (5) develop innovative technology to more efficiently mine and process essential minerals needed for a strong economy and the national security; and (6) develop the scientific foundations prerequisite for solving future mineral problems. In support of these general objectives, the Bureau's mineral resources technology program is structured into three subprograms: advancing mineral science and technology, conserving domestic mineral resources, and developing domestic mineral resources.

Advancing Mineral Science and Technology

This subprogram promotes a sound economy and national security by expanding the minerals science base and developing efficient extraction technology. Expanding the minerals science base provides the basic scientific information required for renewing and augmenting the knowledge base from which future mineral problems will be solved. Developing techniques and tools for efficient extraction technology increases efficiency of energy, capital, materials, and labor across the minerals cycle. These developments also strengthen the competitiveness of our nonfuel minerals industry.

Research objectives to expand our minerals science base include developing a handbook for controlling massive rock failures, generating thermodynamic data for use in devising energy-efficient mineral processing sequences, and conducting fundamental studies on the physical chemistry of mineral-reagent interactions in sulfide flotation.

Research objectives to develop efficient extraction technologies include developing a low profile portable crusher; evaluating innovative concepts for ore and waste disposal systems; operating a process research unit for recovery of phosphate from Florida phosphate matrix; demonstrating pit slope caving; completing investigation of hydrometallurgical techniques for recovering platinum, cobalt, nickel, copper, lead, and zinc from gabbro-type deposits; operating a process research unit for sulfuric acid treatment of sphalerite to produce zinc and elemental sulfur; developing more efficient continuous mining for soft and medium hard rocks; and developing casting methods to produce sulfur concrete components for use in corrosive environments.

Conserving Domestic Mineral Resources

This subprogram promotes a sound economy and national security through maximizing mineral resources recovery and scrap and waste utilization. Maximizing recovery percentages of valuable constituents from reserves currently mined and processed reduces waste from current operations. Improvements to existing processes are sought to maximize resource recovery and reduce waste. Scrap and waste utilization provides technology to reclaim valuable constituents from wastes, to recycle useful metals and nonmetallic materials contained in scrap, and to determine requirements for upgrading recovered and recycled materials for use. This research fosters economic scrap and waste utilization as an alternate source of essential minerals.

Research objectives for maximizing resources recovery include demonstrating a retractable bit system; completing research on recovery of cobalt, nickel, and copper from Missouri lead ores; demonstrating well completion techniques for in situ mining; developing flow and chemical reaction models for in situ mining; developing standardized in situ assaying calibration techniques; developing magnetic processes for recovering iron minerals currently being lost in processing plant wastes; and completing evaluation of hydrometallurgical recovery of alumina from dawsonite as a byproduct of oil shale processing.

Research objectives for scrap and waste utilization include developing engineering data for a continuous leaching process for recovering cobalt, nickel, and copper from lead smelter mattes and drosses; devising techniques for sorting scrap stainless steel from aluminum and cast from wrought grades of aluminum scrap; and operating a process research unit for processing of cobalt-nickel flotation concentrate produced from lead smelter slag.

Developing Domestic Mineral Resources

This subprogram promotes a sound economy and national security through use of plentiful resources as substitutes for critical and strategic minerals. Research efforts in this area are intended to increase domestic reserves of critical minerals by developing processing technology for large, domestic, as-yet-untapped resources.. Substitutes can replace critical and strategic minerals with domestically available minerals.

Research objectives for use of plentiful resources include operating a process research unit to recover tungsten from effluents from a plant treating Searles Lake brine, demonstrating recovery of vanadium from low-grade western resources, and evaluating clay as a replacement for bentonite as a binder for iron oxide pellets.

Research objectives for substitutes for critical and strategic minerals include demonstrating a precision casting technique for producing low-cost, corrosion resistant titanium valves; completing studies to develop periclase refractory as a substitute for refractories containing chrome; developing a nickel-cobalt binder for replacing up to 90 percent of the cobalt used in producing cemented tungsten carbide; and completing research on recovering alumina from domestic kaolin and anorthosite by hydrofluoric activated hydrochloric acid leaching.

Department of the Interior
Geological Survey

The program of the U.S. Geological Survey (USGS) in science and technology has as its broad mission that of increasing our fund of knowledge of both the nation's earth resources and the geologic processes that affect their wise utilization. Implementation of this twofold goal involves identifying natural constraints on land use and resource development and analyzing the consequences of alternative policies on resource development, conservation, and environmental protection.

Issues that warrant special attention by the United States during the next five years, and for which the employment of science and technology may provide mitigation or solution, are discussed under two general headings: basic and applied research and application of existing technologies.

BASIC AND APPLIED RESEARCH

Several issues of national significance require substantial increases in research. These issues are energy availability, increased food and fiber production, management of wastes, availability of world and domestic hard minerals, prediction and mitigation of geologic hazards, and development of environmentally sound extraction policies for energy and minerals. On the following pages each of these issues is described individually and some interactions that need further research are then enumerated.

Energy Availability

Increased reliance on imported petroleum in recent years has been mirrored by severe balance of payment problems. These related conditions have alerted most of the country to our imperative need to develop alternative Earth-based energy resources such as coal, oil shale, nuclear fission source materials, and geothermal energy. Important research aspects of developing these as well as our remaining petroleum resources include

- Research in petroleum geology, including resource occurrence, origin, migration, entrapment detection, and assessment of oil and gas;
- New developments in secondary oil recovery;
- Determination of physical and chemical characteristics of petroleum reservoirs;
- Development of uranium/thorium exploration techniques;
- Understanding the nature, origin, and occurrence of geothermal areas;
- Marine geotechnical studies related to drilling and platform stability;
- Outer continental shelf oil and gas resource appraisal; and

- Drilling on the continental margin to assess the potential for oil and gas resources of the continental slopes and rises.

Increased Food and Fiber Production

Large areas of the United States are blessed with a climate and geography suited to food production. In the long run our agriculture will probably provide the sustained export commodities by which to balance our imports. Optimum agricultural productivity in most areas requires supplemental water and mineral fertilizers. Thus a trend toward greater agricultural productivity will result in greater use of water and fertilizer. Research areas, some of which are already planned over the next five years, needed to help sustain long-term optimum agricultural productivity include

- The impacts on stream-flow of the development of groundwater in humid areas of the country;
- The effect of irrigation on net water supplies and water quality in the humid eastern areas of the United States;
- Accurate measurement of water use by agriculture in the United States;
- The impacts of geologic processes on agricultural lands and the development of techniques to mitigate the effects;
- The impact of abandoning irrigated agriculture in water-short areas of the west;
- Identification of potential new sources of mineral soil supplements and the availability and concentration of fertilizers; and
- Determination of the regional distribution of trace elements in soils of agricultural lands.

Management of Wastes

One of this nation's most serious problems requiring earth science input for solution is the management of wastes, both toxic and nontoxic. Wastes out of sight or smell, in a stream, underground, in the atmosphere, or in an ocean haven't disappeared. They are in our environment and may function at one end of the scale, as a valuable amendment to some environments, or at the other end, as a toxic pollutant. The following areas of research will provide guidance to the nation as it struggles to cope with a growing waste problem

- Determination of the sources of nonpoint wastes in urban, suburban, and rural environments, and the impacts of nonpoint sources of wastes on surface and groundwater;
- Movement of chemical contaminants in streams—in solution and with sediments;
- Management of the wastes from the nuclear power cycle, including a comprehensive understanding of potential underground and marine repositories, and their geologic integrity;
- Movement of radionuclides in the hydrogeologic environment, and waste–media–water geochemical interactions;
- The fate of wastes placed in our river systems;
- The impact of wastes on the ecosystems of estuaries; and
- Movement of wastes by groundwater, both in unsaturated and saturated porous media.

Availability of World and Domestic Hard Minerals

As population increases with its attendant demand for mineral resources, it will become increasingly important to know what our mineral resource inventory is and what environmental restraints may be imposed on development. The economic well-being of our nation depends on our mineral resource base. Continued provision of a data base for the formulation of public policy to alleviate or deal with the mineral shortages of the future will result from research in

- Development of concepts and techniques to improve our capability for identifying and evaluating mineral deposits on both local and regional scales;
- Definition of domestic and worldwide mineral potential;
- The technology of extant exploration techniques;
- The development of new exploration techniques in geology, geophysics, and geochemistry;
- Marine mineral surveys, including manganese, iron, nickel, and cobalt;
- Mineral deposits modeling—understanding the physical chemistry of the formation of mineral deposits and their geologic settings; and
- Mineral information systems development.

Prediction and Mitigation of Geologic Hazards

The Survey has been and will continue to be a primary focal point for activities, in which the application of geoscience can be used to minimize loss of life and property caused by earthquakes, volcanoes, landslides, subsidence, rock deformation, swelling clay-shales, and other natural hazards. Among the many research topics in this area are

- Delineation and evaluation of geologic hazards;
- Development and testing of technologies for the prediction of earthquakes and volcanic eruptions, including identification of precursor phenomena;
- Characterization of the magnitude and nature of earthquake hazards on a region-by-region basis;
- Determination—in advance of reservoir impoundment—of the potential for reservoir-induced earthquakes;
- Prediction of volcanic eruptions, landslides, subsidence, and other natural hazards, thus providing a basis for mitigation;
- Effects and mitigation of volcano hazards in Hawaii, the Pacific Northwest, and Alaska;
- Safe siting of nuclear power plants; and
- Assessment of geologic hazards to surface mining and mine reclamation.

Development of Environmentally Sound Extraction Policies for Energy and Minerals

As energy and mineral needs increase and new sources are found, new environmentally sound exploitation guidelines must be developed. The Survey has a longstanding expertise in providing the data for the development of such guidelines, and increased research activities will be needed. Among them are

- Research on the consideration of environmental problems and constraints that accompany development of the energy of geothermal areas and extensive extraction of coal, oil, gas, uranium/thorium, and hard minerals, both on shore and outward to the continental margins;
- Research to develop new geophysical techniques to characterize the subsurface environment without violating ground integrity by drilling;
- Assessment and prediction of the impact of surface mining on water resources;
- Assessment of the long-term effects of burning fossil fuels on the global climate, including impacts on hydrology; and
- Marine geoenvironmental studies.

Interactions

Each of the six major research issues discussed above affects or places demands on our mineral and water resources. It is obvious that competition for mineral and water resources is increasing and will continue to increase as long as the population grows. Increasingly, reliable information on mineral and water resources and on earth processes is needed on a national scale by those making trade-off decisions for the allocation of finite resources. Among the research topics emerging from these issues are

- Identification and definition of a useful national assessment of mineral and water resources—including the data required for such an assessment;
- Determination of the present quantities, spatial distribution, and purposes of water and mineral use and waste products;
- Determination of the sustained volume and quality of groundwater available for use nationwide;
- Determination on a national scale of the resource savings assignable to implementation of natural hazards mitigation;
- Determination of past climatic conditions as an aid in the identification and projection of the effects of current climatic trends;
- Improvement of the knowledge of the geologic framework of the nation to provide the "baselines" for future activities; and
- Improvement of techniques of technology transfer to developing nations to help them define their resource potential, assure adequate supplies, and supplement our mineral needs in the future.

APPLICATION OF EXISTING TECHNOLOGIES

Technologies whose implementation or application will aid the solution of issues previously listed include: systems for encoding, manipulating, and analyzing spatial (geographic) data in digital form; a digital cartographic data base; remote sensing from space of the Earth's mineral, water, and vegetative resources; and an airborne profiling of terrain system.

Systems for Encoding, Manipulating, and Analyzing Spatial (Geographic) Data in Digital Form

The USGS has been increasingly concerned with the problem posed by accelerating scientific and technological developments in the techniques of gathering and handling information in geology, geography, topography, water resources, and related disciplines.

Prior to 1960, the only form of storage for spatial data was graphic. The manual application of statistical analysis and map interpretation assisted by drafting and photographic processes were the only spatial data handling techniques utilized. Since that time, rapid developments in computer technology have opened the door to nongraphic, digital storage, handling, and manipulation of spatial data. Activities in the Survey relating to the design and implementation of sets of spatial data in digital form have become major programs that are national in scope and involve the development of a new science of spatial data analysis. A measure of the growth of the importance of spatial data is suggested by the fact that, as of March 1976, there were more than 50 spatial data handling activities in the Survey, with about 366 million bits of spatially referenced data stored in operational systems. Current projections indicate that the stored data may exceed 1,365 billion bits by 1981.

The accelerating pace of requests for the transfer of spatial data system technology demonstrates that such technology and data are increasingly in demand. Data are needed in support of research projects, for users such as water managers and land-use planners, and for regulatory purposes in environmental impact analyses and regional energy and resource planning. Those requirements have already created an urgent need for multifile spatial data base management systems, improved access via remote terminals, and linkages to systems maintained by other organizations in the Federal Government and regional, state and local organizations.

Digital Cartographic Data Base

A national digital cartographic data base is being developed by the USGS to serve all users in the Federal, state, and private sectors. Current information digitized from published maps meeting National Map Accuracy Standards forms part of the data base. In addition, a digital cartographic system is being designed to produce both computer files and map graphics directly from mapping-quality aerial photographs. The basic objectives of the program through 1985 are to develop and keep current a data base of digital cartographic data for coastal areas, coal lands, public lands, and many urban and forest areas; and provide timely digitizing, editing, analysis, and display services to Federal and state agencies. Uniform standards for data formats and codes will be developed to provide the interchange and use of common data.

Digital cartographic data are used by state and Federal planners and managers responsible for managing natural resources and public lands to solve complex problems of computer-based analysis. Digital cartographic data can also be used in automation of traditional mapping processes, including revision. The automated digital system will be more efficient and less labor intensive than methods currently used to produce maps, and the digital data will not interfere with other geographic data systems.

Remote Sensing from Space of the Earth's Mineral, Water, and Vegetative Resources

The purpose of viewing the Earth from space is to take advantage of those unique characteristics afforded by space technology that enable us to extend our view and learn more about the world around us; to discover, develop, and manage its many resources; to improve our environment; and to ensure the future by improving the ability of mankind to obtain food, shelter, and better health.

Satellites carry various types of sensor systems designed to measure reflected and emitted radiation from the Earth's surface. All surfaces and objects on the surface of the Earth reflect, absorb, and emit radiation in varying degrees, and the fact that many are unique and separable from others enables us to refer to their measured expression as a "signature." These signatures can be measured by taking advantage of the various spectral bands which range from visible, through near-infrared and thermal infrared, to microwave frequencies. In most cases, the signatures of vegetation, soil, water, and rock change with time due to season and weather. Agricultural and range crops change continually and are synchronized with a crop calendar. The synoptic and repetitive characteristics of space observation enable man to extend his observations from single fields or groups of closely spaced fields to counties, states, and groups of states, and even make comparative studies of continents.

Meteorological satellites (Nimbus, Tiros, NOAA, and Geos) have clearly demonstrated their value in mapping weather patterns, sea surface temperature, and major ocean currents. They are used to predict weather and forecast the movement of hurricanes, tornadoes, and snowstorms.

Landsat, launched first in 1972, has demonstrated that multispectral, synoptic, and repetitive views of the Earth can be beneficial in many ways. Geologists have found that they can extend their knowledge gained from small field areas to broad regions. Agricultural specialists can monitor crop growth and estimate productivity by multistage sampling techniques using field, aircraft, and satellite observations. Range managers can estimate forage availability; foresters can rapidly assess the rates of change due to forest exploitation, fires, and insect infestation; hydrologists can measure areas of surface water and the extent and comparative effectiveness of sprinkler and ditch irrigation systems; and land use planners can obtain current land use maps.

As space technology moves from experimental to operational systems, the acceptance and use of such data will grow. Because they are computer compatible they will form the basis of computerized resource inventory and management information systems of Federal, state, and local government agencies. Private industries, especially the major oil companies and future commodities firms, have already adopted such systems and consider them nearly operational.

We expect that continued research will enable us to further refine spectral measurements from space or combine them in new ways to improve our ability to discriminate and map surface materials. The computer instrumentation used to process these data will improve, become cheaper to purchase and utilize, and, therefore, become more accessible to the average user. With this, the user community will continue to grow as will the application of data to serve their needs.

Airborne Profiling of Terrain System

USGS scientists are guiding the development of an Airborne Profiling of Terrain System (APTS) by the Charles Stark Draper Laboratory of Cambridge, Massachusetts. The system is in the component fabrication stage, with laboratory testing planned during 1980, followed by flight testing in 1982.

When completed, the system will permit rapid collection of real-time three-coordinate position information. Using low-flying aircraft, with a laser profiler system linked to geodetic control points, the APTS is expected to provide accurate surveys of the terrain more efficiently than current technology. Positional accuracies are to be ± 2 feet (about 60 cm) horizontally and ± 0.5 feet (about 15 cm) vertically.

The APTS will provide data for topographic mapping, determining stream-valley geometry for flood plain studies, determining surface subsidence, monitoring strip mine operations, detailed coastal mapping, and studies of geology.

Department of State

The mission of the Department of State (DOS) is to coordinate our foreign policy and promote the long-range security and well-being of the United States. Science and technology play a significant and increasing role in the formulation and conduct of U.S. foreign policy, and in our relations with other countries in general. Science and technology provide tools with which to progress toward solving global problems; even problems predominantly of a political nature are often based on subtle scientific or technical factors, and, of course, certain applications of science and technology in turn create problems for our foreign policy. On the other hand, science and technology both provide bases for international cooperation and opportunities for improving relations with and understanding of other countries.

This report focuses on areas of science and technology related to the significant concerns of the Department of State and on opportunities and constraints for the use of science and technology in achieving foreign policy goals.

INTERNATIONAL PEACE AND SECURITY

Many of the factors capable of disrupting or preventing stability on a regional or global basis are grounded in technology.

Weapons

A credible military capability requires a rather substantial degree of technical training in equipment operation, routine maintenance, and repair. The development of new weapons or even the new acquisition of traditional weapons can alter the military balances among countries and prompt destabilizing responses. This can occur not only among super powers, but among the poorest of the developing nations. We anticipate a continued interest on the part of many countries in acquiring conventional weapons. In the next five years an increased interest in such acquisitions could be demonstrated by lower tier developing countries. It is the policy of this Administration, of course, to try to reduce worldwide arms sales through a multilateral effort.

Nuclear Technology

The spread of nuclear technology continues, and the diffusion of technical information potentially applicable to weapons development cannot be halted or reversed. As international interest and commerce in nuclear power continue to grow, and sensitive technologies such as reprocessing and enrichment become more widely available, it will become increasingly important to strengthen the international machinery for safeguarding nuclear materials, equipment, and technology. We will need to develop a more proliferation-resistant environment, both technical and institutional. Prudence dictates that attention also be devoted to the possibility of contending with subgovernmental groups that are likely to be immune from traditional political constraints; in this context, improved technical controls may be particularly important.

World Development

Focused rather than random efforts in science and technology are needed to further develop and strengthen the developmental process in the third world. It will require increasing cooperation among developed nations, financially as well as scientifically, to pursue such potentially big, expensive, internationally important research as that having to do with energy resources, ocean margin drilling, climate, etc. World development can be characterized both in relative terms—i.e., by widening or narrowing gaps among the countries—and in absolute terms. For many matters of relevance to the world economy, the description in relative terms is most pertinent. However, in the quest for international peace and security both relative and absolute terms are important. This is apparent in military capability, but it is also the case in view of increasing demands by the developing nations for energy, raw materials, food, and prestige. These factors will certainly call for increased attention in the near future.

China

The decision of China to embark on a program of technological modernization based in large part on cooperation with the industrialized democracies is changing the nature of our relations with that country. More broadly, the role that China assumes in the world will be very much a function of its development in military and other technologies. Her success, or lack of it, may have a profound impact on the global political balance.

Treaties and International Law

The United States continues to believe that the development of a regime of treaties, agreements, and customary international law, particularly through the United Nations and other intergovernmental organizations, is the best way to foster peace and security among the world community. Such a cooperative framework is essential to meet the challenges of the future and the problems of the present.

Two major areas of developing law in which technical factors are particularly relevant are the sea and space. While great strides were taken at the Third United Nations Conference on the Law of the Sea, in securing agreement on new jurisdictions and on regulating novel and sometimes competing uses of the seas, there is a price to pay. The technically feasible mining of the deep sea bed, which has not yet been susceptible to organization under a politically acceptable international regime, remains a source of conflict between nations. The development of the sciences of the sea and of the air, which is becoming increasingly dependent on international cooperation, is also increasingly constrained by national and international regulation. Efforts in outer space, however, have resulted in the negotiation within U.N. forums of four multilateral instruments to which many countries, including the United States, subscribe. However, additional and more contentious problems remain to be resolved. In the next five years we can anticipate that technological developments especially relating to resources, energy, industrial applications, and scientific research will pose new problems for the international legal regime.

NATURAL RESOURCES AND ENVIRONMENT

Society is (we hope) entering a phase of transition from the mindless consumption of natural resources and the production of harmful wastes to a relationship of symbiosis with the planet and its ecosystems. Among our key concerns as a civilization are that we not deplete essential resources while we grow economically and socially, that we maintain our environment in a healthful state, and that we not diminish the genetic richness of life in its manifold forms.

Energy Resources

The exploitation of energy resources is unavoidable for any level of economic activity, including provision of basic human needs. The demand for energy will rise steeply as world population increases and economic activity continues to grow. The problem exists, growing ever more serious, that conventional hydrocarbon sources of energy (oil, natural gas, and wood) are being depleted at an alarming—and accelerating—rate; aside from the implications for energy supply, this depletion also threatens the availability of these resources for other uses, such as specialized materials, pharmaceuticals, and fertilizers.

Alternative Energy Sources

Alternative energy sources are under development, but most technical estimates of their availability indicate that it may be two decades before renewable or essentially inexhaustible sources of energy will contribute significantly to energy supply. In the near future, it may well be that coal will undergo an expansion of its role and that fission reactors (possibly including some version of the breeder) will become more widespread under controls adequate to ensure their safe and secure operation. More intensive exploitation of such sources of energy will raise important issues of environmental quality, human safety, and (in the case of breeder reactors) nonproliferation. We have come to realize that many energy exploitation activities have international and even global effects, and we will have to develop more effective mechanisms to influence other countries to adopt modes of action less detrimental to the environment. The developing countries will present a special problem in this regard as they move into the early stages of industrialization with emphasis on near-term cost minimization at the expense of long-term environmental interests.

Efforts to develop additional energy resources may create foreign policy issues in the near future, even though full development of these resources may be decades away. For example, consideration of a solar power satellite system inescapably involves questions of impact on global environment, prospects of multinational sharing of the benefits, allocation of frequencies for transmitting energy to earth, and use of geostationary orbit positions. More generally, international or multilateral energy research and development, especially in large demonstration-scale projects, offers the prospect of reducing costs for each participating country, accelerating needed efforts, widening the base for innovative thinking, and avoiding wasteful duplication. Such multilateral efforts can also serve the broader purpose of enhancing overall international cooperation. On the other hand, effective project management can be seriously hampered by a proliferation of participants—especially foreign participants—in large, complex, and costly efforts which are inherently uncertain of success due to the innovative nature of the technology being implemented. There will often be a trade-off between technical and diplomatic goals.

Energy technology can seriously influence bilateral relationships and can serve as the basis for much broader economic ties.

Forest Reserves

Until the beginning of this century, the predominant energy resource was wood. As population densities grew, and energy demands for heating and cooking as well as for construction materials increased, forests in a number of areas of the world began to be depleted. As a result, large areas were eroded by rain and wind and rendered less fertile. The use of coal and then natural gas and oil came in time to save at least a few areas of extensive timberland. With recent shortages of petrochemical supplies and increasing demands for energy even in the least developed countries has come a renewed attack on the forests of the world. This will have grave consequences for the continued development of poor countries because they will too quickly deplete their forest reserves and will subject large areas of once fertile land to erosion and loss of agricultural utility.

The loss of forest lands is becoming so extensive that its effects are certain to be regional and even global in scope. Forests may play an essential role in the chemistry of the Earth-atmosphere system, and their further depletion could lead to significant long-term effects on the composition of the atmosphere. Moreover, the burning of wood, without special precautions, introduces into the atmosphere a variety of pollutants including particulates. Concerted action is needed to solve this global problem, and the next decade or so may prove to be crucial.

Deserts

The Earth's surface changes over time with changes in climate and man's activities. The dynamics of deserts are not well understood, but recent indications of their growth are worrisome. The reclamation of desert land is costly, but there have been some successful experiments, particularly in the Middle East. Moreover, it has been shown that certain range management techniques can increase the productivity of desert land. A number of countries have large desert areas, and the enhanced utility of such land could be important in their economic development. One key factor is the supply of fresh water. Although deserts are quite dry at the surface, there are underground water reserves in a number of places. The technology necessary to locate and develop these sources of water would promote the utility of deserts. However, new wells have often made the problem worse by opening unsuitable lands or resulting in expanded grazing herds.

Water Resources

The Earth's water cycle is global in scope and involves the oceans, atmosphere, and water on and below the surface of land masses. The entire cycle is fueled by sunlight. While it is difficult to detect long-term trends in the global availability of fresh water, there are important variations in rainfall of local or regional scope both from year to year and for longer times. The variation of precipitation during the course of the year is also substantial, but to a large extent we have been able to adapt our society to this level of variation. Over the next five years we might expect to see improvements in the technologies of locating and exploiting underground water systems and desalinating sea water.

The improved availability of fresh water is of key importance to a number of countries, generally developing countries, and its achievement will have important effects on their growth. Schemes for diverting the flow of surface waters could lead to improvements in fresh water availability for some countries, but could also deprive neighboring countries of some water. From time to time, unconventional ideas, such as towing icebergs from polar regions to the shores of a needy country, are suggested. These may at some point prove worthwhile. Over the next five years, we expect to see increased attention paid to the problems of fresh water availability and techniques for the discovery, exploration, and management of water resources.

Weather and Climate

The United States has engaged extensively in the development of weather sciences in recent decades. Satellite sensors provide synoptic views of large-scale weather systems while a network of Earth-based sensors provides information on a more detailed scale. Considerable advances have been made, particularly in long-range and medium-range forecasting; short-term forecasting, while important, is still frustratingly unreliable in the view of the "man on the street." In the next five years we should see continuing efforts to improve our understanding of weather phenomena and more attractive opportunities for gathering information on a global scale through the involvement of many countries.

Weather modification research was initiated some 50 years ago in this country. The potential benefits of modification make it of continuing interest even though it is not yet certain that it can be safely used in all its forms. Progress in this area is coming slowly, partly because the research is difficult, and the benefits and the risks have to be balanced. International guidance and controls are likely to be a matter of increasing concern during the next five years.

Similarly, interest in studies of climate, its long-term trends, and its implications for the economic health of countries and regions is becoming more intense. As the relationship between climate and other factors becomes better understood, interest in cooperative research efforts could focus foreign affairs activities on this area.

Living Resources in the Oceans

In addition to playing a central role in the dynamics of the world's weather and climate systems, the oceans of the world nurture a tremendous variety and quantity of living resources. These resources are finite in extent, and we are in serious danger of depleting certain species to the point of extinction. To preserve the richness of life's varieties and to avoid traumatic disruption of the delicate balances of life in the ocean, we will need to carefully manage our fishing, commercial, recreational, and other ocean-related activities to optimize their benefits for mankind and at the same time conserve the marine environment and its resources. Many nations are more heavily dependent on the oceans for food than we are. Some have adopted fishing practices that threaten certain species or that interfere with the take of other countries. The trend toward extended coastal nation jurisdiction over living resources in adjacent ocean areas may permit more effective management of fishery resources and thereby improve the ocean's productivity in the long run. We also anticipate that our increasing ability to understand the ocean and monitor changes in it will improve our ability to manage its living resources.

Other Ocean Resources

The oceans of the world hold great promise for mineral and hydrocarbon production as well as for alternative sources of energy. One of the most exciting prospects is the recovery of polymetallic nodules (more commonly known as "manganese nodules") from the floor of the deep sea bed, but it is not yet fully known what other resources might be exploitable.

The technology for deep-sea mining is in the final stages of development prior to a prototype mining operation, and it is highly likely that national and/or international rules will be in place in the very near future to provide a legal framework for actual operations. Exploitation of the ocean areas beyond national jurisdiction was a key issue at the Third United Nations Conference on the Law of the Sea, and it is not yet clear how this issue will be resolved. The United States and other countries, however, have continued to stress that such an activity on the high seas is a freedom that cannot be abridged without consent of all parties. Further developments in undersea technology and the commencement of actual mining operations on the high seas could make this issue an important foreign policy concern in the next five years. Oil production, especially in the industrial countries, is shifting offshore, and this requires complex and costly technology.

Mineral Resources

Global interdependence in supplies of many mineral resources has become a fact of life in this half of the century. Few countries are self-sufficient in resources considered essential for maintaining national security and an advanced economy. Demands for raw materials will surely continue to increase as more countries develop their economies and as those economies become more complex. But supplies are not unlimited, nor are our relations with certain suppliers wholly reliable. There is a limit also to our ability to stockpile effectively the quantities of key materials needed by the U.S. economy for more than a period of months or, in some cases, a few years. This makes us vulnerable to a number of possible actions that would interrupt or increase the cost of such supplies. We expect this problem to intensify in time, and over the next five years we may see significant impacts on our supplies of certain materials arising from the efforts of developing countries to gain political influence through control of their natural resources.

The natural free market response to this situation would probably be the stimulation of alternative suppliers and the development of substitute materials. To accomplish the first, new techniques of resource discovery, extraction, processing, and even transportation will find useful application. Satellite remote sensing surely will be increasingly used in locating and managing natural resources; efforts to draft principles governing remote sensing of the earth's natural resources and environment are underway in the United Nations. Extraction and processing require a certain skill level as well as capital investment; to promote the development of resources, we may face difficult demands for providing technology and capital equipment to countries not yet able to use them properly. Furthermore, this phase of the raw material supply process can be both environmentally damaging and sociologically stressing. Some developing countries evince little concern for the quality of their environment, but pollutants deposited in air and waterways will affect neighbors and even distant peoples to some extent. Additionally, the requirements for creating a labor force and moving it to the location of an activity can stress the family and cultural life of a significant segment of the population and create economic disparities that may wreak havoc on a country's economy. Within the next few decades the occurrence of these effects will undoubtedly increase.

The development of substitute materials to replace materials whose supply is not reliable is not expected to occur by itself in the private sector until the supplies are actually interrupted. Time will then be a key factor. Government action to prepare for this kind of situation can be initiated at an early date.

Pollution

The international aspects of pollution of the environment are now seen as very important. The physical and chemical processes involving land masses, oceans, and the atmosphere recognize no political boundary. We are finding, moreover, that while certain effects of pollutants are local in scope there are many subtle but important effects that are global in range. Controversies exist over the precise characterization of phenomena like depletion of atmospheric ozone, changes in atmospheric carbon dioxide, acid rain, trace metal effects, and so on, but little doubt exists that they are important phenomena that are influenced by mankind's activities. Even a phenomenon like eutrophication, which we are accustomed to think of as occurring in small bodies of water, can be of significance in international relations in regions such as Europe. Environmental insults are expected consequences of virtually any large-scale activity of a society. Technical advances in assessing and ameliorating environmental impacts continue in the United States and other countries; however, some developing countries do not welcome their application because there are frequently significant costs or constraints implied.

Environmental monitoring, aimed merely at improving our knowledge and understanding of the dynamics of pollutants and their sources and sinks, also constitutes a matter of some international controversy. This controversy can be expected to increase as monitoring becomes more necessary and the technology more capable. Efforts will be underway to encourage other countries to assess in advance the impact of their activities on neighboring countries or on the global commons.

Waste Management

The United States is the major consumer of resources in the world, a fact that has not gone unnoticed in recent times. Other countries urge us to be more conservationist in our use of resources and may see low-waste technology as a necessary development. This technology is most relevant to developed countries, and we expect to see increased attention paid to cooperative efforts in this field in the near future. At the present state of technology, there are unavoidable wastes that are not recyclable (for technical or economic reasons). Such wastes must be disposed of in an environmentally sound way that does not threaten human life or health. The disposal of radioactive wastes is an example that is both controversial and challenging. Suggested disposal sites include geologic depositories on land or in the sea bed and outer space. As waste materials continue to build up, the pressures to solve this problem will increase.

Natural Hazards

Among the global problems that have been with us for eons is the occurrence of natural disasters—floods, earthquakes, fires, droughts, volcanic eruptions, storms, tidal waves (tsunamis), etc. They have always taken a toll in human life. As the Earth's population increases and our societies become more complex, the effects of natural disasters are increasingly severe. Science and technology hold promise for providing eventually a reasonable capability to anticipate their occurrence, to plan the location and nature of settlements in order to minimize vulnerability to them, and to alleviate their consequences once they occur. Interest in natural disasters is very strong among developing countries. The technologies of direct relevance include remote sensing, communications, air transport, building techniques, and water diversion, among others.

Food Supplies

Basic demand for foods continues to increase, and advances in nutritional science have altered consumer demands for certain foods. Among the factors pertinent to the supply of food grown on land are soil fertility, weather conditions, quantity of arable land, harvesting methods and level of effort, food distribution, and food processing. Plant diseases and spoilage are major factors in reducing food supplies. A continuing effort in this area needs to be bolstered in the future if we are to make any headway against the increasing severity of the problem. Not only are technical advances needed, but innovative organizational and administrative approaches, particularly in delivering fertilizers and water to the farmers and distributing and storing their products, are needed.

The living resources of the world's oceans and fresh water bodies offer the opportunity to alleviate the demand for foods high in protein. The trend toward extended coastal nation jurisdiction over living resources in adjacent ocean areas may permit more effective management of fishery resources. Increasing attention is also being given to ocean species that are presently underutilized. Finally, significant advances continue to be made in both freshwater and saltwater aquaculture, whereby fish and vegetation can be grown and harvested in a highly productive manner in a controlled environment.

Basic Research

Basic research in sciences ranging from archeology to zoology should have little negative impact on our foreign relations and should continue to provide opportunities for international cooperation. Such research is the source for future technology. The development of scientific information that pertains to the natural resources of countries or that gives rise to proprietary rights will probably continue to prompt assertions of sovereign interest by some countries.

INTERNATIONAL COMMERCE AND THE ECONOMY

Economic Development

The large disparity in levels of economic development among countries shows every sign of continuing to increase. We have seen that the stability of the world economy is jeopardized by the stresses generated through this disparity in level, and we have become aware that in this age of interdependence we are disadvantaged both politically and economically by increasing the disparity. The solution offered by developing countries is to rapidly erode and dissipate the positions of advantage held by advanced countries. On the other hand, an approach that aims at long-term development of all countries, while not attractive to developing nations in general, may be the most effective and palatable approach from our standpoint.

Technology is crucially important to the development process; but for a country to make progress in its ability to accommodate, use, and eventually develop technology, it must have an appropriate infrastructure. The lack of suitable infrastructure in most of the developing countries has thwarted a number of well-intentioned efforts to help them through the transfer of technology from the United States and other advanced countries. As focus on the question of technical infrastructure is improved, greater demands for technologies of communications, transportation, and construction, and for education in science and engineering, will become apparent. But this cannot be simply a matter of transferring them in the form they take in the United States. Technologies must be adapted to the purposes they are to serve in the host country, and they must be appropriate not only to the technical job to be done but also to the societal and cultural setting. In the next five years we can expect a further increase in the appetites of developing countries for an immediate diet of high technology. We must begin to work more effectively with these countries in the longer term, but more sound, approach to developing technical infrastructure.

Telecommunications

Rapid technical advances in telecommunications are being realized. Large-scale integration of electronic components is permitting economies in a variety of equipment and promises to spawn the development of new capabilities. A number of other technical advances are underway in this area. They imply that the ease with which we communicate both domestically and internationally will grow while the costs will diminish. With such developments we can anticipate that countries will become more active in communicating, and this can lead to overcrowding of the radio-frequency spectrum and of the satellite orbit space in geostationary orbit. In addition to increases in communications between fixed points, we anticipate similar increases in the use of broadcasting, particularly of television.

Because television is so effective a medium for the communication of ideas, many countries fear uncontrolled broadcasting from stations outside their sovereign jurisdictions. The broadcast of television from satellites to home receivers in other countries is of concern to many countries, and efforts are underway in the United Nations to deal with this problem. In the next five years or so we expect to see several countries progress toward establishing domestic satellite television broadcasting systems, and at least one regional system may be in place. This will raise the level of interest or concern among a number of countries heretofore rather unconcerned about the issue.

Free Flow of Information

The principle of free flow of information and ideas regardless of national frontiers is fundamental to American society. Recent international reaffirmations of this principle by UNESCO have not settled the issue, and some countries continue to urge regimes under which controls over the flow or even the content of transborder communications would be acceptable. A number of countries have adopted regulatory schemes for dealing with the flow of data that might affect the privacy or other rights of their citizens. International efforts to deal with the protection of rights and at the same time the preservation of freedom of information flow will be increasingly important in the next five years.

The technology to deal with information gathering, storage, processing, and dissemination is making rapid strides. Most countries seem to recognize that this is an area in which the gap between developed and developing countries is growing. Increasing demands for "informatics" technologies and for adequate policies are being made, and responses will be urged more fervently in the near future.

Commerce

Commerce among the developed countries and with the Communist countries of Asia and Eastern Europe is an important factor in the economic health of the United States. While agricultural products will surely retain a predominant position among our exports, trade in technology products and transfer of technology will probably increase. This will have the effect of increasing the technical competence of the importing countries, with concomitant military implications in some cases, and economic competition. The next years will raise issues of increasing complexity in the promotion of trade under circumstances that preserve national security interests and our economic health.

HUMAN RESOURCES AND WELFARE

Population

Population increases underlie many of the problems we have identified. Assistance in family planning and improvement of living standards are expected to help in dealing with the problems. The pressures in the next five years will urge creative and incisive approaches. Distribution of population plays a key role in a number of areas including urban blight, rural development, labor force, standard of living, education, and a constellation of other social and cultural problems.

Medicine and Health

Advances in medicine and health care are expected to include improvements in diagnostics, treatment, disease control, rehabilitation, and the delivery of health care. These advances can play an important part in raising living standards around the world. Health care is an area in which developed countries can bring benefits of science and technology to developing countries with minimal (but significant in some cases) impact on the cultural and social integrity of the recipient. Consequently, this area is a prime candidate for expansion of international activities in the years ahead.

INTERNATIONAL MEETINGS

During the next five years there will be a large number of international meetings dealing with matters related to science and technology. These will be intergovernmental meetings and meetings involving a broad range of organizations as well as individuals. These meetings are sponsored by the United Nations (and its member organizations), such other intergovernmental bodies as the Organization of American States and the Organization for Economic Cooperation and Development, as well as individual countries, professional societies, etc.

Illustrations of the most significant of such meetings include the 1979 World Administrative Radio Conference (WARC) to revise the international radio regulations, including the table of frequency allocations, technical sharing criteria, coordination procedures, etc.; the 1979 U.N. Conference on Science and Technology for Development; the 1981 U.N. Conference on New and Renewable Sources of Energy; the U.N. Conference on Exploration and Peaceful Uses of Outer Space proposed for 1982; and the 1983 Regional Administrative Radio Conference (RARC) to plan the use of broadcasting satellites at 12 GHz in the Western Hemisphere.

INTERNATIONAL COOPERATION

There is an extensive program of bilateral and multilateral cooperation in science and technology involving a number of countries and a number of U.S. Government agencies; regional institutes and special organizations, as well as agencies of multilateral intergovernmental organizations, are involved. One example should suffice to illustrate these opportunities:

For almost a decade, the Deep Sea Drilling Project (DSDP) has been an extraordinary scientific undertaking. Its major objective has been the exploration of the Earth's crust beneath the oceans. This has been accomplished through the use of the drill ship *Glomar Challenger*, which obtains cores of sediment and rock from the deep ocean floor. The scientific results of DSDP have revolutionized the earth sciences and led to general acceptance of the plate tectonics model. This model has immeasurably enhanced our understanding of the Earth as a dynamic system possessing a continuously changing distribution of oceans, continents, and mountain ranges. As interpreted by means of the plate tectonics model, the results of DSDP have provided evidence crucial to the reconstruction of the Earth's geology, climate, ocean circulation, and evolution of living organisms during the past 200 million years. At the same time new insights into the geologic processes associated with the formation of mineral and hydrocarbon deposits have been provided. Scientists from approximately 30 foreign countries have participated in DSDP, and five countries have entered into formal agreements with the National Science Foundation concerning their participation.

The oceanographic institutions that conduct DSDP have proposed a major new program for the 1980s which would be a logical evolution from DSDP. Whereas DSDP was concerned almost exclusively with drilling in the deep ocean basins, the proposed program would be concerned primarily with a detailed investigation of the outer continental margins of world oceans. Although these margins are of central importance to a better understanding of the plate tectonics model, their geology and geophysics are relatively unknown. At the same time, the outer continental margins are probably the largest unexplored areas in which oil and gas may be found. The margin studies will require a more advanced drill ship that can drill in an environmentally safe manner in areas where hydrocarbons may be encountered. The proposed studies provide an opportunity for the continuation of the excellent international cooperation that has occurred under DSDP. On the other hand, the definition of the seaward limit of coastal state jurisdiction over the continental margin and the nature of coastal state jurisdiction over marine scientific research are two unresolved issues under consideration by the United Nations Conference on the Law of the Sea. As long as they remain unresolved, they may impose constraints on any future drilling program on the outer continental margins of the world.

In addition to multilateral cooperation on specific projects, the United States participates in bilateral cooperative agreements with many countries covering large topic areas.

The work of a number of multilateral intergovernmental organizations includes significant opportunities for the discussion and the use of science and technology. In the North Atlantic Treaty Organization, the Organization for Economic Cooperation and Development, such U.N. bodies as the World Meteorological Organization and the Food and Agriculture Organization, the Organization of American States, and other organizations, these topics are of continuing interest and are discussed from their technical, economic, and political aspects.

Department of Transportation

The Department of Transportation (DOT) has organized and planned its research and development programs to help solve the current and emerging transportation problems identified through the political and socioeconomic research processes. While DOT concentrates on short-term research, the transportation infrastructure has a very long life. An example is the persistence of excess capacity in the railroad industry, which in many corridors dates back to the early 1900s. While research may be aimed at solving short-run problems, impacts on the infrastructure may be felt far into the future. Long-range trends in transportation demand and technology must be considered in terms of what effects they will or could have on the current infrastructure. Even though our research and development program might be called "short term," we are faced with a number of opportunities and constraints upon which new technological capabilities can have a profound impact. Although the policy of short-term R&D remains valid, it could be a constraint to innovation in transportation. To bring new ideas into transportation, we intend to reverse what has been a steady deterioration of long-range R&D funding by including a long-range component in our advanced programs.

The Department has conducted a number of studies that identify transportation trends and problems. These studies looked at the alternatives for the near and distant future and are used as a guide to the development of our R&D program. Perhaps the most significant problems facing our transportation system are the increasing energy shortages, especially liquid petroleum, and anticipated high growth rates in certain modes, particularly aviation. In addition, the freight transportation industry continues to suffer from financial problems that could impair its ability to meet future demand.

In modal terms, railroads and pipelines may be increasingly relied upon and must adjust their capacity to meet the new demands safely and efficiently. Urban transportation may see a shift in job and housing patterns, with shorter work trips, but often in corridors of increased congestion. Innovation is required to reduce the amount of energy consumed by commuting. On our waterways, continued growth of energy shipments could lead to serious safety problems as ports and waterways become more congested. Likewise, if aviation demand grows in the face of higher energy prices, congestion of air terminals will pose greater safety problems. Our research and development program strategy is structured to resolve the problems created by these trends. A number of new or emerging technologies provide opportunities to solve not only modal but the so-called cross-cutting or transmodal problems as well.

To provide an integrated approach to this outlook, we have divided our response by the following categories: marine, aviation, road and highway, rail, urban, pipelines, and transmodal.

MARINE TRANSPORTATION

In the next five years, the development of offshore resource platforms at distances greater than 25 miles from the coast will present the major marine problem.

The safeguarding of personnel operating those platforms; the safety of shipping, especially oil and liquefied gas tankers, through offshore waters; and the protection of the environment during this activity are problems of national significance. The development of offshore platforms will add to the complications of solving the following problems.

Collisions of Commercial Vessels

The number of collisions, rammings, and groundings of commercial vessels in U.S. waters must be reduced. This problem is pursued by developing a Marine Safety Information System, establishing crew qualifications, defining vessel and vessel equipment safety requirements, and increasing damage resistance of vessel hulls. Vessel navigation, control, maneuvering, and communication systems will be enhanced by improving the effectiveness of short-range aids to navigation, developing a system for the management of vessel traffic, and the application of loran-C to making harbors and harbor entrances safer for mariner and nonmariner alike.

Oil Spills

We must detect, identify, contain, and dispose of oil and hazardous substances spilled in U.S. waters. A wide range of scientific and technological options to solve these problems are being investigated. These include development of sorbent applications for oil recovery, dispersant application systems, and oil slick burn-off systems. The improvement of techniques and equipment necessary for the prevention or reduction of oil discharges from stricken vessels is being sought. The special requirements imposed by the arctic environment are also being investigated.

Enforcement of Laws and Treaties Affecting U.S. Waters

Many technological capabilities can be utilized to improve marine surveillance and communication capabilities. These include the use of satellites and new vehicles such as Small Water Area Twin Hull (SWATH), Surface Effect Ship (SES), or Lighter than Air (LTA), and advanced communication techniques (digital and burst).

Dangers From Carrying Hazardous Substances

There are ways to minimize dangers to personnel and facilities from hazardous substances carried in maritime commerce. The implementation of the Ports/Waterways and Tanker Vessel Safety Act is a major problem to be addressed. The investigation of fires and explosions aboard tankers is being conducted to identify technologies that can be used to reduce these dangers. Improvements in shore-facility fire-fighting capabilities are included in the current program. Other major efforts are the development of criteria for cargo containment systems and standards for the safety of vessel hull structure. Future efforts will be expanded to include the minimization of dangers from offshore nuclear platforms and from bulk chemical carriers.

AVIATION

Historically, the Air Traffic Control and Navigation System has been improved and expanded on an evolutionary basis. The significant investments in avionics by the users and the practical necessity to provide continuous services require the continuation of this process.

Upgrading the Current Air Traffic Control System

The projected growth of aviation was evaluated in the late 1960s and a plan was developed to provide capabilities to meet the demands for services through the mid-1980s. Many of these capabilities, particularly enhancements to the enroute and terminal automation systems, have been implemented. Other major improvements are nearing the end of the development cycle and will be placed in the operational environment in the next several years.

The rapid technological progress made in integrated circuits, minicomputers, distributed minicomputers, high-speed microprocessors, and solid-state radio frequency (RF) technologies have provided significant opportunities to meet the objectives of these major improvement programs. The impact can perhaps be best appreciated in light of the increased levels of sophistication required in aircraft systems to carry out curved approaches utilizing the microwave landing system, data processing and display of collision avoidance information, data link-transmission, and area navigation. These technologies have the capabilities to provide an even higher level of service to aviation in a cost-effective manner. Additionally, they can be used to upgrade, in an economically acceptable way, the equipment of the smaller aircraft that compose the bulk of the general aviation segment of the civil fleet.

The future, the mid-1980s into the early 1990s, will present DOT's research, engineering, and development program with the formidable challenge of achieving one of the largest systems engineering efforts ever to be accomplished in the United States. The automated systems currently employed in providing services in the enroute and larger terminal areas are based upon the technology of the mid-1960s. The technology of the communications systems serving the entire air traffic control process dates from the 1950s through the 1960s. The automation and communications systems lack the capacity and speeds to meet the demands for air traffic in the post-1985 era. The Air Traffic Control Upgrading program will include design, fabrication, and testing of virtually all hardware and software components and the in-service changeover of 20 Air Traffic Control Centers, 63 of the busiest terminal areas, and the replacement of the entire switching and network facilities for voice and data communications. Implementing these changes will be difficult since operations cannot be interrupted.

ROAD AND HIGHWAY TRANSPORTATION

Reducing energy consumption, improving safety and the environment, reducing construction and maintenance costs, and improving the ability of the existing system to meet demands are the major highway problems. Potential solutions are described below.

Use of Waste and New Materials Sources in Highway Construction and Repair

There have been, for many years, chronic shortages of materials suitable for aggregate use in both asphaltic concrete and portland cement concrete in some areas of the country. Research activities initiated in the early 1970s to develop alternative sources of highway construction material, including domestic, industrial, and mineral waste, have now begun to produce useful outputs. Two processes have been developed to permit the utilization of municipal incinerator residues. Both of these products hold promise for reducing the cost of road building in major metropolitan areas with large municipal incinerators.

Techniques are being developed for recycling asphaltic concrete in the reconstruction of deteriorated pavements. Efforts have been made to reduce the demand for petroleum products used in highway construction by shifting to asphalt emulsions rather than petroleum-based solvents, and using sulphur as an extender or replacement for asphalt. The development of techniques to convert sulphur itself into a material with adequate engineering properties to serve as a binder has perhaps the greatest potential for reducing the demands of highway construction on petroleum supplies. Basically, the method consists of a new technology of polymerization that permits the long chainlike sulphur molecules in molten sulphur to be cross-linked so that conversion to the crystalline state upon cooling is inhibited.

Measures to Accommodate Heavy Vehicles on the Highway System

Much of our highway system is rapidly wearing out. Congress has recognized this problem by the institution of the bridge replacement and pavement resurfacing, restoration, and rehabilitation programs. To protect the large investments in these programs, the determination of the appropriate loading levels for highways is urgent.

Increasing truck efficiency by allowing greater payloads raises three major issues: safety, protection of the existing highway system against excessive damage, and policies for the future design loadings for highway pavements and bridges to optimize the combined cost of cargo movement and highway investment. The latter question also gives rise to major intermodal policy issues due to a possible change in the competitive position of the trucking industry and revision of tax arrangements to ensure equitable distribution of highway costs.

The questions of economics and optimization of cost and energy considerations require refinement of existing damage functions for highway pavements and bridges in relation to axle and gross vehicle weights. Analytical and experimental work is in progress to permit correction, refinement, or replacement of the damage functions developed in the AASHO (American Association of State Highway Officials) Road Test in the late 1950s to make them more relevant to the types of pavement now in use.

Another aspect of the heavy vehicle problem is the control of deliberate overloading in excess of legally prescribed axle and gross weights. Damage functions increase dramatically as loading exceeds those weights for which the pavement or structure was designed. To assist the states in their efforts to control this problem, research and development programs are in progress to upgrade existing weighing-in-motion systems designed for installation in pavements. A new system that permits ordinary short span highway bridges to be utilized as weighing platforms is also under development.

Relief of Urban Congestion and a Safer Urban Driving Environment

Over the past 10 years a variety of independent techniques has been developed to permit more efficient use of existing highway capacity such as digital computer control of traffic signal systems, express lanes for high occupancy vehicles, variable message signs on expressways to alert motorists to road conditions, and radio information systems to provide routing information. All of them have limitations. The computer control of traffic signals in a central business district, for example, does not permit adequate cooperation by the individual driver when localized congestion occurs because he lacks information about its effect on his intended route. Therefore, there should be some information feedback to the driver so that he can make choices that will improve the general flow, increase safety, and reduce his driving time. Research activities now in process are working toward this objective and will also examine

- Improved methods to expedite the movement of public transit vehicles with but minor effect on or even positive improvement in the general flow;
- Special routing and scheduling techniques for cargo vehicles in urban areas to permit more efficient movement and less impact on rush hour traffic; and
- Simulation techniques to permit local transportation officials to select the most effective mix of known traffic management strategies to optimize flow for their particular networks and conditions.

A parallel effort is aimed at intercity corridor management. A major demonstration of these techniques is expected in the Long Island corridor east of New York City over the next several years. It will involve the use of a mixture of automatic and manual traffic situation monitoring systems, and a central computer to analyze the data and control the driver information systems to guide motorists around main points of congestion. Results from a less extensive intracity corridor management system in Dallas, Texas, indicated measurable improvement in both flow efficiency and accident reduction.

Sophisticated urban traffic management can have a major impact not only on traffic flow and safety but on economics and public health. Substantial reductions in automotive emissions can be obtained by the maintenance of smooth flow rather than turbulent stop and go traffic. There is an obvious need to examine the economic tradeoffs involved in the reduction of automotive air pollution, especially since we are now approaching the point where emission levels will present health hazards only in localized urban areas of intense, combined industrial and transportation activity.

Automated Systems

The enormity of the highway safety problem, the long-range implications of the petroleum crisis, and the rapidly escalating cost of highway maintenance, all suggest that major new thinking will be required to meet future needs.

Automation of highways and highway vehicles has frequently been suggested as a way in which significant improvements in safety and flow efficiency could be achieved. Such automation requires not only technical hardware development but major institutional changes and, eventually, a painful and perhaps impossible revision in public attitude toward vehicle operation. We have pursued both aspects of the problem in our programs to date, and there are indications that automation, or at least partial automation, of some densely traveled intercity corridors and major radial routes in metropolitan areas might be cost effective. Preliminary results indicate that lane capacity could be doubled, and, with adequate investment in the reliability of components, there could be an enhancement of safety. Automation has an attractive side benefit in that it lends itself to the use of alternative power sources such as electricity, utilizing new developments in conveying such power from the guideway to the vehicle without the use of exposed third rails. Because the range of application appears restricted to certain high-volume intercity and urban expressway facilities, it has been our position that highway automation must be both evolutionary and "downward compatible"; that is, the automated vehicles must be capable of traveling on the nonautomated portion of the highway network.

Transferring vehicle control from human to automatic has a large safety potential. Rapid advancements in both sensors and computers are making such a transfer practical. A study by the Jet Propulsion Laboratory in 1977 to assess this possibility estimated that a continuation of past trends in technical progress could make such automatic autos both technically and economically feasible within this century.

The automotive industry has been moving rapidly to replace the complex mechanical linkages, formerly used to control fuel mixture and a variety of other functions, by electronic systems. It is quite possible that within 10 years most automobiles on the road will have a microprocessor onboard for such functions. If the cost is low enough, the microprocessors could be used to perform a variety of tasks that would enhance safe operation of the vehicle. For example, many erratic maneuvers at major highway intersections are made because of the driver's confusion about appropriate routes or exits. The technology already exists to provide the individual motorist with navigational information tailored to individual requirements. Precise and timely instructions to the motorist can greatly reduce the incidence of dangerous maneuvers and can also reduce unnecessary travel, producing fuel savings. Early prototype hardware produced in the program has been selected by both Japan and Germany as a component of the trial systems they already have in operation.

Onboard microprocessors have other potentials in the highway safety field. They may alert motorists to hazards on rural two-lane highways where some 60 percent of our highway fatalities occur. Conceptualization of possible applications is just beginning, resulting largely from frustration with the limitations of visual signing and marking systems.

Impervious Concrete

One of the developments emerging from the major effort to control damage to concrete bridge decks produced by de-icing chemicals has been the invention of a clever system for producing essentially impervious concrete. This consists of adding finely divided wax pellets to the concrete mixture which is poured and cured in the normal fashion. While the material was originally developed for bridge decks, it has applications in many other situations where concrete is exposed to salt, water, or other corrosive liquids. A number of highway departments are considering its use for bridges in tidal waters, and we have inquiries from the U.S. Navy and the U.S. Army Corps of Engineers concerning other applications.

Vehicle Safety and Fuel Economy

The industry has historically taken the position that improvements in safety and fuel economy are not important factors in marketing cars. In recent years, Federal safety and fuel economy regulations have forced improvements. Readily available technical solutions are now being applied by the industry to satisfy the Federal regulations by 1985. Research during the next five years should be directed to finding additional technical solutions to the social problems created by the pervasive use of the automobile.

This research should include advanced engines; alternative engine, fuel, and drive train combinations; applications of microelectronics and other expanding technologies to the automobile; and development of advanced automobiles that integrate improvements in aerodynamics, structures, engines, and other systems. We are currently initiating an Integrated Vehicle Systems Program to serve as a basis for evaluating how advanced technology can be developed and applied to solve the social problems of the automobile. Test vehicles and the associated technology development programs will demonstrate how to construct automobiles that will improve safety and reduce the national dependence on foreign oil.

The Secretary of Transportation has called for a program to establish research directions to provide advanced automotive technology options for "socially responsible" motor vehicles of the future. A conference held in Boston on February 13 and 14, 1979, attracted widespread representation from the automobile industry, government, and academia. The conference identified promising technology alternatives that would lead to advancements in engines, fuel, drive trains, structure, and other vehicle system concepts. A summit conference between the Secretary and industry leaders was held on May 18, 1979, to address the future direction of the program.

Control of Diesel Emissions

The use of light duty diesel engines in passenger cars offers the potential for significant improvements in fuel economy. Preliminary research, however, indicates that the exhaust products could create health hazards, particularly if diesel cars make up a significant portion of the motor vehicle fleet. At present, no technology exists for the control of all the harmful products in diesel exhaust. Research approaches should include prevention of the formation of harmful products through improved control of the combustion process and removal of harmful products through exhaust treatment devices.

RAIL TRANSPORTATION

Improving safety, reducing petroleum fuel consumption, and improving finance are problems that beset rail transportation. Handling growing shipments of irradiated nuclear fuel casks is an emerging concern, and the current standards for the casks need to be evaluated. Various track and safety programs now underway will enhance railroad safety and reduce costs. Electrification of high tonnage routes is an attractive way of reducing costs and petroleum usage. Two specific areas of research are described below.

Safe Rail Movement of Hazardous Materials

Transfer of sizable quantities of radioactive material from nuclear plants must begin shortly, and most of the transport burden will fall upon the railroads, since this mode is the most cost effective. Rail cars carrying nuclear wastes could possibly suffer the same kinds of accidents in which tank cars have been involved as they will be carried on existing track under normal railroad operating conditions. Yet, standards for testing nuclear packaging may not be consistent with those required for tank cars, especially when prescribing the requirements for fire resistance. A multimodal, interagency research and test program is needed to ensure that radioactive materials and additional shipments of other materials caused by shifts in energy consumption can be safely and economically moved. Accordingly, within the next five years, analysis and testing of nuclear cask designs should be conducted.

Results from the continuing track structure inspection program will be used to minimize accidents and reduce costs. The program will lead to derailment reduction by establishing operating limits for vehicles.

Rail Energy Conservation

Intermodal shipments can reduce energy consumption by switching freight to less energy-intensive modes. The

Intermodal Demonstration project uses existing conventional equipment and aggressive marketing of improved service to attract more trailer-on-flat-car traffic in specific corridors. The Intermodal Systems Engineering project is assessing innovative intermodal equipment and identifying possible terminal improvements. A third project is concerned with the design, fabrication, and testing of a low-profile intermodal car that can operate through the restrictive clearances existing in the northeast. The objective is to move more long-haul trailer traffic by rail instead of highway. Not only will this reduce consumption, it will also improve safety and reduce highway congestion and pollution.

With increasing intermodal traffic and more unit trains, traffic density has reached levels to justify electrification in some rail corridors. Since an electrified railroad can use energy generated by a variety of sources, it reduces dependence on petroleum. Research is focused on opportunities for improvement in locomotives and power distribution systems.

URBAN TRANSPORTATION

Solving the problems of passenger transportation in the city and suburbs requires the development of systems that will reduce congestion, pollution, and energy consumption while increasing mobility, especially for those without auto access. Three such systems are described below.

Paratransit

Areas of low population density cannot be served economically by fixed route systems. Paratransit services are the best known method of extending public transit to such areas and of providing service to the handicapped and elderly. An aggressive effort is underway to find solutions to the institutional problems impeding the integration of paratransit and public transit services and to develop tools to enable the operator to deliver high levels of service at reduced costs.

One deterrent to the effective integration of paratransit with public transit is the lack of an appropriate vehicle designed to meet the special requirements of paratransit services, including the transportation needs of the handicapped. The Federal Government's role is to stimulate the motor vehicle industry to develop such vehicles at a price affordable by the operators. Up to three designs will be developed, tested, and evaluated in engineering prototype form.

Downtown People Movers

The Downtown People Mover program is intended to provide a national demonstration of how fully automated systems can improve the circulation and distribution of people in the center of cities. The goals are to determine the cost savings, to assess the economic impact on the central city, and to test the feasibility of people movers as circulators, distributors, and substitutes for fixed guideway systems such as subways.

Energy Conservation and Propulsion Technology

A joint DOT/Department of Energy (DOE) program has been initiated to develop more energy efficient and environmentally acceptable mass transit vehicles. A near-term program provides for an in-city demonstration of the gas turbine engine in buses to verify its capability to operate on a broad range of fuels, cleaner emissions, and lower maintenance costs. A longer range program is the development of flywheel energy storage systems (FESS) (pure flywheel and diesel/flywheel hybrids) as alternative propulsion applications. It seems probable that flywheel-propelled vehicles can reduce the dependence of transit vehicles on petroleum fuels without significantly increasing life-cycle costs, reduce noise and emissions, and offer electrically propelled mass transit buses with comparable operating economies and route flexibility to the current diesel bus. Two different engineering prototypes will be designed, fabricated, tested, and evaluated in separate 40-foot buses.

PIPELINE TRANSPORTATION

During the past three decades, there has been substantial growth in pipeline transportation systems, with hazardous materials pipeline systems tripling in mileage. These systems transport natural gas, crude oil and various petroleum distillates, and certain other chemical commodities.

The basic strategy for protecting the public from the adverse impact of failures of pipeline transportation systems has two parts. First, efforts should be made to prevent failures of operating pipelines, and, second, efforts should be made to minimize adverse impacts if pipelines fail.

One area where research and development is needed is in finding better ways to inventory existing pipelines, both onshore and offshore. This would include development of techniques and equipment that determine the operating condition and useful safe service life of existing pipelines. Corrosion control, including such specific items as stress-corrosion-cracking of existing pipelines, also requires extensive research. In a related area, there is a need for R&D to test and evaluate the suitability of various materials used in the construction of pipeline transportation systems, such as those contemplated in the arctic and offshore. This would include assessing the safety and suitability of construction and operating methods. There is also a need for a range of studies to analyze and evaluate the safety, economic, and environmental criteria used in the development of pipeline safety regulations, as, for example, in fracture mechanics analysis of pipeline welds and proposed arctic and offshore pipelines and terminals.

Cryogens, such as liquefied natural gas (LNG), will increasingly be transported and stored in strategically located facilities. Research on LNG spills, besides being necessary for future pipeline safety, would have applications for siting problems associated with LNG storage. Examples of specific problems would be: How do the vapors disperse? Under what conditions do they ignite?

Continuing needs are to find effective ways to train and coordinate emergency response personnel and to develop improved response techniques and equipment to combat large pipeline ruptures and resulting fires. There is also a need to find ways for analyzing the safety effectiveness of the various pipeline technologies and the key parameters that influence their safety performance.

TRANSMODAL PROGRAMS

Several technologies offer the opportunity to make a contribution to more than one mode of transportation. The following paragraphs describe important areas for research.

Systems Applications R&D—Automatic Control/Automation

The DOT has a vital interest in the development of the necessary models and systems to take advantage of the rapidly expanding capabilities of microcomputers and their input/output devices.

Microcomputers have emerged as a dominant technology of the 1970s. Their costs, weights, and sizes are decreasing drastically while their computing capabilities, versatility, and reliability are steadily increasing. These revolutionary changes can have a significant impact on energy efficiency, automation, safety, and traffic control. It is essential that the impact of these devices on all phases of transportation be assessed so that the Department can begin to formulate policies and structure programs to take advantage of the opportunities they provide.

Microcomputers will find applications in many diverse areas, including both onboard and fixed-base systems. Onboard applications could include displays, position-indicating systems, and computation of optimum power and fuel settings. Examples of fixed-based applications are traffic control systems and displays, and freight and passenger control and logistics. The role of DOT is not to develop computer technology, however, but to assure the orderly and safe flow of traffic. There are many ways that this flow can be enhanced and optimized through either ground-based or vehicle-borne high-speed computations.

The Department is developing models for commodity flow, air traffic, and urban transportation which will be able to utilize microcomputers. These models or systems and their implementing hardware and software will begin to have significant impact on transportation in the next five years. Areas of research to be conducted or encouraged by the Department include not only the systems application studies and experimentation but the development of the appropriate sensors and actuators that interface with the microcomputers. The challenge is to develop low-cost, batch-fabricated, reliable sensors and actuators that easily interface with digital systems. Integrated silicon sensors have shown promise for measuring pressure, strain, acceleration, temperature, and radiation. Integration of signal conditioning circuitry with the sensor is an attractive possibility. The development of low-cost display technology that is reliable, easy to interface with digital circuits, and has low luminance noise is also high on the list of technologies to encourage.

Energy—Substitute Fuels

The thrust of the Department's energy program is toward conservation, particularly by reducing demand through carpooling, 55-mph speed limits, vanpooling, fuel economy standards, and so on. Although the supply or production side of the energy picture is primarily a DOE responsibility, DOT is also vitally interested. The economics of new sources of energy could have a profound effect on the national transportation system. In addition, any shift in the movement of energy materials, i.e., feed stocks, gasohol, and output byproducts, is a transportation issue. The Department is investigating the problems of coal transportation impacts (coal roads) and substitute fuels (syncrude, alcohol). We intend to participate in the National Alcohol Fuels Commission's investigation of the long- and short-term potential for alcohol fuels from biomass and coal. We foresee increasing use of substitute fuels beginning in approximately five years, although we do not see a major switch from petroleum-based fuels in the foreseeable future. It is imperative that we expand our research to develop a broad spectrum of technology options that will be available over the longer term.

Lagging Productivity

There is increasing concern about the slowing rate of growth in both capital and labor productivity in the national economic system. There are opportunities for substantial productivity improvements in transportation that are not being exploited or, in some cases, are barely explored.

In particular, the application of the new technologies that permit noncontact suspension and propulsion of surface vehicles offer the promise of higher speed and reduced guideway maintenance that would impact on time-sensitive commodity flow, intercity passenger travel, airport access, and urban people movers. Although many of these research efforts will not achieve tangible results within the next five years, it is necessary to start research now on systems and components.

In several European cities, capsule-carrying pipelines are being used for garbage disposal and apartment delivery. Such possibilities have hardly been explored in this country.

Commodity Flow R&D

Research will be conducted to identify systems that could improve the commodity flow system by providing faster, safer, and less expensive service to the markets of the future. Sources of inefficiency will be identified, and system design alternatives will be evaluated. Design alternatives to be considered range from an entirely new system, perhaps using only existing rights-of-way, to advanced concept intermodal systems. The contributions that might be offered by advanced command, control, and communications technology to meet future shipper requirements will also be identified and evaluated.

Hazardous Materials Transportation R&D

Recent trends have been for an increasing amount of materials that are explosive, unstable, toxic, corrosive, flammable, or radioactive to flow in commerce. These trends are expected to continue during the next five years, and minimizing the threat to public safety will pose major challenges. Protecting the public from the adverse impacts of releases of hazardous materials has two essential parts: first, to prevent the accidental releases; second, to minimize adverse impacts if and when they occur.

To help prevent hazardous material releases, R&D is needed to find better methods of packaging to preserve the integrity of containers that are exposed to severe environmental conditions during transport. There is also a need to find better ways of keeping track of the various types of hazardous materials. This includes information on their physical and chemical properties, the quantities of each moved, shipment origins and destinations, modes, and routes, with instantaneous location capability. In addition, studies must analyze and evaluate the impacts of safety regulations, especially with respect to facility siting and operation, containerization, and routing.

To reduce the impact of hazardous material spills, there is a need for R&D to find ways to more effectively train emergency response personnel and improve response techniques and equipment. This includes selecting a management information system that will enable emergency response planners and teams to minimize the damage from accidental releases of hazardous materials in transit, as well as to prevent them in the first place.

There is a continuing need for R&D to find ways of analyzing and evaluating the safety effectiveness of the various modes of transportation. This includes identifying key parameters that influence safety performance and measuring sensitivities of the parameters. Improved planning methodologies for forecasting future trends in the volume and flow patterns of hazardous materials and to assess their implications for safety regulation are necessary. Research to develop better ways to identify and evaluate improved hazardous materials transportation systems would be included. Finally, there is a need to assess the problem of terrorism and sabotage and to develop effective countermeasures to protect these systems from such threats.

Environmental Protection Agency

The main function of the research program of the Environmental Protection Agency (EPA) is to provide a sound scientific base for promulgation, enforcement, and review of environmental standards and regulations. Primary research results provide information to characterize pollution threats, define health and ecological dangers from polluting sources, and, when possible, develop, test, and evaluate appropriate means to control pollution.

Inherent in our research is an appreciation and careful consideration of the costs, in both capital outlay and quality of life degradation, that may result from controlling or not controlling identified harmful pollution.

Part of our research program also anticipates potential threats, particularly from new types of pollution that may result from our nation's shifting energy posture to meet the looming energy crisis. Diesel cars with their high fuel economy may capture a significant share of the American automobile market but may, at the same time, contribute substantial amounts of hazardous pollutants to urban air. Similarly, oil shale may become a key domestic energy supply, but it too may produce new types and concentrations of pollution. The philosophy applied in both of these cases is representative of many of our other projects, that is, that gaining scientific knowledge about a pollution source before it becomes a problem will avoid costly control retrofits. Hence, we work to keep the environment clean at the lowest cost.

This summary focuses upon our fiscal year 1979 research plan for eight separate but interrelated and ongoing EPA programs supporting the Agency's regulatory mission. A ninth section, on global pollution, is included to illustrate the anticipatory thrust of our activities, wherein we study basic environmental processes, identify potential environmental problems, and seek over the long term to pinpoint areas for possible regulation.

TOXIC SUBSTANCES

Modern industrialization has resulted in a dramatic increase in the production, variety, and use of chemicals, some of which can be directly toxic to man and the environment. Other chemicals have delayed deleterious effects of subtle, little-understood consequences. With the recognition of the long-term, less-than-lethal hazards comes a dilemma that is key to the future role of chemicals in our society. The dilemma is basically how to derive maximum social benefit from modern technology while limiting attendant risk to an acceptable level. Inherent in the concept of determining acceptable risk are social judgments rendered singly or collectively through our various institutional processes. It is EPA's task to regulate toxics in the environment to ensure that acceptable levels of risk are not exceeded. It is the generation, evaluation, and continual update of the scientific and technological data base necessary to support that regulation that forms the fundamental mission of the toxic substances research program at EPA.

To fulfill our support mission in chemical regulation over the next several years, we have established two research goals: to characterize the nature and extent of risks posed by potentially hazardous chemicals, and to develop control strategies, technologies, systems, and management practices that will prevent, interdict, or at least minimize exposure to hazardous chemicals. The first goal involves research in three areas: effects of toxic substances on humans and the environment, assessment of the magnitude and mechanisms of population exposure to toxicants, and production of integrated risk assessments and criteria documents to support future regulatory decisions. The second goal encompasses two areas for research: characterization of toxic emission sources by developing a data base that describes the chemical industry, and development and evaluation of alternate control strategies and technological systems to mitigate human exposure to hazardous chemicals.

Each of these tasks requires the development and application of validated measures and techniques to serve as quality assurance of research results. Documentation of the precision and accuracy of the data base used for risk assessment and regulatory decisions will be achieved through a quality assurance program that covers personnel, methods, equipment, and data handling procedures. The quality assurance program must be applied to each of these components individually and as they function in an integrated research system.

AIR POLLUTION

Significant progress has been made in cleaning the air of the United States, but the more that is learned about the complex mixture of fine particulates and other pollutants that are in the air, the deeper the concerns regarding their potential adverse effects on health and welfare. Our atmospheric research program will concentrate on three particularly troublesome pollution problems: interregional transport and transformation of air pollutants, the fate and effects of air pollutants, and their diffuse and naturally occurring sources.

Interregional Transport and Transformation of Air Pollutants

EPA's evaluation of the chronic health risks associated with particles that are subject to interregional transport will be achieved through animal toxicology studies on chronic effects of exposure to particles of varying chemical composition, including sulfates and nitrates. Subsequent studies will assess the effects of exposure to mixtures of particles of known chemical composition. Concurrently, long-term epidemiological studies are being designed and conducted. Our research will deal with inhalable particles (defined as having an aerodynamic diameter of 15 micrometers or less). Research into the improvement of air quality simulation models (AQSM) to predict the long-range transport of particles is necessary. Our research plan also includes studies to provide standardized monitoring systems that will assure comparable data throughout the system.

The transport and transformation of ozone/oxidants and precursors also need thorough investigation. We are beginning to develop and validate models that can address the interregional transport processes of ozone/oxidants. The first step in the development of these models is to quantify multiday rather than multihour dispersion and transport. The second step is to identify and quantify removal processes for ozone and its precursors, and the third step is to define and quantify multiday chemical mechanisms associated with transport. The transport and transformation of secondary organics is the final problem to be addressed in this category. We need to develop an understanding of the prevalence, concentrations, formation, and transport of primary and secondary organic particles and vapors in urban air. Our research will develop practical methods for identifying and measuring specific organics in ambient air. Also included will be field studies to determine the presence of secondary organics in urban ambient air.

Fate and Effects of Air Pollutants

The fate and effects of trace elements are of concern to our research program because these elements can cause biological damage to organisms, ecosystems, and humans. Our trace elements research will quantify emissions from stationary and mobile sources and will evaluate control technologies. We will also develop methods to identify the chemical form of trace elements in emissions and in the ambient environment. Our research will provide the scientific basis needed to assess potential health and environmental risks from these emissions. Transport and transformation studies will be carried out to provide better methods for predicting the ultimate form and distribution of trace metals, including the potential for uptake into the food chain.

The fate and effects of airborne particles which degrade visibility are also of concern, since reduced visibility has the potential for producing human and material costs in the form of automobile accidents, transportation delays, and mental depression. In our research we intend to determine the chemical composition of the time particulate size fraction primarily responsible for visibility degradation. We will also provide the data base required to evaluate possible control strategies. Additionally, we will provide methods to evaluate the human and economic cost of visibility degradation.

The diffuse sources of particular concern to our study are fugitive emissions, including fugitive dust. These complicate the problem of revising present ambient air quality total particulate standards. We do not have the data to assess the relative contribution or the health risks associated with particulates from fugitive sources. We have, therefore, established a research plan to better address this problem. We will develop instrumentation to provide size spectrum and chemical composition data for ambient air particulates. These measurement methods are required to discriminate among the various kinds of particulates, their sources, and their potential adverse health effects. The emphasis will be on the design and evaluation of alternative control strategies and will include the development of industrial fugitive process particulate (IFPP) controls to reduce emissions.

Naturally Occurring Pollutants

The naturally occurring pollutants of primary concern are hydrocarbons. We hope to provide emissions data to assess contributions of natural sources of hydrocarbons to ambient concentrations of ozone/oxidants in metropolitan areas. Finally, we hope to provide characterization and effects data to determine the relative importance of natural versus anthropogenic air pollutants in causing allergenic discomfort.

INDUSTRIAL WASTEWATER

Industrial wastewater can contain a variety of toxic and conventional pollutants such as phenols, benzene, chlorinated hydrocarbons, metals, suspended solids, and oxygen-demanding compounds. The industries of prime concern to EPA's industrial wastewater research program are the organic chemical, pulp and paper, iron and steel, and petroleum refining industries. The organic chemical industry produces wastewater with the highest concentration of toxic pollutants; the other industries produce less polluted wastewater but at higher discharge volumes.

Recognizing the major threat industrial wastewater poses to the maintenance of water quality, Congress made the elimination of pollutant discharges to U.S. waters a national goal by passing the 1972 amendments to the Federal Water Pollution Control Act (FWPCA). As a result, industries must adopt successively more stringent control measures aimed at the removal of 65 toxic substances from their effluent. Because it is either not feasible or prohibitively expensive to treat their wastewater, many industries have responded to these strict effluent limitations by trying to recycle as much wastewater as possible. Recycle and reuse programs have also been stimulated by decreasing supplies of water and the resulting increase in cost. An important side benefit of wastewater recycle and reuse has been that as products are recovered from the effluent stream less solid waste is produced. But to achieve effective recycling and reuse, certain major control technology problems must be resolved. The diverse composition of industrial effluents demands technology that sufficiently renovates water for reuse in many sensitive industrial processes. Designs are needed for centralized treatment plants that will be fed by small industrial concerns unable to afford their own water recycle plants. And, economic viability and operational feasibility must be demonstrated to ensure commercial acceptance.

To provide the control technology base for water treatment and recycle and reuse, research will have to proceed on a number of fronts. Our research perspective can be defined as developing control concepts and prototype technology, demonstrating the commercial viability of the technology, and, finally, making the technological information available so that industry can install the technology and meet wastewater discharge laws. Research must also have the continued participation of industry in paying part of the enormous cost of developing effective control technologies. The industrial wastewater research program, therefore, has three prime goals: technology for treatment of toxic pollutants, technology for reuse and recycling of industrial wastewater, and information for environmental assessments.

Identification and Treatment of Toxic Substances

EPA needs to rapidly refine its list of 65 toxic substances identified under the 1977 Amendments to the FWPCA. Many potentially toxic compounds not on the list are being identified in industrial wastewater samples analyzed by EPA's Effluent Guidelines Division (EGD), and, conversely, several of the 65 substances have yet to be found by the EGD survey. By 1980, development of analytical methods to identify and measure contaminants in wastewater will be completed. The first chemical analysis methods developed will be appropriate for the 65 toxic substances.

Because the repeated application of any analytical method involves major costs, we intend to develop sets of generic or surrogate parameters representative of groups of individual pollutants. Existing cost-effective analytical methods such as nonspecific detector analysis will be complemented by the surrogates. Following the development of the methods to analyze water contamination, we will be able to test the adequacy of conventional (i.e., biological) technologies for treating wastewater effluent. We also plan to document effective practices currently in use for pollution abatement associated with industrial operation, maintenance, and housekeeping functions.

Reduction of Water Use

Although it is used primarily as a coolant, water also serves as a process chemical and solvent and thereby becomes a major contributor to industrial water pollution. Therefore, our research will be oriented toward the development and demonstration of water use reduction and closed-loop operation for process and solvent application. First, we will develop recycle and reuse methods for reducing water use in those industries that are high water consumers. Since it may be impossible to reduce water use significantly or to achieve closed-loop operation for some industries, we also plan to examine alternative low pollution processes as well as modifications (such as changes in raw input materials) in an effort to limit sources of toxic contamination of wastewater. This research will focus primarily on the textile and organic chemical industries.

Environmental Assessment Studies

Environmental assessments are important tools in EPA policymaking. Under this research program we will develop and implement research to estimate the quality of environmental measurements important to regulation and pollution control. We will also consolidate, or develop, linkages between the important data sources on industrial effluents. We will then develop a list of the most severe discharges for the textile, iron and steel, and other industries. This list will be used to direct future industrial wastewater research.

WATERSHED MANAGEMENT

Water can be polluted directly by point source discharges of contaminants, particularly toxic chemicals, or indirectly by diffuse, or nonpoint, sources of contaminating materials. To minimize the detrimental effects of point and nonpoint sources on an ecosystem, comprehensive management of the watershed is required. To accomplish this, we must first develop a fundamental understanding of the dynamics of the physical, chemical, and biological interactions in the watershed. We can then devise the predictive capability required to estimate the extent of ecosystem modification induced by point and nonpoint sources of pollution. Through monitoring and analysis, the pollution discharged into a watershed and the relationship between pollutant loading and water quality can be established. Finally, using this information and taking into consideration socioeconomic and energy factors, the appropriate pollution control measures can be specified.

The Water Control Management Research Program will concentrate on nonpoint sources of pollution. Our first goal will be to understand further the complex relationship between these discharges and the quality of the affected waterways. Our second goal will involve the development and evaluation of cost-effective management methods to limit pollution from nonpoint sources. Finally, we will focus on the development and demonstration of effective implementation strategies for nonpoint source control methods. The end result will be the design of improved voluntary acceptance approaches to control nonpoint sources. To aid in this process, we will develop relevant economic impact data, analyze existing institutional mechanisms, and investigate opportunities to integrate nonpoint source control programs with other pollution control programs.

DRINKING WATER

In recent years, modern techniques and equipment have become available to detect and measure very small quantities of contaminants in drinking water. The application of these sensitive methods has identified potentially hazardous organic contaminants in drinking water already treated by filtration and chlorination. With the explosive growth in the use of chemical compounds in industry, homes, and farming, chemicals are also finding their way into our drinking water sources. Some of these chemical contaminants may cause cancer, genetic mutations, or birth deformities. Protection of our drinking water from chemical as well as microbiological contaminants, is, therefore, an important preventive public health measure. EPA drinking water research goals are to develop detection, measurement, and monitoring methods to further identify and quantify drinking water contaminants, develop treatment technology to reduce contaminant levels to acceptable concentrations, assess the health effects of contaminants, and establish scientific and technical bases for action to protect groundwater quality.

Methods to Identify and Quantify Contaminants

The importance of the first goal rests in the fact that without reliable contaminant measurement and monitoring methods, it will be impossible to provide scientifically valid and legally defensible data to support and enforce the regulations of the Safe Drinking Water Act. Consequently, primary emphasis will be placed on continuing our quality assurance program in EPA, state, local, and contract laboratories. Specifically, our research will support these laboratories by providing validated measurement systems, reference standards and samples, interlaboratory performance tests, procedures and criteria for certification of laboratories, and other quality control materials and services required to document the precision, accuracy, and intercomparability of monitoring data. A companion objective will be to develop analytical methods for measuring classes of organic compounds.

Treatment

In treatment technology research, the second area of emphasis in the drinking water program, our first task will be to develop treatment processes to control organic compounds.

Health Effects of Contaminants

To accomplish our third goal, the gathering of data on the health effects of contaminants, we will focus upon the carcinogenic potential of certain organic compounds and the relationship between the presence of certain inorganic compounds in drinking water and cardiovascular disease. Our health effects research into organics will pursue two courses. First, we will examine those compounds that have already been identified as potential human health hazards. Second, our investigations will focus on various groupings of compounds selected on the basis of observed concentrations, frequency of occurrence, and preliminary hazard evaluations from either available data or short-term bioassays.

As with organics, we will develop information on the health effects of inorganic compounds. Here, we will perform a comprehensive review of a number of inorganic maximum contaminant levels, including those for fluorides, nitrates, selenium, lead, cadmium, and arsenic. We will also specifically assess the health effects of asbestiforms. Additionally, we will attempt to determine if the water softness/heart disease correlation is caused by benefits from hard water or detriments from soft water. The fact that some chemical compounds, suspected to be carcinogenic, are formed during chlorination of drinking water creates a dilemma. It is important to limit human contact with suspected cancer-causing agents, but it is also essential to keep waterborne infections at their current low levels. Consequently, we will continue to monitor outbreaks of waterborne diseases and to characterize the causative agents to develop data necessary to maintain desired levels of contaminant-free drinking water.

Protection of Groundwater Quality

A primary objective of the fourth goal of the program is to identify major problems in protection of groundwater and to provide the appropriate assessment methods to states and local communities. Work on assessment methods will include development of biological and chemical indicators of groundwater pollution, methods to detect pressure increases resulting from well injections, and protocol for determining pollution potential of activities in a specific area. Research on a corollary objective, to produce the information upon which to base scientific and engineering guidelines for source control criteria, will focus on petroleum exploration and development, application of waste to the land, artificial recharge of groundwater aquifers, and agricultural practices. In addition, we intend to complete our assessment of the severity of groundwater pollution nationwide, by adding 16 states to the 34 already studied.

ENERGY AND ENVIRONMENT

The profile of U.S. energy development and use will undergo major changes in the years ahead. Although only slowly evolving, it appears that our national energy policy will call for a widespread conversion of utility and industrial power facilities from scarce oil and gas to plentiful coal, decreased fuel consumption, and, in the longer term, the use of technologies that are only now beginning to emerge for the production of liquid and gaseous fuels from coal and oil shale. These shifts in energy development and use pose potential threats to human health in the next two decades, and a multitude of information is needed to avert massive future environmental impacts.

Our energy and environment research, development, and demonstration program has the concurrent goals of assessing the magnitude of health and environmental problems to establish a data base for the formulation and refinement of useful standards and developing effective, commercially acceptable control technology. The major energy-related environmental concerns requiring intensive research are, in order of priority, to minimize impacts of increased conventional combustion, minimize impacts of increased energy extraction, quantify cancer-causing potential of diesel soot, and minimize impacts of emerging energy technologies.

Conventional Combustion

Our research in conventional combustion will focus primarily on the development of control technology to reduce the environmental impact from expanded coal combustion. Here, we will concentrate on developing and demonstrating a low-NO_x burner capable of controlling emissions from new and existing utility and industrial coal-fired facilities. We will attempt to resolve the remaining operational problems associated with current flue-gas desulfurization processes. And, we will also perform environmental assessments of conventional utility, industrial, commercial, and residential combustion sources with emphasis on currently unregulated pollutants such as heavy metals and organic compounds.

A second research objective will be to improve the efficiency and cost effectiveness of current particulate control technology. As part of this work, we will evaluate the applicability of fly-ash conditioning agents to various types of coal, develop and demonstrate economical control technology for disposal of large amounts of fly ash and sludge, and evaluate the impacts of these combustion products on terrestrial organisms and ecosystems.

Finally, we will develop effective physical and chemical coal-cleaning processes for sulfur control and flue-gas denitrification technology capable of high levels of NO_x control for utility and large industrial applications.

Energy Resource Extraction

Priority research also will deal with the environmental problems associated with energy resource extraction. One particular area of concentrated research will be the identification of specific health damages resulting from coal mining in general and the quantification of the effects of western coal and oil shale extraction on surface and groundwaters. Research on environmental damage caused by coal extraction will focus on methods of limiting pollution through improved sediment and mine drainage control techniques and on a better definition of Best Management Practices (BMPs) for use by the regulatory programs of EPA and the Office of Surface Mining in the Department of the Interior. Water quality effects in the west will be treated by quantifying the relationships between pollution from strip mining and oil shale mining and the damage to aquatic organisms and their habitat. As part of this research, existing water monitoring networks in western energy extraction areas will be augmented to collect data on annual water quality trends.

Another research avenue we will pursue is the assessment of the potential environmental damages from oil and gas extraction and handling. In addition, we will assess the impact of uranium extraction on surface and groundwaters and will develop appropriate control technology. We will also demonstrate revegetation and reclamation techniques for western surface-mined areas. This work will include assessment of the impact of mining and reclamation practices on western surface water hydrology, forests, and range lands. Methods will be developed for the ecological recovery of toxic spoils generated by surface coal mining. Finally, we will establish energy-related water quality baseline data in major eastern coal areas that have been identified for new development in the late 1980s.

Diesel Engines

Anticipated market penetration of diesel cars will make the potential health problems associated with the use of diesel engines a major issue in the near future. To obtain a health effects assessment for diesel soot, we have initiated whole-animal exposure studies using intratracheal instillation, inhalation, and skin painting. Simultaneously, we plan to continue our research dealing with evaluations of the carcinogenic potential of diesel particulate emissions from different fuels and engines. We will also make ambient measurements to estimate the background levels of "soot" (i.e., prediesel levels) in population centers throughout the United States. Additional work will be aimed at predicting the exposure of various population segments to diesel soot as a function of market penetration. Information generated from this work, coupled with available health effects data on diesel soot, will allow risk assessments to be made.

Research relating to the effectiveness of various control devices will concentrate on identifying ways to reduce substantially the quantity and potency of particulate emissions while still retaining the superior economy characteristics of diesel fuel. Here, we will evaluate current catalytic converter technology as well as advanced control techniques with the aim of identifying cost-effective control methods.

New Energy Technologies

The multi-billion-dollar energy technology development program now underway will present environmental problems as the technologies become commercialized. EPA will continue to apply research to the broad spectrum of evolving energy-related processes, but will change the research direction and emphasis as necessary when specific new technologies approach commercialization. We see as our first task the environmental assessment of the major first generation in situ and surface oil shale processes. Our assessment will characterize input materials, products, and waste streams; identify potential environmental impacts; describe performance and costs of control alternatives; and suggest appropriate discharge limits and control technology. Similar assessments will be performed for major coal gasification, coal liquefaction, and fluidized bed combustion processes. In addition, we will produce environmental assessments of other important new energy technologies including geothermal, waste-as-fuel, solar, advanced power cycle, biomass conversion, and energy conservation techniques.

SOLID WASTE

Solid waste is produced in this country at a rate of 487 million tons per year. This enormous amount of solid waste creates a number of environmental problems, principally the protection of public health. The EPA research response to the solid waste problem has grown out of the recognition of its enormity, the severity of public health risks, and a congressional mandate provided by the 1976 Resource Conservation and Recovery Act (RCRA). In addition to charging EPA with protecting public health, the Act also calls for EPA to develop means to conserve natural resources through waste reduction, recovery, and reuse of materials now discarded. Thus, the EPA solid waste research program is characterized by two parallel directions of investigation: development of safe disposal management practices, and development of resource recovery methods.

Solid Waste Disposal

There are four high priority objectives which are steps toward better methods for waste management. First, we need to develop new or improved methods for landfill site selection, design, operation, and maintenance. We plan to develop methods to control and accelerate the decomposition of landfill wastes, to estimate potential impacts of gas and leachate migration, and to control such migration. The influence of mixed wastes (industrial and municipal) on the decomposition processes and on gas and leachate production at codisposal landfills will also be determined. Further, our research will identify changes in municipal solid waste management practices required to deal with anticipated increases of industrial waste and sewage sludge that are to be disposed of in municipal landfills.

Second, we need to evaluate the technical, environmental, and economic strengths and weaknesses of methods available for large-scale processing and treatment of hazardous and municipal wastes. Where deficiencies are found, we will develop new processing and treatment methods. Emphasis will be placed on methods that involve cost-effective biological, chemical, and physical treatment processes applicable to both homogeneous and heterogenous hazardous waste streams.

Third, we need to develop management practices to handle the large volumes of solid waste produced by the utility and mining industries. Utility solid wastes include ash residues from combustion of fossil fuels and flue-gas desulfurization sludges as byproducts of sulfur control. Mine wastes include tailings, waste rock, low-grade ore, and overburden.

And fourth, we need to develop and implement a total quality assurance program for all aspects of waste reduction, recovery, reuse, and disposal.

Another priority will be to develop remedies for adverse land disposal impacts at existing landfills. Beginning in fiscal year 1980, our research will determine the quantity and quality of industrial wastes and will subsequently seek to develop industrial process modifications to change the waste characteristics sufficiently to permit environmentally acceptable disposal. We also plan to examine alternatives to landfill for solid and hazardous wastes, including deep well injection, placement in underground mine cavities, and land spreading.

Solid Waste Recovery

Research on better methods for waste reduction, recovery, and reuse has been stimulated in recent years by the combined need for new energy resources and waste disposal techniques that would conserve land and reduce costs in metropolitan areas. Use of waste as fuel has been successful on a small scale. However, much of the equipment presently available has not been specifically designed for use on municipal solid waste, and, thus far, operating experience has been insufficient to provide for the design and selection of optimum equipment configurations. Our research program will attempt to supply the needed data. We will evaluate existing techniques of recovery to see where technology is lacking, and we will develop technology to meet the needs. Our research projects will examine municipal refuse resource recovery operations, including shredding, magnetic separation, air classification, screening, drying, and densification.

We will also evaluate the efficiency and economic viability of materials handling systems such as conveyors, dust collectors, cyclone separators, electrical controls, storage, and fuel retrieval bins as well as systems with provisions for energy recovery. In conjunction with equipment and process evaluations we will study the potential for development of new and marketable products and chemicals that could be produced economically from municipal and industrial solid wastes. The focus will be on developing methods for the biological and chemical conversion of wastes to such things as chemical feed stocks, protein, alcohols, and ammonia.

A second order of emphasis will be research into product and industrial process modifications that can reduce waste generation or enhance recovery and reuse. Basically, we intend to determine if product designs can include provisions for waste reduction by making it easy to separate potentially valuable materials such as certain metals, glass, and paper from other solid wastes. At present, the difficulty of separating the reusable commodities is a major shortcoming of resource recovery.

The study of mining wastes is our third priority in resource recovery research. We will evaluate available methods for extracting resources from mine waste disposal sites. Our work will focus primarily on mining wastes generated in the early 1900s that are still relatively rich in minerals.

NONIONIZING RADIATION

Proliferating communication and radio-navigation systems and electronic devices for home and industry have steadily increased the intensity of public exposure to nonionizing radiation at microwave frequencies. Furthermore, forecasts indicate that the upward trend will continue. While it is well known that intense levels of exposure to nonionizing radiation can produce thermal effects in the form of severe heating within body tissues, little is known about the possible nonthermal human health risks posed by lower levels of exposure. East European countries have standards set at a level which recognizes findings that they attribute to the nonthermal effects of nonionizing radiation. The United States, on the other hand, bases its occupational exposure standard solely on the thermal effects. As a result, allowable exposures in the United States are 1000 times greater than those in the Soviet Union. Research is needed to develop scientifically sound information that can be used to resolve the differences in exposure standards. Further, information is required to assess potential health effects of continuous low-level exposures at frequencies used extensively in our society.

A Health Effects Data Base

The nonionizing radiation research strategy has both a near-term and a long-term plan. In the near term, our research will contribute to decisions on the need to establish environmental guidelines applicable to the general population. Research projects will be devoted to expansion of the current health effects data base to determine whether a standard developed solely from thermal effects data provides adequate protection of public health.

Expansion of the health effects data base will come from two sources: projects that determine the impact of low-level exposures on selected biological processes in animals, and studies of retrospective and prospective epidemiologic data. Activities here will focus on determining health parameters associated with types of nonionizing radiation exposures known or suspected to cause adverse effects. We will continue to conduct experiments on animals using both continuous low- and moderate-level exposures at FM (100 megahertz), near UHF-TV (425 megahertz), microwave (2450 megahertz), and other prevalent frequencies. These experiments will focus on teratologic, hematologic, immunologic, and behavioral effects. To supplement the data from animal experiments, we will attempt to correlate exposure data to adverse human health effects by examining historical epidemiological data on persons who have been occupationally exposed.

Effects on Reproduction, Genetics, and Epidemiology

We also plan to expand our research work into areas of reproduction, genetics, and epidemiology. As part of this effort, we will conduct prospective clinical and epidemiologic studies of people who have just entered a situation of nonionizing radiation exposure and, within limits, will attempt to identify the specific frequencies that have a high potential for causing biological problems.

Other research activities will include studies of the more complex and subtle problems from nonionizing radiation. For example, experiments on rodents will be used to determine the effect that continuous, moderate-level exposures have on lifespan and on potential carcinogenicity. We will also attempt to identify the interactions between nonionizing radiation and ambient environmental factors such as temperature and humidity.

In addition, we plan to investigate the potential for synergistic effects resulting from simultaneous exposure to multiple frequencies. EPA's long-term research will be devoted to an understanding of the functions of biological trigger mechanisms that result from low- and moderate-level exposures. Our first priority will be to refine existing theories on the interaction of nonionizing radiation and biological systems. Here, we will be examining theories which describe how damage occurs to cell membranes and to cell biopolymers (e.g., enzymes and protein) as a result of exposure. We will also examine newly observed phenomena that have the potential to contribute to the development of new theories. Work in this area will be aimed at explaining what "type" of nonionizing radiation can cause cellular damage. Specifically, we will perform studies to determine if frequency and power density combinations known to possess a high interaction potential with one biological system might also have similar high interaction potentials with other biological systems. Additionally, we will examine the interaction of extremely low AM frequency exposures with the central nervous system and identify those frequency bands having a high biological resonant absorption. Finally, our research will attempt to identify exposure effects phenomena in laboratory animals. These research activities will deal primarily with identification of "hot spots" within tissues and organs that are particularly sensitive to nonionizing radiation exposures. We will then develop indices for evaluating the degree of stress induced by such exposures.

GLOBAL POLLUTION

The increasing awareness of the transport of pollutants across political and geographical boundaries and the observable harmful effects from the deposition of pollutants in distant regions has led to the recognition of pollution as a global problem. Close cooperation among many nations will continue to be necessary to resolve the specific problems associated with the spread of pollutants throughout the air and waters of our planet. Presently recognized environmental problems that are truly global in nature include acid precipitation and other effects of intercontinental transport of power plant emissions, ozone depletion in the stratosphere, climatic changes related to increases of carbon dioxide and other aerosol products in the atmosphere, and marine pollution.

The goals of global pollution research efforts are to understand the processes by which pollutant emissions give rise to global impacts, to develop the capability to predict the movements and concentrations of pollutants in the global biosphere, and to identify and assess resulting effects of climate change on human health and welfare.

Long-Range Transport of Pollutants

The effect of long-range transport of air pollution is our highest priority global pollution research area because of the national policy of increased coal consumption, coupled with the development of large power plants and the trend to utilize tall stacks. Energy and environment relationships make this a subject of increasing importance. Understanding how to protect ozone levels in the stratosphere is a secondary research area because, to evaluate the need for and the impact of regulatory options concerning essential uses of fluorocarbons, much more definitive knowledge is needed on the tropospheric and stratospheric behavior of other halocarbons intended as substitutes, and on the effects of increased ultraviolet band radiation.

Climate

The effects of climate change and variability will be given less emphasis because such changes are so little understood that meaningful socioeconomic assessments cannot yet be undertaken. Climatologists need to provide quantitative information on the characteristics of expected regional and global climate changes (wind patterns, temperatures, precipitation, and seasonal and annual weather patterns, etc.).

Marine Pollution

Marine pollution will also be researched at a relatively moderate level because severe episodes are local in character, and the time required for localized pollution to affect the global marine environment is relatively long compared with atmospheric transport times. Marine pollution problems are also not so readily identified or so urgent as those of air pollution.

National Aeronautics and Space Administration

The National Aeronautics and Space Administration (NASA) conducts a program of research and development aimed at contributing to the nation's broad objectives in aeronautics and space as defined in the Space Act of 1958. In addition, NASA expertise is being applied to such other problems as developing technologies to support national energy policy.

Philosophically, NASA's programs are shaped and guided by two overriding and governing criteria—usefulness to human life and intellectual importance. Simply stated, intellectual importance eventually translates to long-term usefulness. Within that philosophical framework, NASA efforts in space and aeronautics comprise four general categories or mission thrusts:

- Scientific exploration of the solar system to improve our understanding of our planet and its phenomena, coupled with observation of the universe in the continuing effort to understand the place of Earth and life in the cosmos;
- Investigation of the Sun-Earth relationships so basic to the whole biosystem in which we live;
- Use of near-Earth space for global studies and monitoring of the Earth, its resources, and its atmosphere, communications, and other purposes directly beneficial to human welfare; and
- Generation of advances in aeronautics technology that support the development of more economical, efficient, and environmentally acceptable air transportation systems, the continuation of the nation's competitive position in the international aviation marketplace, and the maintenance of superiority in the nation's military aircraft.

The following eight sections discuss science and technology needs and opportunities related to NASA's mission thrusts. They also describe the way in which the mission thrusts will be made possible and practical by the Space Transportation System and the emerging capability for acquiring, handling, and analyzing massive amounts of data.

BASIC RESEARCH

As required by the Space Act, one of NASA's principal missions is the conduct of basic research relevant to aeronautical or space activities. NASA recognizes the value of the foundation that basic research provides to all of its missions. Examples include applied mathematics and computer science, including computer-human interactions and artificial intelligence; the fundamental understanding of matter and natural processes needed in the development of multipurpose sensing instruments and measurement techniques; pioneering research in the life sciences, ranging from human to primordial; and basic investigations in fluid mechanics essential in the design of efficient, quiet, and nonpolluting aircraft.

SPACE AND TERRESTRIAL APPLICATIONS

The current Space and Terrestrial Applications Program has three principal elements: observe and study the Earth from space to understand and forecast environmental behavior, assess the productivity of the Earth's surface for both renewable and nonrenewable resources, and develop a "standard" model of the dynamic Earth; pursue selected technology development to provide more effective, efficient, and higher capacity communications systems using satellites; and demonstrate to the scientific and commercial user communities the capabilities of the space environment for enhanced control of materials processing. Collectively, these program elements seek to utilize NASA space capability to address the many challenging problems related to human welfare and to the improvement of the quality of life on Earth.

The benefits to humanity are already manifest worldwide, and the potential for further benefits is still unfolding. Data sensed from space can provide global information regularly, and can do so more rapidly than other methods. Communications can be conducted over almost a hemisphere at a cost that is independent of the distance between communicators. Space flight is the only method for obtaining weightlessness over extended periods of time, and it provides the additional advantage of a high vacuum for materials processing needs. Extensive cooperation is required among Federal, state, and local government agencies, private industry, and universities in developing solutions to problems and in transferring the resulting techniques, technology, and information to decisionmaking officials.

Observing the Earth's Environment From Space

In the last several years it has been successfully demonstrated that space observations and techniques can make substantial contributions to understanding processes related to communications using satellites, weather forecasting, climatic fluctuations, manmade changes in the environment, tracking of severe storms, and motions of Earth's crust. Space techniques also have potential for assessment of agricultural productivity and land use changes.

The Environmental Observation Program applies space observations and techniques to aid in the ability to understand and forecast environmental behavior. The scale of this behavior includes the extremes: global to regional—from global weather, climate, and ocean circulation to regional severe storms and air and water quality. The program encompasses and integrates the diverse scientific fields of meteorology, climatology, atmospheric chemistry and physics, and oceanography. A result could be development and demonstration of key aspects of a global environmental data acquisition and processing system.

The Earth's environment is a complex system. This complexity is largely caused by the interactive processes among life, the atmosphere, the land, and the oceans. These basic processes must be understood before the crucial factors that constitute and adversely affect the environment can be assessed and delineated.

In the next five years, the major program thrusts in the Environmental Observation Program will be to study jointly with the Department of Defense (DOD) and the National Oceanic and Atmospheric Administration (NOAA) a national oceanic satellite system for limited operational demonstration of ocean monitoring capability; to develop jointly with NOAA the next generation of weather satellites, to be shuttle-compatible and to assure continuity of NOAA meteorological operations; and to participate in the National Climate Program by exploring expanded use of space for climate observations.

Observing Earth Resources From Space

The Resource Observation Program develops and demonstrates the application of space observations and space techniques in meeting national and global needs for improved management of food, fiber, water, and land resources; for improved effectiveness of mineral and energy resources exploration; and for understanding the dynamic characteristics of the Earth, including processes related to earthquakes.

In the next five years, the Resource Observation Program will conduct research jointly with the U.S. Department of Agriculture on using data acquired from space to improve production of major agricultural commodities in primary producing areas. The program will also expand space capability for geological delineation of surface features for resource assessment; develop, over a period of 10 to 20 years, models of the contemporary motion of major tectonic plates; model crustal strain accumulation in seismically active areas; and identify causative relationships for changes in polar motion and Earth rotation.

Satellite Communications

NASA's new directions in satellite communications technology development respond to a growing need for new and improved commercial services. In the next five years, NASA will pursue selected research in advanced communications technology.

Projections of the demand for satellite communications indicate potential saturation of the geosynchronous orbit-spectrum space in useful view of the United States. Users of point-to-point and broadcast services would then strongly compete for orbit and spectrum space. Predicted saturation coupled with continued demand growth leads to a requirement for new technology to effectively use or reuse more of the limited orbit and spectrum resources. The K-band (30/20 GHz) was allocated in 1971 to increase satellite communications capacity, but it has been inadequately explored and developed for use in the

United States. Meanwhile, flight R&D programs in foreign nations are proceeding with developments of technology that could reduce U.S. participation in a growing worldwide market.

As in the past, NASA R&D activity can significantly advance the state of the art for K-band technology to meet the expanding communications demand and provide U.S. industry with the technology to meet competitive requirements in a world market. Because the private sector's R&D must be based on short-term profit potentials, the government must accomplish the long-term basic R&D that will ultimately enable industry to confidently offer new and expanded services.

Materials Processing in Space

This program emphasizes the fundamental science and technology of processing materials under conditions that allow detailed examination of the constraints imposed by gravitational forces. The studies are directed toward certain selected materials and processes which will best elucidate the limitations due to gravity as well as demonstrate the enhanced process control that may be possible in weightless environments. One activity that will be emphasized during the next five years will be consolidation of the data and supporting studies associated with the specific areas of materials science and technology being addressed in the initial space flight investigations. Experiments will be initiated in materials processes using shuttle-spacelab capabilities. Emphasis will also be placed on the continued identification and understanding of design-driving experimental requirements together with initiation of technology development studies to reduce the risks associated with future hardware development.

SPACE SCIENCE

The Space Science Program is important both for the long-range benefits it can be expected to provide to life on Earth and for the challenge and stimulation it provides to the human spirit and intellect.

The program has four elements—the Planetary Program, the Astrophysics Program, the Solar Terrestrial Program, and the Life Sciences Program. These elements collectively seek to determine the origin and history of the universe and the physical and chemical processes occurring in and governing it; the origin, history, and distribution of life in the universe; the structure and dynamics of the Sun-Earth space system, including the magnetosphere, ionosphere, and upper atmosphere and the physical and chemical processes that shape and control the system; and the best uses of space as a laboratory to study physical, chemical, and biological processes and to provide knowledge for improving conditions on Earth.

Realization of the potential benefits of the Space Science Program depends upon not only direct support of these activities but also the development and operational deployment of the space shuttle, spacelab, and other space transportation system elements. Tracking and data handling capabilities must be upgraded to meet orbital and deep-space science requirements. Also, looking to future missions that will challenge the technological state of the art, space research and technology activities must pursue the new technologies emerging from the laboratory and bring them to such a state of readiness that space system developers can exploit their benefits.

The Planetary Program

The objectives of planetary exploration are to extend our understanding of the formation and evolution of the solar system through study of solar system bodies and the interplanetary medium, and to apply that understanding to continual assessments of past, present, and future processes on Earth. Our strategy for accomplishing these objectives is built around reconnaissance missions to seek and identify major characteristics of the planets; exploration missions to discover and understand the processes, history, and evolution of planets; and intensive study missions focused on specific problems of high importance.

The Astrophysics Program

The Astrophysics Program consists of scientific inquiries into questions that have bewildered and inspired humans since our primitive beginnings. What is the nature of the universe? How did it begin, and what is its eventual fate? The philosophical motives for probing the universe are compelling, and the capability that the vast laboratory of space provides for studying and, it is hoped, understanding processes in the universe cannot be duplicated in Earth-based laboratories. For example, if quasars are as remote as their spectral shifts seem to indicate, they are emitting energy at a rate that cannot be explained even by nuclear processes as we now understand them. Studies of such phenomena are at the very forefront of science. While the benefits to humanity cannot always be predicted, the opportunities cannot be ignored.

The astrophysics strategy is first to identify sources and new phenomena with ground-based observatories, aircraft such as the C-141 Kuiper Observatory, balloons, sounding rockets, free-flying satellites, and other space facilities, as appropriate. Then sources will be catalogued by making "all-sky" surveys in all wavelengths with systems such as the High Energy Astronomy Observatory (HEAO-A) and the Infrared Astronomy Satellite (IRAS). Identified sources will then be studied with high-resolution observatories such as HEAO-B and the Space Telescope. Last, correlative studies of the sources will be made in all appropriate wavelengths with free-flying satellites and multidisciplinary shuttle-spacelab missions in order to develop a complete understanding of the phenomena and objects involved. In addition, continuing studies in such areas as relativity will exploit the unique characteristics of space as a physics laboratory. The knowledge gained from all of those activities will improve our general theories about the nature and origin of the universe.

The Solar Terrestrial Program

The objectives are to understand our local star, the Sun, as an astronomical body, as the source of Earth's light and heat, and as the major source of the Earth-space (magnetosphere, ionosphere, and upper atmosphere) system; to understand the form, character, and variability of the Earth-space system; to understand the physical processes that shape, control, and link the individual elements of the system; and, eventually, to predict the potential effects of naturally occurring and human-caused perturbations on the Earth-space system. These objectives are becoming increasingly important for a resource-limited world whose ability to feed and shelter its inhabitants may be reduced by small changes in the climate and whose environmental quality may be degraded, perhaps irreversibly, by humanity's activities.

The strategy employed by the Solar Terrestrial Program includes remote observations of the Sun and the various manifestations of the solar cycle. Pioneering exploration of the structure and dynamics of the solar wind, extended magnetic fields, and cosmic rays is also included, as is a multifaceted attack on the mysteries of Earth's complex neutral, charged-particle, and magnetic-field environments and their interfaces. These activities involve in situ observations, a developing opportunity for active experiments, ground-based measurements, and a comprehensive program of theory and modeling. In addition, we conduct research into the magnetospheres of other planets in the solar system.

The Life Sciences Program

The objectives of the Life Sciences Program are to ensure human health, safety, well-being, and effective performance in space flight: to utilize the space environment and space technology to advance knowledge in medicine and biology; and to understand the origin and distribution of life in the universe.

Past life sciences flight observations have focused on evaluating possible hazards to crew health and on identifying factors that could impair crew performance. While those activities will continue to be important, the shuttle-spacelab will enable us to conduct a greater number of types of flight experiments. The flight experiments will exploit the unique zero-gravity aspect of the space environment to study the fundamental responses of humans and other organisms to gravity, and will develop ways to apply the resulting knowledge to extend the duration of safe manned space flight.

SPACE TRANSPORTATION

The nation's space program is evolutionary. Scientific discoveries and technological progress produce new opportunities and requirements that lead to new capabilities to meet national needs and external challenges. The Space Transportation System Program focuses on providing the new capabilities needed for efficient, economic space transportation and related space services. It has two basic goals.

Vehicle Development

The first goal is to provide easy, low-cost access to, from, and within space for the payloads and systems that NASA and other users, both public and private, develop to achieve their space objectives. The primary vehicle for satisfying that goal into the 1990s will be the Space Transportation System (STS), consisting of the space shuttle, spacelab, and upper stages.

The space shuttle and spacelab will promote increased use of the unique features of space both for potential applications that are understood today and for applications not yet conceived. Consequently, a key objective of the Space Transportation System Program is to define a feasible evolutionary approach to satisfying future requirements for space systems and operations and to focus on more advanced elements of the STS. These include the solar electric propulsion system and the orbital transfer vehicle.

Support Systems

The second goal is to develop support systems that have the potential of enhancing the STS's unique capabilities and, therefore, its role in the use of space for the benefit of humanity and in the maintenance of U.S. leadership in space. Operations of the space shuttle and spacelab will generate requirements for new space services and operations. Possibly the most pressing requirement will be for development of free-flying and shuttle-attached power packages to provide electrical power in the 25-kilowatt range to support longer shuttle missions and a host of independent applications and science missions. These packages could form the basis for larger and higher power platforms, which could then provide logistic support for potential space operations, e.g., space manufacturing.

These advanced systems could also serve as staging platforms for operating and servicing large Earth-orbiting spacecraft such as large astronomical telescopes, solar-terrestrial observatories, and advanced communication systems. In addition, space manufacturing and processing techniques and systems will require levels of electrical power and times in orbit not available with the basic space shuttle. The objectives will then also include the development of remote maneuvering, teleoperators, structure fabrication devices, habitats, and waystations.

Economic Uses

Use of space for economic and useful purposes can create economic new markets and products and enhance the national economy, as well as maintain the leadership of the United States in space capability. The role of the government in expanding the use of space for economic purposes is one of identifying promising initiatives and carrying out the activities necessary to establish feasibility. When feasibility has been demonstrated, industry can then invest private capital for space ventures. Satellite communications systems and the myriad of space technology spinoffs are past examples of industry's commercialization of products stemming from government support for early space capabilities.

Continued progress toward space goals will depend on successful development and timely achievement of operational status of the Space Transportation System, and its future growth. Adequate funding must be maintained for current development projects, supporting facilities, and complementary systems. Studies and advanced technology programs needed for future capability extensions must be initiated with appropriate lead time and conducted thoroughly so that new system development decisions can be informed and timely. In addition, the needed new elements of the STS must be developed so that they support the foreseen applications in a timely and economical fashion.

SPACE TRACKING AND DATA SYSTEMS

Today's tracking and data acquisition network facilities are the result of a complex, interrelated technological and programmatic evolution. We developed our first network to track the first U.S. satellite as part of the Vanguard program.

The decision to put a man on the Moon before the end of the 1960s led to development of the specialized Manned Space Flight Network (MSFN) to support that national effort. Another entirely different network became necessary when we started sending automated spacecraft beyond Earth orbit—the Deep Space Network (DSN).

The goal of the Space Tracking and Data Systems program is to develop, implement, and operate the tracking and data acquisition systems required for sophisticated and complex flight missions of manned and unmanned orbital spacecraft, sounding rockets, deep-space probes, and research aircraft.

Data Relay in Earth-Orbiting Missions

For Earth-orbiting missions, one continuing objective is to increase mission coverage while reducing operational costs. The Tracking and Data Relay Satellite System (TDRSS) will take a significant step toward meeting that objective when it begins operation in 1980. It will provide a sixfold increase in mission coverage for all satellites in low Earth orbit. It also will provide launch, landing, and nearly continuous orbital support to the space shuttle. Plans are to phase out most of the existing ground-station network as TDRSS becomes operational in order to stabilize the overall cost of supporting the tracking and data acquisition requirements of Earth-orbiting missions.

Another major objective is to provide the optical tracking capability required by Earth-dynamics and ocean-dynamics projects such as Lageos, Geos, and their successors. Investigations of geodesy, tectonics, and sea state require precise determination of orbits to ensure that onboard sensors collect useful data. The required precision will be provided by portable laser-ranging stations that NASA has developed and that have an accuracy of about 10 centimeters.

Data Relay in Planetary and Deep Space Missions

Our principal objective for planetary and deep space missions is to improve their navigational accuracy and telemetry data return to maximize the scientific value of the missions. Our continuing program is increasing antenna sensitivity through improvements and modifications that reduce system noise and increase receiver capability. Navigational accuracy is being improved by new tracking techniques such as dual-station tracking. Use of higher frequencies inherently increases telemetry data rates. The current conversion of the Deep Space Network will allow greater data return and ease the saturation of the network's antennas.

Increases in antenna aperture—provided by either a large, single antenna or an array—create opportunities for quantum jumps in science return. The NASA Space Science Program is emphasizing the imaging and radar mapping of planets and their satellites. Flights to study the distant outer planets and their rings and satellites are of long duration and are characterized by brief, intensive flyby encounters and orbiting missions, both of which require high data rates. Obtaining maximum return from missions planned for the 1980s will require an advanced, groundbased, large-aperture antenna system. For missions beyond the 1980s, NASA is contemplating an orbiting deep space relay station to provide new capabilities. We will investigate the system's promise thoroughly through systematic study and technology development.

Advances in Data Processing and Communications

The increasing data rates discussed above will be accompanied by comparable increases in data processing workloads. The shuttle era will require an order-of-magnitude growth in data processing capabilities. Processing loads from spacelab data alone will exceed current capabilities. Consequently, new data processing facilities and equipment are needed. The combined effect of new modes of operation—such as interactive, experimenter-controlled operations—and the sophistication of advanced generations of sensors will create increases in data output. The increases will require both the use of even more advanced techniques for data handling and management and the use of new technologies and equipment. Only by those means will we be able to ensure the greatest return of scientific information for the lowest cost.

The advent of adaptive missions, in which activities are modified to reflect knowledge gained in earlier activities, has created the objective of providing the real time capabilities and flexibility demanded by the mode of operation. The increased on-line data flow required to support adaptive missions will be accommodated by increases in the bandwidth of communication links and by use of advanced techniques, such as shared computer resources, to provide for real time data and control.

SPACE TECHNOLOGY

Investments in technology are beneficial and effective in enhancing current and planned missions and in enabling missions that require capabilities beyond the anticipated state of technology.

The principal objective of the Space Technology Program is to generate advanced technology for application to future space projects and flight missions. The program is designed to support and enhance approved programs, to provide technology options for planned near-term programs, and to enable missions not yet conceived.

Planning for the Space Technology Program derives from missions identified by the flight program offices as well as representative missions for the 1980 to 2000 period, which will require technological advances. Technology requirements of those missions are structured in three major thrusts: information systems, spacecraft systems, and transportation systems.

Information Systems

In information systems, the thrust is to develop capabilities for the acquisition, processing, and dissemination of data in a form responsive to specific user needs. The objectives are to provide broader synoptic coverage, wider spectral range, more rapid access to useful information, and, ultimately, lower cost systems. User needs are derived from the science and applications missions of the Agency. Major elements are sensing and data acquisition, sensor support systems, communications, data reduction, and data dissemination.

Spacecraft Systems

The thrust in spacecraft systems includes technology in structures, materials, guidance, navigation and control, automated systems, power, onboard propulsion, and planetary entry. Much of this thrust is concentrated on technology for large space systems that are expected to require new techniques and components to enable their on-orbit construction or deployment, new long-life materials, distributed control systems, and novel concepts for providing shared, onboard support services.

Transportation Systems

The transportation systems objective is to provide the technology required for an enhanced space transportation system. The current program includes technology work in structures, materials, chemical and electric propulsion, and aerothermodynamics; it is supported by shuttle flight experiments in both the orbiter experiments (OEX) and the spacelab programs.

AERONAUTICS

Aviation in the United States is the principal mode of intercity public transportation, a major element in national defense, and a source of economic strength. The United States dominates the free-world marketplace in transport and military aircraft and holds a leading position in other types of aircraft. Continued preeminence in aviation will demand advances in the nation's aircraft and system capabilities, in the productivity and profitability of the U.S. aviation industry, and in the environmental acceptability of U.S. aircraft. The maintenance of U.S. technological leadership is, therefore, a matter of national concern and importance.

During the next five years, major research and technology (R&T) objectives of the Aeronautics Program will include completion of the Aircraft Energy Efficiency Program; further reductions in aircraft environmental impact; continued programs in rotorcraft technology and in avionics, human factors, safety, and terminal area operations; further progress toward validation of supersonic cruise technology; continued contributions to the technology needs of military aviation; and further strengthening of the generic R&T programs in the several fundamental disciplines upon which future achievements will be based.

The potential benefits from continued U.S. technological leadership in aeronautics are of major importance to the nation. NASA's program will strive for a balanced combination of in-house and university research, industry contract activities, interagency cooperative efforts, and the regulatory process.

Civil Air Transportation

Research and technology activities in NASA's Aeronautics Program are shaped by perceived technology needs and opportunities. With respect to civil air transportation, current and projected needs are for safer, more economical, efficient, fuel-conservative, and environmentally acceptable aircraft. These objectives apply to subsonic transports, both large and small, and to supersonic transports. Corresponding NASA R&T activities include experimental investigations of advanced engine components and alternative fuels, wind-tunnel and flight research on improved airfoil sections and wing-platform configurations, demonstration of active control systems and electronic cockpit displays, and validation of composite primary aircraft structures. A substantial portion of this effort, directed at providing the technology for large improvements in transport aircraft energy efficiency, is conducted in close cooperation with the transport manufacturing and operating industries. Special efforts of a relatively long-term nature include supersonic cruise research, in which the emphasis is on aerodynamic configurations and titanium structures, variable-cycle engines for civil and military supersonic cruise aircraft, and laminar flow control to reduce drag.

General Aviation and Rotorcraft

In addition, NASA R&T is focused on the particular needs of general aviation and rotorcraft. Growth rates in these types of aviation are greater than for commercial transports, and world market competition is especially active. In general aviation, the major technology needs are in efficiency, noise reduction, safety, and utility. NASA R&T emphasizes small engines, including internal combustion engines; propellers; stall-spin prevention; single-pilot instrument operations; and aerial applications of agricultural materials. In rotorcraft, the objectives of the technology program are to improve efficiency, reduce vibration and noise, lower maintenance costs, and remove operational constraints due to weather.

Military Aviation

With respect to military aviation, NASA works closely with the Department of Defense in identifying needs and planning research and technology activities, many of which are jointly conducted. Primary emphasis is on the integration of flight and engine controls, flight testing of advanced configurations, and the development of a broad data base for use in future military aircraft design.

ENERGY SYSTEMS

NASA is authorized to conduct programs in support of the nation's energy research and development needs. This authorization is contained in the Energy Reorganization Act of 1974, the Department of Energy (DOE) Organization Act of 1977, and the National Aeronautics and Space Administration Act of 1958, as amended. The DOE-NASA Memorandum of Understanding, dated March 21, 1978, describes the general conditions under which DOE-NASA cooperative efforts will be formulated and conducted.

The NASA program strategy is to utilize NASA aeronautics and space technologies, experience, and facilities in support of the energy research, development, and demonstration program needs of DOE and other government organizations. The overall goal of the program is the effective use of NASA capabilities in the accomplishment of specific technical and programmatic goals resulting from national energy policy.

Current programs using both NASA funds and funds from other agencies encompass three major areas of activity: space utilization, which includes nuclear waste management and satellite power systems; solar terrestrial applications, which include solar heating and cooling, photovoltaic conversion, wind turbogenerators, solar thermal energy conversion, and energy storage; and conservation and fossil energy applications, which include advanced ground propulsion, advanced coal extraction, industrial gas turbines, and phosphoric acid fuel cell systems. Additional applications of NASA technology, expertise, and facilities to energy problems can be expected.

Significant benefits to the United States in this critical area can be anticipated in the next few years when the technologies that are being developed are moved to the marketplace and begin to provide alternative supplies or energy conversion techniques. A close working relationship between NASA and the other agencies involved is essential if these benefits are to be realized.

National Science Foundation

The National Science Foundation (NSF) promotes the progress of science in the United States through the support of research and science education. Unique among Federal agencies in its dedication to basic research and its responsibility for future scientific research capability, the Foundation seeks to press fundamental scientific inquiry on all fronts.

The broad support of research is especially important because it promotes the germination of unexpected discoveries and breakthroughs. Such unanticipated developments frequently lead to advances in applied science and in technology that in turn are useful in addressing the important practical problems our society must deal with. The connection between basic research and useful application is sometimes long and tenuous. Nevertheless, the discussion of problems and opportunities relating to science and technology must not overlook the need to provide for the continuing development of science in general.

In this report, NSF has chosen to describe 11 special areas of basic research in which advances over the next five years can be expected to contribute to the understanding and amelioration of important national problems. These particular topics have been selected because the information and insights resulting from long-term fundamental research on them would improve the basis for national policy decisions and directions and enable concerned governmental and private sector organizations to carry out their responsibilities more effectively. The descriptions for each of the topic areas characterize the national concern, point out scientific opportunities, and mention the readiness of the research community to take advantage of the opportunities. There are, of course, many additional topics that could have been emphasized. Among these are important matters of continuing concern. For example, maintaining the supply of scientifically trained people is of particular interest to NSF. There continue to be challenges to our ability to identify, motivate, educate and employ wisely the human scientific resources needed to keep our scientific enterprise healthy.

In the course of selecting the research topics, another type of concern became clear—the need to stimulate and encourage a much higher level of scientific and technological literacy throughout the general public. The nation's ability to deal with such issues as nuclear reactor development, toxic chemicals, future space programs, world population growth, and new medical technologies requires an educated public that understands the pros and cons of various actions as well as a large number of educated decisionmakers who can weigh scientific factors with confidence to reach responsible and wise decisions.

A higher level of scientific and technological literacy would contribute to

- Better preparation of all segments of society to deal with technological developments that affect everyone's lives;
- A higher level of competence by workers in high technology fields such as electronics, communications, and nuclear power;

- More adequate understanding by decisionmakers of the merits and limitations of scientific information available to them; and
- Better understanding of risks facing society and of the consequences of major decisions.

One only need contemplate the difficulties and ramifications of such matters as SALT, "test-tube babies," development of alternative energy sources, and the Three Mile Island incident to realize the importance of this topic to the nation's future.

NEW APPROACHES TO MODIFICATION OF PLANTS AND INCREASED CROP PRODUCTIVITY

The increasing world population and the rising expectations of nutritionally deprived people emphasize that the quantity and quality of economic crops must be significantly increased. Another urgent reason to increase crop productivity is the inevitable reliance on biomass to replace waning petroleum sources and dwindling material resources of other kinds. Research in this area could also have an impact on materials available from forest, agriculture, and aquatic systems that are in demand in building, paper, textiles, and chemical feed stocks.

Research opportunities abound. Recent advances in plant tissue culture and recombinant DNA research provide the basis for a foreseeable new technology in which plants can be modified and desirable new traits introduced. Genetic techniques can be used to improve drought tolerance, winter hardiness, and nitrogen fixation. As this field develops, the time scale in which these techniques will be applicable will be much shorter than the long-term (multiyear) breeding program of conventional agriculture. Another important focus might be on the study of phosphate metabolism of plants to decrease dependence on large applications of phosphate fertilizer.

The development of salt-tolerant plants is another research area of promise and readiness. Saline water is a pervasive problem, and the increasing salinity of irrigation water is decreasing the land available for agricultural production. Research needs include work on membrane structure and function in salt-tolerant plants, temperature effects on water content in cells, symbiotic relationships between plants and microorganisms, effects of increasing saline water on osmosis in plants, uptake of toxic ions and the effect on plant nutrients, and methods for removing excess salts from plants.

An ecological approach to the design and management of crop and livestock systems also holds promise of increasing the efficiency of food production. Many opportunities exist for the reintroduction of ecological controls to agroecosystems, particularly to achieve control of pest and disease organisms and to restore processes of nutrient retention and regeneration on food and fiber producing lands. Efficiency may also be improved by the development of crop and livestock combinations or rotation sequences that optimize yields of food and fiber with respect to energy and nutrient subsidies. Ecological engineering can also be used to improve the adjustment of water and nutrient supplies to plant and animal needs during the production cycle. These approaches are essential to provide a framework for incorporating genetically improved stocks into an efficient agroecosystem management plan.

PHOTOCHEMISTRY, PHOTOBIOLOGY, AND PHOTOPHYSICS

The transformation of solar energy by plants (photosynthesis) is the start of the process that eventually leads—over eons—to fossil fuels. Existing photosynthetic processes are obviously much too slow to provide immediate energy supplies. Among the most promising avenues for increasing energy sources are improvements in efficiencies of biological photosynthetic activity and the construction of artificial systems to increase the capture and conversion of light energy.

The use of biological photosynthesis to generate convenient energy resources is limited at present by an incomplete understanding of the relevant chemical and biological processes. Basic knowledge in such fields as chemistry, enzymology, and plant physiology is needed to understand the mechanisms by which plants and bacteria collect, store, and convert or release the energy of the sun. A complementary and promising approach to this task is the use of new genetic techniques in which the selection and stock maintenance of hybrid plants with increased photosynthetic efficiency is pursued.

Opportunities are also available within the fields of photochemistry and photophysics. Of particular interest is the effort to use light-induced reactions to produce high-energy, unstable materials that can be reconverted to their original forms under controlled conditions. During this process stored energy is released in a usable form. Another objective is the conversion of light energy directly to electrical power through either photovoltaic or photogalvanic cells. Progress along each of these routes is expected to be aided over the next five years by results from basic research. Knowledge of organic and inorganic sensitizers and catalysts for photochemical reactions is expected to compensate for the low solar radiation sensitivity present in key materials and to provide key binding sites to promote low-energy reaction pathways. Basic research in new photoreactions which may lead to the production of high-energy materials is also of major importance.

Difficulties to be overcome in the photo-assisted electrolysis of water to produce hydrogen for fuel include poor wavelength response, low quantum efficiency, high electrode expense, and uneven stability. In photogalvanic cell development a major difficulty is the inadequate understanding of interfacial phenomena. The construction of boundaries such as solid-liquid interfaces or membranes may offer the means for the prevention of energy-wasting back reactions. Recent advances in instrumentation promise rapid progress in this area. The development of new materials systems, e.g., amorphous semiconductors and organometallics, demonstrates long-term potential for the achievement of low-cost photovoltaic conversion using noncritical materials. Other new materials systems of interest include hydrogen in metals and superionics.

Recent advances in theory and instrumentation are yielding new levels of understanding of photochemical processes. Laser spectroscopy, in particular, has permitted the detection and characterization of short-lived species occurring as transients in light-induced reactions. New spectroscopic techniques have markedly increased capabilities to obtain knowledge of the structure and function of enzymes, surfaces, membranes, and other composite molecular assemblies involved in photochemical, photophysical, and photobiological processes. Research in engineering is directed toward the provision of the framework required to ensure future implementation of feasible energy conversion techniques and the achievement of a balanced and comprehensive examination of potentially promising long-term approaches.

CHEMISTRY OF BRAIN AND BEHAVIORAL PROBLEMS

Better understanding of the chemistry of the nervous system can help in dealing with behavioral problems that lead to obesity, drug and alcohol abuse, and such disorders as schizophrenia, senility, and mental retardation.

Recent advances in immunocytochemistry have allowed us to gain increasing knowledge about transmitter systems, which has advanced our understanding of the fundamentals of nerve interaction. By raising antisera to various molecules, labeling them radioactively, and injecting them into cells, various molecule systems can be traced by identifying the individual cells containing these molecules by the fluorescence they produce in the light microscope.

Research needs over the next five years include: biochemical centers to expand the production of antisera to be used as markers so that many more molecular systems can be explored, neurochemical studies addressing changes in brain neurotransmitting systems associated with various behaviors, improvement in biochemical techniques so that less of a substance is needed to determine its composition and structure, and improvement in purification techniques to aid in the isolation and purification of receptor molecules that activate a neuron after it receives an electrochemical signal from another neuron.

Very little is known about basic molecular mechanisms that mediate behavior. Research must address cellular and subcellular mechanisms pertaining to behavior. The effects of hormones on the development and maintenance of the nervous system from the embryo through senescence must also be better understood.

Separating out the chemical components of the nervous system and developing an understanding of how they interact with the structural components is a major effort. Numerous diseases of the nervous system are due to known chemical imbalances with a genetic origin. Behavioral disorders due to simple chemical imbalance should continue to reveal themselves with the constant improvement of noninvasive techniques that may allow chemical analysis of human tissue in the future. More innovative techniques are needed to bridge the narrowing gap between chemical and anatomical approaches to nervous system function.

The neuroscience research community nearly doubled between 1975 and 1978 (to about 10,000 active scientists), and there is an abundance of well-trained, highly sophisticated researchers with interdisciplinary backgrounds.

POLITICAL ECONOMY OF CRITICAL RESOURCES

Economics and political processes that influence the price and availability of critical resources in international markets are not well understood, nor are the economic and political effects of changes in the price and supply of such resources. This lack of knowledge limits the adequacy of economic projections and the design of effective government policies.

Recent events have dramatized the impact of political as well as economic forces on the price and supply of critical resources: world oil prices have quadrupled, world capital is undergoing a shift favoring the OPEC countries, other blocs are likely to follow the OPEC example, and economic forecasts have been rendered inaccurate by unforeseen shifts in the price of oil. The supply and price of oil and other natural resources necessary to meet economic growth and national security objectives are subject to rapid change, and effective responses to impending price shifts and international "commodity politics" depend on inadequately developed knowledge. Economic forecasting, important in dealing with inflation and unemployment, is currently based on an inadequate understanding of international economic and political processes.

There is an immediate need for, and capability to conduct, additional research that emphasizes:

(1) Models of optimal exploration and production of nonrenewable resources, and the impact of governmental action in such models;

(2) Incorporating bargaining models, coalition strategies, and international economic factors into the analysis of cartel behavior to determine the extent to which observed market action deviates from optimal cartel behavior and to project cartel response to modified economic and political conditions;

(3) Use of techniques adapted from decision theory, control theory, game theory, and the economics of arms races in the analysis of strategy and bargaining among blocs of national governments, with the objective of ascertaining the welfare effects of bloc action and optimal rules for policy intervention, negotiation, and conflict resc. ·ion;

(4) Research on the effectiveness of international marketing arrangements and agreements (e.g., buffer stocks) to smooth out fluctuations in prices, and exploration of the political and economic conditions that influence their effectiveness; and

(5) Improvement of models for projecting world and domestic supply and demand for critical resources by incorporating international political factors along with selected sets of realistic assumptions about new sources of supply, resource substitution, shifting technological requirements, and market conditions.

The research outlined above would assist in developing more accurate short-term economic projections and forecasts of the prices of critical resources, more adequate long-term forecasts of resource flows and prices (with implications for planning), and more effective policies relating to the production of critical resources, international trade, and international negotiations.

BIOGEOCHEMICAL CYCLING

An understanding of biogeochemical cycles is crucial to the management and maintenance of our Earth's ecosystems. Studies of many elements have been conducted, but the most concern has developed about carbon (C), nitrogen (N), sulfur (S), and phosphorus (P). Major perturbations of the composition of the atmosphere and biosphere have resulted from society's use of ancient deposits containing C, S, and, to some degree, N (e.g., increased fossil fuel use, more intensive use of fertilizers, increased destruction of tropical forests, and increased use of marginal lands for grazing and subsistence agriculture). Byproducts of their use are being released at increasing rates into modern environments, causing severe disturbances. We need accurate information about these elements everywhere within the boundaries of our biosphere and atmosphere—in ocean depths as well as at the limits of the stratosphere.

Three global problems of immense magnitude are identifiable: (1) increased ozone-layer interactions and possible ozone depletion because of penetration into the upper stratosphere of N_2O and chlorofluoromethanes that react with ozone under the influence of ultraviolet light; (2) increased CO_2 levels in the atmosphere which, if continued, might lead to a perceptible warming of the earth's surface with consequent changes in patterns of world climate; and (3) increased acid rain, which is injurious to most freshwater ecosystems and could possibly harm crops, people, buildings, and machinery. These major problems overshadow many other environmental events. They also may lead to such related problems as loss of agricultural production, increased rates of desertification, destruction of water quality, and general lowering of the quality of life.

We need to know background levels in natural ecosystems in order to assess the principal perturbing effects. Thus, there is need for a long-term intensive study of the cycles of critical elements and the interrelationships among them in biota, atmosphere, and terrestrial and aquatic systems of the world. Further long-term studies of critical communities in selected biome regions of the world should be undertaken.

One means is through the development of sites at which long-term ecological research projects may take place, measuring key abiotic and biotic parameters by standardized methods for long periods of time. With such work there is hope that scientific research will be able to provide the information necessary to determine the type and magnitude of effects on the biosphere and atmosphere.

CERAMICS, POLYMERS, AND OTHER NONMINERAL MATERIALS

The high level of consumption of a large variety of materials in modern society results in a reduction of the availability of natural resources from which they are derived. To the concern for the diminution of the stock of natural resources, there have been added the related concerns of the various costs of materials exploitation and production (such as energy and transportation costs) and associated health, safety, and environmental regulations; availability of foreign sources of supply; and domestic industrial viability. The result has been a combination of social, political, and economic pressures to conserve, substitute, and recycle materials. In the area of substitution, basic research offers a particular opportunity to improve the understanding of the relationships among the structure, composition, and properties of materials.

Two principal objectives have been identified that will allow expansion of the supply of products and processes. One is the design of materials with longer useful lifetimes and which can be recycled or remanufactured, e.g., aluminum beverage containers. The second is the development of substitute materials for nonrenewable, energy-intensive, or scarce materials, e.g., high-strength polymers to replace metals.

The pursuit of these objectives can be implemented through the conduct of long-term basic research in the following areas, among others: ceramics (to replace high-temperature metallic alloys); high-strength polymers (to replace metals); synthesis of polymers from nonpetroleum, renewable resources; and toughening the surfaces of low-cost amorphous materials with small amounts of special materials to enhance durability.

Other efforts involving basic research can focus on the production of potentially "new" materials for high-technology applications. Examples of such materials and their potential applications include amorphous "metallic-glass" substitutes for low-loss transformer cores, amorphous semiconductors for use in electronic device applications, photosensitive polymers and inorganic materials for dry lithography, and high-temperature superconductors for power generation and transmission.

IMPROVED CATALYSIS

Although catalysts are of vital importance to a number of industrial processes, particularly in the chemical and petrochemical industries, an understanding of how catalysis works is available in only a few isolated cases. More efficient catalysts could result in industrial processing techniques that use less energy and raw material and result in less waste and pollution of the environment.

Basically catalysts are substances that modify or increase the rate of a chemical reaction, without themselves changing chemically at the end of the reaction. Industrial catalysts are now designed mostly on a trial and error basis. Basic research is needed to discover new concepts based on fundamental understanding of structure, bonding, and reactions that will rationalize and extend these empirical correlations.

Examples of the possible research topics involve both homogeneous and heterogeneous catalysts. More research is necessary, for example, for understanding the electronic structure of an ordered solid in chemically useful terms. Other topics of potentially great investigative relevance include surface construction as well as electronic surface structure, structural studies of absorbate-substrate systems, and in situ studies of chemical reactions on surfaces. Further work on catalytic processes that proceed at lower temperatures and pressures and require less energy would be very useful.

Progress in these areas can help alleviate several critical national problems in addition to the obvious waste of material that accompanies a reaction of less than desirable specificity and yield. Energy must be expended in purifying and separating the product. Capital is devoted to more sophisticated plant construction than necessary. With an adequate basic research effort, U.S. industry could use the fundamental information to drastically modify existing processes within the next decade or two. These innovations will benefit not only individual companies but the U.S. economy as a whole and contribute positively to our balance of payments in international trade.

Techniques and instrumentation now exist to carry on very advanced research in this field. Although research in some of these areas is manpower-intensive, nuclei for research groups are now available in many academic institutions. In addition, cooperative research projects jointly supported by several countries are possible and would be useful.

COMPOUNDS AND MIXTURES FOUND IN BIOLOGICAL AND INDUSTRIAL WASTES

Waste products from industrial physical, chemical, and biological processing consist of complex compounds and mixtures of substances. A better understanding of them is likely to lead to new concepts for recovery of their resource values. Solid waste alone, from urban, industrial, agricultural, and mineral processing sources in the United States, is produced at a rate of more than 4 billion tons per year—approximately 100 pounds per person per day. These wastes require a vast commitment of human, material, and energy resources for their refinement, fabrication, and ultimate disposal. They present a potentially valuable source of resources as well as difficult and somewhat frightening environmental problems (e.g., toxic chemical waste). Further significant advances in waste recycling technology are likely to depend upon new fundamental research knowledge.

Research needs range from better understanding of the chemical, physical, and biological characteristics of waste-stream composition, to the equally complex social, economic, and behavioral problems of the evolution of policy, cost-benefit analysis, and attitudes toward secondary materials.

The issues expected to be addressed in the search for self-sustaining systems that would biologically stabilize organic wastes and those that would lead toward cost-effective extraction of valuable resources from industrial residues are those that historically have benefited most from scientific research. These research areas should attract substantial cooperative industry-university efforts.

The problems related to resource and energy shortages and waste disposal can be expected to increase in magnitude and severity unless remedial actions are mounted on a timely basis. The potential for the present stock of fundamental knowledge to contribute importantly to needed solutions through recycling of waste has been substantially exhausted. The scientific capability to mount such a research effort is available.

GEOPHYSICS AND GEOCHEMISTRY UNDERLYING MINERAL EXPLORATION

Advances in science and technology can directly and indirectly assist in alleviating U.S. problems of mineral supplies. These problems are the result of increased demands upon supplies for new and increased quantities of products and services as well as difficulties in locating and exploiting mineral values. While some of the supply concerns have been and will continue to be alleviated by attempts to conserve, substitute, and improve product quality, major reliance on existing minerals will remain for a considerable time. This circumstance underscores the need for research and development to locate, identify, exploit, and process the desired mineral deposits. Research undertaken with these objectives in mind can be expected to benefit related areas of fossil fuels, water resources, and waste disposal.

The importance of research in minerals exploration is heightened by the approach of conditions in which many critical minerals are no longer found in any appreciable quantities at surface or immediate subsurface sites. Future domestic exploration must focus on deposits at increasing depths where surface indications are more obscure. To aid in the discovery of such deposits, research must be conducted on the natural processes of mineralization within native geologic environments and on methods that can pierce increasing thicknesses of cover.

Advances in knowledge of global plate tectonics have already provided the opportunity for increased understanding of the genesis of mineralization, while the development of new techniques such as aeromagnetism will allow improved examination of hidden geologic settings. Requirements exist and opportunities abound for research to explain mega-hydrological features and paleomarine conditions. The reconstruction of paleoclimatic belts for given time periods holds promise for helping to identify favorable regions in which laterites may be deposited. To aid in the development of exploration methods, research is needed on multiple-data-set modeling to permit integration of geophysical and geochemical data.

The organizational and technical tools are available for concerted efforts that can make major contributions to the understanding of mineralization processes and to the identification of the signals that permit deposit location. Programs could be mounted to undertake research on chemical migration in the Earth's crust, focusing primarily on processes that concentrate ore minerals. Related research would deal with fluid movements important in the accumulation of oil, and selected concerns in water resources and waste disposal systems. Continental drilling could be undertaken to gain more information on the origins of known ore deposits. Such a program could involve cooperation among several Federal agencies, academic institutions, and industry. Complementary studies in submarine geology and geophysics, margin drilling, and the development of the applications of multichannel seismic reflection techniques can also be expected to contribute valuable information on the location of economically useful mineral deposits.

COMBUSTION

Combustion of coal, oil, gas, and other fuels accounts for about 96 percent of the energy conversion processes in the United States. Of the 50 million barrels oil-equivalent consumed per day, only 50 percent ends up as useful energy. The balance is wasted, mostly due to conversion losses. Even a slight improvement in combustion processes will mean a substantial economic and energy savings.

Increasing reliance on coal and high-sulfur oil in the foreseeable future dictates that we must effectively eliminate pollutant and soot formation during the combustion process. With 120 million cars expected to be on the road by 1985, the environmental impact of combustion research is self-evident. In addition, the increasing availability of new materials for clothing, buildings, furnishings, and other uses also requires a fundamental understanding of their combustion characteristics so as to ensure product safety.

Combustion is the process whereby materials are oxidized at a rapid rate with chemical energy released by the severance, rearrangement, and formation of chemical bonds between substances in the reacting system. In practical situations, the system is of such dimensions that both the chemical reaction rate and the physical exchange of energy and reacting species within the system and with the environment are significant. Both the chemical and physical aspects of combustion need to be further investigated. Important general areas of research are turbulent reacting flows, hydrocarbon and coal reaction kinetics, formation of oxides of nitrogen and sulfur, flame behavior near the fuel-lean and fuel-rich extinction and ignition limits, spray combustion, soot and carbon formation; and combustion of composite materials (e.g., water-in-oil emulsions and coal-in-oil droplets). Many of these areas would be suitable for research cooperation among industry, university, and government laboratories.

PROPERTIES OF INFORMATION

A major imperative of modern society is the creation and maintenance of capabilities to obtain data, synthesize information, and make timely decisions. Ever-increasing quantities and complexities of data require greater efforts at selection and integration of elements in a reasoned and timely fashion. Powerful achievements in the field of computer science have made it possible to control numerous processes and ease communication difficulties to the benefit of all sectors of society.

The expansion of the role and importance of information has long since led to the establishment of the field of information science. Information scientists have attempted to focus on a number of concerns among which the understanding of the properties of information and their translation to practical application are particularly significant. To pursue these objectives, four research tasks of major importance have been identified:

(1) Improving understanding of the structural properties of information collection and of access systems;

(2) Increasing understanding of the connections between form and content as well as the relationships between the structures of language and information;

(3) Identifying abilities and limitations of user cognitive processes and their relation to external stores of information; and

(4) Developing comprehensive, quantitative simulation models of the information economy to predict the impact of technological, economic, and other changes on the transfer of information.

Rapid progress in communication and information technology has strongly affected the nature and direction of computer science research. In the field of computer system design, submicron technology is expected to provide new stimulus for basic research on computing instruments, to be aided by the availability of intelligent graphics terminals. The analysis and design of software systems and computer data bases have been advanced through the techniques of distributed computer processing. Recent achievements in both hardware and software have yielded a new threshold from which a variety of new and important research directions can be expected to unfold over the next five years.

To meet the growing demand for economic, efficient, high-capacity communication systems, new efforts are required in basic engineering research that deals with optical communication systems, large-scale computer communication networks, and very-large-scale integrated electronic circuits. Much more emphasis is also required for the development of integrated optical circuits similar in concept to integrated electronics. To combat effectively the limitations that are inherent in the finite nature of the usable frequency spectrum, much attention will have to be devoted to the study of multiuser systems.

Many of the research topics could provide potentially valuable results in protecting such basic rights as privacy, copyright, and confidentiality. These issues will continue to weigh on and influence the course of research so that new technologies can be exploited with maximum benefit and minimum harm.

Nuclear Regulatory Commission

The Nuclear Regulatory Commission (NRC) is an independent agency whose statutory responsibility is to regulate civilian nuclear activities so that the public health and safety, national security, and environmental quality are protected and the antitrust laws are obeyed. NRC regulates the construction and operation of production and utilization facilities such as nuclear power plants; regulates source, byproduct, and special nuclear material; inspects facilities, licensees, and power plant operations and enforces the regulations and license conditions; cooperates with the states in their licensing responsibilities; prepares standards and regulatory guides; and engages in regulatory research.

The NRC is actively working on certain problems deemed to be of national significance. Much of the effort lies in the areas of research and technical assistance. The specific problems described in the paper are grouped into the areas of nuclear reactor safety, waste management, safeguards, health effects from low-level radiation, and nonproliferation.

NUCLEAR REACTOR SAFETY

The Nuclear Regulatory Commission is addressing itself to the resolution of, or to the better definition of, a broad spectrum of interdisciplinary issues. Of particular interest are the following general areas:

Stress Corrosion Cracking

Stress corrosion cracking (SCC) leading to loss of material integrity has occurred in major coolant system components of nuclear reactor facilities; for example, in steam generator tubes (Inconel) of pressurized water reactors and in major primary coolant piping components (Stainless Steel-304) or both pressurized and boiling water reactors. The fundamental mechanisms and physical processes associated with SCC, including the conditions necessary to initiate it, should be thoroughly investigated.

Two-Phase Flow Phenomena

To better predict the behavior of many safety systems in a nuclear reactor, it is necessary to understand the coolant flow behavior occurring in both the liquid and vapor phases. Thus basic properties and phenomena associated with two-phase flow should be thoroughly investigated.

Human Behavior as It Relates to Reactor Safety

The reactor operator has frequently been considered the weak link in operational reactor safety. However, there are indications that human flexibility can enhance reactor safety because of the ability the operator has to react to unexpected circumstances. The positive and negative aspects of the role of the reactor operator in operational reactor safety should be objectively assessed.

Enhanced Uranium Usage

It has been proposed to develop fuel elements for light water reactors which will perform more efficiently and will lead to a better overall use of our existing uranium fuel resources. The NRC will review the safety aspects of enhanced uranium usage.

The Role of Risk Assessment Techniques

The Reactor Safety Study (WASH-1400) has been reviewed by a study group appointed by the Commission, and the conclusions and recommendations of that group have been accepted by the Commission. A policy statement on how to implement the group's recommendations has been issued. Risk assessment techniques will be used in various aspects of nuclear power plant regulation. Design of these techniques will concentrate on external sources such as fires, floods, and earthquakes, the development of improved methods for analyzing human factors, and the development of component and system data bases. (Risk assessment techniques will also be extended to activities such as waste management and fuel cycle facilities.)

Level of Funding

The fiscal year 1978 research and technical assistance contractual support for nuclear reactor safety was approximately $130 million. That is expected to rise to $150 million in fiscal year 1979 and $180 million in fiscal year 1980. The amount is expected to remain the same during the next five years. Approximately 10 percent of the total reactor safety budget is spent on advanced reactor concepts.

WASTE MANAGEMENT

The licensing of a geologic repository for the disposal of high-level radioactive waste is expected to be unlike any licensing that the NRC has previously undertaken. First, the Commission will be making a finding, based on its review of the Department of Energy (DOE) application, that the facility is or is not capable of protecting the public health and safety over long time periods. Second, the applicant will be a U.S. Government agency, DOE, and third, the public perceives the successful demonstration of a capability for managing high-level waste as the most critical nuclear power issue.

Each step of the licensing sequence for the first geologic repository is going to receive close scrutiny and participation by the public, industry, and other government agencies. It will be NRC's role to evaluate an application to make a finding of "safety" in a well-structured way that is open to public inspection. The public's confidence and acceptance of a decision to proceed with repository development will depend on its perception of the job that NRC is doing in the public interest.

Site Studies

In April 1976, the NRC initiated site suitability scoping studies for geologic repositories in bedded salt to develop an awareness of those aspects of a site that are important to safety and of those parameters important to the migration of radionuclides. The scoping studies were also intended to gain insights into geologic processes that might be encountered by a repository over its lifetime. Additional studies have considered repository design and waste form performance. The results of these studies have been used in developing lists of technical information requirements.

Models

A critical part of the NRC technical program is the development and testing of models. The six purposes of model development are to:

(1) Assist in the development of an understanding of the relative importance of various siting, design, and waste form considerations in the overall performance of the repository; this understanding is critical to the development of regulations and technical positions for waste management;

(2) Place technical performance boundaries on anticipated repositories over a credible range of scenarios to support a finding that the repository can reasonably be expected to perform in a manner that will not adversely impact public health and safety for periods of time measured in thousands of years.

(3) Assist in determining those facets of surface facility design or subsurface systems or structures that should be considered important to safety;

(4) Identify major data deficiencies and determine the importance of reducing those deficiencies (through uncertainty and sensitivity analysis)—these insights will become increasingly important during a license review for identifying areas where data refinement is needed and where it is not needed (i.e., where data refinement would have a minimal impact on repository performance);

(5) Develop insights to those aspects of the operational and isolational phases that are sensitive to design changes; and

(6) Facilitate a timely and efficient license review and provide timely guidance to DOE.

Technical Information Requirements

The technical information requirements to be developed by NRC for the preliminary site review concern data acquisition, qualitative site characteristics, evaluation of alternative sites, natural resources, climatology, hydrogeological characteristics and processes, design considerations in siting, retrieval philosophy, monitoring philosophy, borehole guidance, preliminary site review, model development, radiologic impacts, environmental impacts, other impacts, and inventory.

Before provisional construction authorization is granted, there are four major steps, each of which includes several technical studies or activities. The initial systems modeling includes these requirements: use of modeling in the licensing process, use of probabilistics in the licensing process, sensitivity analysis, uncertainty analysis, fracture flow modeling, model documentation requirements, model sophistication, radionuclide transport modeling, and alternative impact evaluation. The evaluation of waste/package performance must cover waste package retrievability, waste/rock interactions, waste package quality control, waste classification, waste form/package, and waste package durability. The provisional construction authorization repository design requires study of repository excavation and construction, rock mass sealing, borehole sealing, systems important to safety, monitoring, quality assurance, operational phase testing and design verification, and surface facility decommissioning. In addition, siting requires technical information about natural resources, climatology, and treatment of demography.

Research in management of waste and spent fuel will be significantly accelerated with particular emphasis upon mid-term methods that can bridge the time period until ultimate decisions on waste disposal are made.

Level of Funding

In fiscal year 1978 research and technical assistance contractual support for waste management amounted to $13 million. That level remains for fiscal year 1979, but fiscal year 1980 levels are expected to be $18 million. Levels may increase in fiscal years 1981 through 1983.

SAFEGUARDS

Safeguards are necessary to prevent theft of licensed materials or sabotage. In the near term, the major research effort will be concentrated on completion and transfer to operational activities of methods that will assist performance evaluation of safeguards. At present, these evaluative methods are being field tested and modified to incorporate user comments and to increase their utility. Limited tests to date have produced promising but inconclusive results.

When the current national program to evaluate alternative fuel cycles and new approaches to preventing proliferation is completed, the NRC plans to consider studying the impact these evaluations portend for domestic, foreign, and international safeguards. NRC will also examine the possibility of transferring safeguards technology developed in the United States to other nations.

The current program of research to decrease the vulnerability of reactors to sabotage should be completed by the end of the five-year period, and it will have continuing impact on reactor design.

Level of Funding

Approximately $11 million was spent on research and technical assistance contractual support for fiscal year 1978 in the area of safeguards. In fiscal year 1979 the level will be near $13 million and near $14 million in 1980. Levels for fiscal years 1981, 1982, and 1983 are expected to be reduced slightly.

HEALTH EFFECTS FROM LOW-LEVEL RADIATION EXPOSURE

Recent epidemiological studies suggest that current radiation protection standards may not be adequate. The adequacy or inadequacy of these standards could have considerable significance for the use of nuclear power, medical use of x-rays and radioisotopes, and other widespread industrial and research applications of radioactive materials. Additional studies are required to determine whether these results are correct and to better determine the health risks associated with current radiation protection standards and, more specifically, the health effects associated with low levels of radiation exposure. Because of the nature of the problem, such studies are expected to be of extended duration and expensive to perform.

Human epidemiologic studies on low-level radiation effects would involve the expenditure of millions of dollars over the next several years. It is not certain which organizations should take the lead in this research, and, due to Privacy Act limitations, the effort might require new legislation to allow the collection and use of health and other data. The current Interagency Task Force on health research and radiation protection, under the leadership of the Department of Health, Education, and Welfare, is investigating this problem; it has made recommendations to the White House on what should be done, who should do it, and how, and what new legislation might be necessary to do it. NRC is performing several research studies to support the government efforts.

Level of Funding

About $1 million was spent in fiscal year 1978 on research and technical assistance in studying radiologic health effects. The fiscal year 1979 level is expected to rise to $2 million, and $3 million is estimated for fiscal year 1980. In fiscal years 1981 through 1983 the levels will probably be about $3 million.

NONPROLIFERATION

President Carter's nuclear policy statement of April 7, 1977, committed the United States to a number of actions motivated by proliferation concerns. These actions include, among others, indefinite deferral of the commercial reprocessing and recycling of the plutonium produced in the U.S. nuclear power programs, restructuring of the U.S. breeder reactor program to give greater priority to alternative designs of the breeder and to defer the date when breeder reactors would be put into commercial use, and redirection of U.S. nuclear research and development programs to accelerate research into alternative fuel cycles that may afford less ready access to materials usable in nuclear weapons. The statement also called for an International Nuclear Fuel Cycle Evaluation (INFCE). That proposal was taken up by an international conference in October of 1977 and instituted as a two-year effort which is currently underway and scheduled for completion in early 1980. The United States also organized in 1977 the Nonproliferation Alternative Systems Assessment Program (NASAP). The goal of NASAP is to support U.S. participation in INFCE and to provide recommendations for the development and deployment, both nationally and internationally, of civilian nuclear power systems that offer improved proliferation resistance, use fuel resources efficiently, and have acceptable public safety and environmental characteristics.

The Department of Energy has the lead role with regard to the U.S. technical studies and analyses in INFCE and NASAP. NRC's role in NASAP primarily involves an assessment of the licensing issues and problems of the principal alternate technologies and the preparation of a report to the President and Congress on the comparative safety, environmental, safeguards, and licensing aspects of those technologies under serious consideration by DOE. NRC's role in INFCE involves the participation of individual NRC staff members on various U.S. support groups that prepare analyses and review U.S. and foreign papers developed for the INFCE working groups and Technical Coordinating Committee.

NRC's fiscal year 1979 Authorization Act makes $1 million available "for studies and analyses of alternative fuel cycles (including studies and analyses relating to licensing and safety, safeguards, and environmental aspects)." Section 9 states that NRC shall monitor and assist, as requested, in INFCE and studies of the various nuclear fuel cycle systems by DOE. The Commission will then report to Congress, semiannually through calendar year 1980 and annually through calendar year 1982, on the status of domestic and international evaluations. The reports are to include, but not be limited to, a summary of the information developed by and made available to the Commission on health, safety, and safeguards implications of the leading fuel cycle technologies.

Level of Funding

The NRC efforts on nonproliferation have been chiefly in support of NASAP and INFCE. About $1 million in research and technical assistance is planned for fiscal year 1979; none was spent in fiscal year 1978. Levels will probably decrease slightly during the five-year period.

Veterans Administration Department of Medicine and Surgery

Research in the Veterans Administration (VA) is concerned primarily with improvement of health care, especially of veteran patients. An integral part of this mission is to provide a better understanding of the phenomena underlying disease, disability, and health. To achieve its objectives, a broad study of life processes is pursued.

Most VA research is "applied" in nature because of the interest in its practical application. However, basic research is also encouraged. VA research is conducted almost exclusively by its own staff within the system's hospitals.

In the following sections, current research problems are described together with promising approaches which may lead to their solution. The constraints on the scope of research to be undertaken and alternative approaches to conducting the projects are discussed.

DYSVASCULAR LOWER EXTREMITIES

Several diseases and injuries alter the blood vessels of the legs and feet so that blood no longer flows normally to them. The impairment can become so great that tissues die, necessitating amputation of the part. Greater disability results as more of the extremity is removed, and surgeons try to amputate as "low" as possible. If the amputation does not remove all compromised tissue, however, the stump is unsatisfactory, and disability is increased. Determination of the exact state of the circulation before amputation is critical in achieving minimal residual disability.

Methods are now available to obtain some information about the state of damaged circulation in the legs and feet. None has proved satisfactory when used alone, and efforts are now underway to coordinate the use of several techniques including variations of the Doppler effect with ultrasound, electromagnetic flow meters, temperature monitoring by infrared thermography, radioisotopic flow measurements, and the determination of blood gases, especially oxygen, through the skin. New methods will be investigated as well. Annual funding is estimated at $500,000 to $1,000,000.

AGING AND ASSOCIATED PROBLEMS

The nation's population includes an increasing proportion of citizens reaching advanced age with attendant physical and mental limitations and with diseases peculiar to the elderly. The veterans of World War II are now approaching their later years and, as the largest group, are rapidly shifting the average age of the veteran population toward the upper decades of life. In the next few years, an increasing importance will be placed on studying the disabilities and diseases of the elderly.

Availability of a large population of aging and elderly patients provides Veterans Administration clinicians and scientists with firsthand knowledge of the problems of aging and a population of volunteer subjects who may profit from research on these problems. Investigation is needed on the fundamental changes that occur during aging. Detailed biochemical and psychological studies of the brain and its reaction to psychologically active drugs should be undertaken. Alterations in heart and blood vessels, defense mechanisms, and metabolism should be studied to provide information that is basic to prevention and treatment of defects. Annual expenditures for aging research are estimated to be $750,000 to $2,000,000.

BIOLOGY OF SCHIZOPHRENIA

Schizophrenia, one of the major groups of mental diseases, primarily attacks young people, for example, Vietnam veterans, and imposes a lifelong handicap. Treatment has improved the quality of life of many schizophrenic patients, but it is still far from perfect and sometimes produces side effects that are serious handicaps in themselves. Current treatment consists of maintenance and continues over years. Curative treatment would be more satisfactory to the patient and less expensive for the provider.

Schizophrenia has been considered both a single but varying disease and a loose collection of diseases that share a few characteristics. It is essential for future research to determine the biological relationships and definitive properties of each clinical condition labeled schizophrenia. The drugs used to treat schizophrenia act on the biochemistry of the brain and also produce side effects that reflect chemical changes in the brain. These and other observations suggest that schizophrenia is associated with chemical changes, as causes or effects. Improved biochemical techniques, especially those involving enzymes, make possible more detailed studies of the schizophrenic brain and may eventually open the way to an improved treatment or cure. One requirement for successful research in this field is the availability of properly preserved brains from schizophrenic patients. The Veterans Administration, with a large number of well-studied schizophrenic patients, could establish a "research bank" of brains from those who die. Annual funding for schizophrenic research is estimated at $500,000 to $1,000,000.

BIOLOGICAL FACTORS IN ALCOHOLISM

Despite years of investigation, the biological characteristics that predispose people to alcoholism or perpetuate the condition remain unknown. Alcohol abuse is one of the most widespread and costly problems in much of the world and is a major cause of ill health, illness, and death among veterans. Alcoholic patients have been reported to differ genetically and biochemically from the balance of the population, but causal relationships have not been demonstrated. Elucidation of these relationships should provide a basis for cure and prevention of the disease.

While the social and psychological problems of alcoholism are being extensively studied, the biological aspects are less thoroughly investigated. The changes produced by alcohol in such body structures as the membranes covering each cell and lining the intestinal tract are beginning to be investigated in order to understand how the substance exerts some of its effects. Other changes in brain biochemistry, and related alterations in the nervous system, may explain why alcoholism is so persistent. Research into hereditary factors may explain why alcohol consumption leads to alcoholism in some individuals but not in others. Annual funding for basic research in alcoholism is estimated at $1,000,000 to $2,000,000.

COMPREHENSIVE HEALTH CARE BY A FEDERAL AGENCY

The provision of comprehensive health care by privately operated Health Maintenance Organizations (HMOs) is now well studied in the United States. No Federal agency has established an analogous system, and no research has been done on the special problems that may arise. It is not even clear what research procedures should be used in the scientific investigation of such an extensive organization. Establishment and validation of the methodology requires basic health services research.

The Veterans Administration has a large and varied health care system divided into districts, and it is possible to change the manner of delivering health care in one district and compare the results with those in another district. The current method of episodic care of veterans for specific illnesses could be replaced in one district by a health care delivery scheme similar to an HMO, with total and continuing care being provided to a more restricted group of veterans or to veterans and their dependents. Comparison of the HMO-like system with the episodic one would require development of criteria and a practicable method of accumulating relevant data to judge effectiveness of health care delivery. Annual funding for health services research, exclusive of the cost of clinical care, is estimated to be $250,000 to $750,000.

TISSUE REPLACEMENT BY TRANSPLANTATION AND REGENERATION

Although transplantation has replaced internal organs, notably the kidney, there has never been a satisfactory method to replace such other body parts as extremities or central nervous tissue. Successful reimplantation of severed limbs and the body's ability to regenerate some tissues suggest that replacement or regeneration may be possible for limbs or other parts. The possibility depends upon the solution of numerous basic biological problems concerning the control of tissue structure and function.

Grafting of a part from one individual to another provokes a set of complicated reactions that lead to the rejection of the part by the recipient. These reactions are becoming increasingly well understood and better controlled. This leads to the hope that it may be possible to transplant limbs; a useful limb, however, requires functional nerve connections with the recipient's body, and this in turn depends upon appropriate nerve regeneration. Transplantation and regeneration thus share common problems that will have to be attacked through basic research. The details of how one cell changes into another specific cell type, either generalized or specialized, and control of these changes must be investigated before attempts can be made to induce them. Basic problems include how to initiate, direct, and suppress cell growth, division, and function. Although clues to the solutions of these problems have appeared, no quick answers can be expected. Annual funding is expected to be $1,000,000 to $2,500,000.

MONITORING OF SPINAL CORD FUNCTIONS

Injury and disease destroy tissue of the spinal cord, the connection and waystation between the brain and most of the body. Interruption in the continuity of the cord results in loss of sensation and of voluntary control over the body below the damage. The results are most evident in the extensive paralysis seen after injuries such as a broken neck. In order to understand exactly how such interruptions affect function, and to treat them more effectively, scientists and physicians must be able to accurately and specifically measure the changes that are occurring in the damaged cord. This is not yet possible.

The spinal cord contains elements that transmit impulses between the brain and most of the body, as well as nerve cells that modify or control many reactions. Since the impulses are, in large part, electrical in nature, electronic techniques should disclose the extent of transmission by nerve cells within the cord. Injury and disease often destroy only a part of the spinal cord, impairing, but not completely blocking, transmission. Accurate measurement of the extent of injury is essential to improve the evaluation of defects and to judge the success of attempts to correct them. The development of practicable methodology is thus an essential step. Annual expenditure on spinal cord monitoring is estimated to be $500,000 to $2,000,000.

RESEARCH CONSTRAINTS

The extent of research in the above areas is currently constrained by restrictions on funds and personnel. To exploit the areas fully, the Veterans Administration would require more research space than is currently available. The Veterans Administration has other research commitments that must be met to fulfill the agency's mission of providing health care for eligible veterans. Diversion of present funds and personnel into the identified areas would impair efficiency in other important fields.

ALTERNATIVES

Most of the research identified here could be conducted by other agencies, under contract, or by a system of grants to academic institutions, rather than being conducted as an intramural function of the Veterans Administration. The value of the intramural method lies, however, in the proximity of the clinicians, who often are the principal investigators, to the laboratory scientists. Even basic research under this arrangement tends to be directed toward useful ends because of the obvious needs of the patients. In the later stages of research and development, the availability of patients under the care of clinician-scientists provides volunteers to test the results of the research. Being involved in such testing is quite likely to benefit the collaborating patient as well.

SCIENCE, TECHNOLOGY AND PUBLIC ISSUES:

Papers on Selected Topics Commissioned by the National Science Foundation

I Science, Technology and The Uses of Information

1 Science and Technology Policy and the Democratic Process

by Dorothy Nelkin*

SUMMARY

This paper reviews the issues involved in the demands for greater public participation in science and technology policy. Numerous controversies suggest that the sources of "participatory politics" lie in the changing public image of science and technology and especially in the changing attitudes toward expertise as a source of political legitimacy. The paper analyzes these sources of conflict and the various "publics" involved, describing diverse fears about science and technology that lead to participatory demands. While surveys indicate that most people perceive science and technology as instrumental in achieving important social goals, specific concerns persist. The central and pervasive issue—the source of many conflicts—is the question of who controls crucial policy decisions.

As various groups seek greater participation in science and technology policy, governments have expanded participatory channels both within the representative system and through administrative reforms In addition, participatory experiments outside of official administrative channels have proliferated. Describing experiments in the U.S. and Western Europe, the paper suggests the range of options available to accommodate participatory demands. It also analyzes some of their difficulties, as the complexity of science and technology and the expectations of efficiency associated with technical planning obstruct efforts to develop participatory procedures.

The demands for participation can affect the application of resources to research, the strength of research institutions and the way scientists choose their topics and conduct their work. And participation can influence the general climate—political, social and intellectual—in which research takes place. Thus its implications warrant special attention in the coming years.

INTRODUCTION

Everywhere symbols of progress are under public scrutiny. Technologies of speed and power—airports, highways, nuclear plants—provoke antagonism as local communities protest against noise and disruption. Science-based programs such as genetic screening are a source of public debate. And important areas of research face persistent opposition: note, for example, disputes over fetal research, recombinant DNA, and studies of the genetic origins of human behavior. Technical problems that were once considered the preserve of experts are increasingly forced into the political arena as many interests seek greater control over the direction of science and technology.[1]

Until recently, most questions of public control focused on technological applications, but public scrutiny has come to rest on science as well. Given the policy importance of many areas of scientific research and the

* Professor, Program on Science, Technology and Society and Department of Sociology, Cornell University.

growing concern in the biological sciences with basic life processes, such scrutiny is inevitable. Indeed the relationship between science and the public appears to be in the process of significant renegotiation.

Despite the intensity of some disputes, few people seek to impose major restrictions on science and its applications; most areas of science and technology are valued as instrumental in achieving important social goals. Rather, the negotiation is over *who* should participate in establishing policies and controls, how these controls will be organized, and how much they should influence decisions concerning the direction and conduct of research. Reflected in these questions is a troublesome feature of contemporary society—the impact of increased complexity and specialization on the decision-making process. The power afforded to those who control technical information reduces public control over many public policy choices, and technical complexity appears to limit effective political choice. This concern is ubiquitous in the controversies over science and technology. The key slogans in these disputes are "public accountability of scientists," "demystification of expertise" and "lay participation." Local groups demand a voice in the location and design of nuclear power plants. Patients seek greater control over their medication. Parents insist on influencing the science textbooks used in schools. Consumers raise questions about the validity of the data that back government regulation of drugs such as cyclamates, saccharin and laetrile. Citizens seek to participate in shaping the rules and standards that control science and technology or the conditions of work in the laboratory.

Participatory demands have focused both on the conduct of research (the recombinant DNA and fetal research disputes) and on its public consequences (research on genetic manipulation or on the sources of criminal violence). Indeed, a new field, "the social assessment of science," is emerging to examine "the involvement of non-scientists in the assessment, judgment and criticism of science."[2]

While the concern in this review is more with policies that affect the conduct of research than with the management of technology, it includes examples of the latter. It is often hard to distinguish science from its applications, and public concern about technology extends to science, especially in those areas perceived to have eventual application. People are clearly more concerned about technology than about science, but when asked in a public opinion survey whether it is more important for society to control science or technology, 59 percent responded that both must be controlled equally (20 percent that neither should be controlled at all).[3]

This paper, then, analyzes the politics of participation in both science and technology policy. It describes the sources of participatory demands, the existing procedures and some recent experiments in accommodating those who wish to participate, some of the obstacles to greater participation, and, finally, some of the potential implications of greater public involvement.

SOURCES OF PARTICIPATORY POLITICS

Demands for greater public involvement i[illegible] and technology reflect the convergence of a heig[illegible] sensitivity to the environmental and social implications of science and technology with a changing view of expertise. These are the pre-conditions for participatory politics as people seek greater control over decisions that affect their lives.

CHANGING IMAGES OF SCIENCE AND TECHNOLOGY

Attitudes toward science and technology are ambivalent: Optimistic expectations about potential social good are mixed with concern about undesirable consequences. Surveys indicate that most people view science as instrumental in achieving important social goals, yet only a small majority (52 percent) of respondents in a 1976 survey said they believed that science and technology have produced more good than harm. By far the greatest perceived benefits were in the field of medicine (81 percent); in other areas—improved living conditions, environmental conservation, energy programs and improved communication—science and technology were reported as beneficial only 10-14 percent of the time. A review of longitudinal data on the public's confidence in people who run institutions found that the proportion of respondents expressing "a great deal of confidence" in the scientific community declined from 56 percent to 43 percent between [illegible] and 1976.[4] However, in gauging the public's esteem for science relative to other occupations in 1976, science ranked second (behind medicine) and engineering third. Public regard for scientists has declined in absolute terms, but this reflects the declining trust in most major institutions. Between 1966 and 1976 those expressing a great deal of confidence in Congress declined from 42 percent to 14 percent, the heads of major corporations from 55 percent to 22 percent, the U.S. Supreme Court from 51 percent to 35 percent, and educators from 61 percent to 37 percent. In this context scientists have fared rather well.

But opposition to specific projects persists. Science and technology are often blamed for environmental pollution, loss of privacy and civil liberties, the impersonal character of medical care, the routine character of work, and the threat of greater social control.

Concerns about science and technology extend beyond the fear of risk. Some critics (e.g., those opposing fetal research or the teaching of evolution) feel that science has ethical implications that threaten deeply held personal beliefs. Others worry about the potential misuse of scientific findings: for example, critics of research on recombinant DNA, on the XYY chromosome and on the relationship between genetics and intelligence question the implications of relating genetically mediated characteristics to human behavior. Biology, they claim, is "a social weapon." "Can future generations cope with the possibilities science opens up today?" Still others are concerned about equity both in the allocation of resources to science and technology and in the distribution of social and environmental costs.

It is often argued that changing attitudes toward science reflect disappointed expectations. Atomic energy and the space program brought science from the obscurity of the university laboratory to the forefront of American consciousness, creating optimistic expectations—the so-called "Moon-Ghetto syndrome": Given enough support, science and technology can solve all problems.[5] Yet the 1960s also brought the realization that science and technology not only failed to solve social problems but often contributed to them.

The changing image of science and technology also reflects reaction to the increased role of the Federal Government in this sector. The rate of technological change, the scale of scientific and technical projects, the rapid diffusion of technology to new areas, and the high costs associated with R&D have necessitated increasing Federal commitment to develop and regulate many technologies and some areas of scientific research as well. These have essentially moved from the private to the public sector.

This has happened at a time when public interest in playing a role in major policy decisions at all levels has reached a high point. Science and technology, of course, have not been immune. Demands for public accountability and questions about the appropriate focus of authority in science and technology policy reflect a more general loss of confidence in established institutions and a concomitant concern with public or consumer influence.

CHANGING IMAGES OF EXPERTISE

As technical knowledge assumes growing policy importance, scientists and engineers are increasingly active in public policy decisions through advisory boards, special commissions, staffs, and consultant groups.[6] Federal Government expenditures for consultants have been estimated at over $1.8 billion a year. Moreover, there has been a general evolution from *ad hoc* advisory activity to the institutionalization of an "intellectual technocracy." Federal Government employment of scientists grew by 49 percent from 1960 to 1970; for social scientists the increase was an even greater 52 percent. (Total Federal Government employment grew by only 30 percent during the same period.)

The implications of the growing policy role of expertise are not clear. Some perceive evidence of a growing technocratic elite—a new source of power with decisive influence over both politicians and the public.[7] Political power, it is argued, has shifted significantly from political and corporate decisionmakers to "knowledge elites" who derive their authority from the cultural emphasis on rationality, efficiency and technological progress, and from the specialization inherent in bureaucratic organization.

An alternative view sees the technical elite essentially as a "mandarin" class, beholden to the existing establishment and commanding little or no independent power.[8] Experts, it is claimed, use their specialized knowledge to serve the established institutions in industrial society. They are said to be effective only insofar as politicians and entrepreneurs allow them to be, and they pose no real threat to entrenched forms of power. It is further argued that existing institutions legitimize themselves by enlisting technical expertise and using it selectively to give the appearance of objectivity, to support predetermined decisions, to maintain secrecy, or to "buy time" when opposition is anticipated.[9]

Both views of the experts' role have increased apprehension about an erosion of democratic values. Metaphors of "the priesthood" or "the new Brahmans" are used to describe the function of the "knowledge elite" in a social system where authority is based on information, and many policy choices are reduced to technical decisions.[10] The power afforded to those who control technical information reduces public influence. Thus, just as people are more aware of the potential harm associated with science and technology, they also feel they have fewer ways to directly influence them. This structural dilemma creates demands for greater public involvement.

THE "PUBLIC" AND THE PARTICIPATORY IMPULSE

Controversies over science and technology involve a variety of groups concerned with greater influence over the direction and control of science and its applications. As they organize, their concerns are translated into participatory demands.

WHO IS INVOLVED?

The most obvious "public" concerned about science and technology are those persons who are directly affected by land expropriation, immediate risk, or rapid local economic, environmental or social change. Residents of urban neighborhoods often claim that technological developments neglect their needs. They question the justice of the anticipated distribution of costs and benefits from a technology: Can any reduction in some citizens' welfare be justified by greater advantage to others? Can the magnitude or intensity of harm borne by neighbors of a noxious facility be reasonably incorporated into cost-benefit calculations? These questions are expressed in organized opposition to the expansion of airports and the siting of highways or power plants. Recently similar arguments have confronted plans for siting biology laboratories to host recombinant DNA research. Indeed, most participatory demands come from the "neighbors" of science and technology, those directly affected by a planned project and unable to avoid its present impact or potential risks.[11]

In addition, a well-defined "concerned public" includes direct recipients of such professional services as health care. Mental or medical patients are directly affected by the availability of new drugs, new biomedical technologies, and trends in the use of health care techniques.[12] They may also be subjects of clinical research. Here the demands for participation, based on the rights of individuals to have some voice in their own treatment, are often expressed in proposals for changes in professional-client relationships.

A far more vaguely defined public is made up of the consumers of the products of science and technology. Here the participatory impulse is often reflected in protest against government regulation of technology. The laetrile dispute and the debates over regulation of saccharin and cyclamates raise questions about government intrusion and the limits of individual choice. In such cases, the opportunities opened up by science and technology and the potential risks involved in innovation exacerbate this classic political dilemma.

Other groups critical of science and technology share rather global concerns based on ideological or moral principles. These include members of major environmental and public interest associations who seek to influence national policy.

Finally, an important source of criticism has come from the scientific community itself. Many scientists became politicized during the 1960s. At that time, they focused on anti-war activities and the issue of military research in universities. More recently, their attention has turned to the environment, nuclear power or biomedical research. These scientists question the potential risks in areas often obscured from public knowledge.[13]

As differences between experts rise to public visibility, science no longer appears as an objective and compelling basis for policy. Disputes among experts, encouraged by the intrinsic uncertainty concerning the impact of science and technology, call attention to this limited ability to predict potential risks. Controversy thus demystifies expertise, exposing the non-technical and political assumptions that influence technical advice and encouraging the transfer of problems from the technical to the political arena.

FROM PUBLIC CONCERN TO PARTICIPATORY DEMANDS

Technological controversies are marked by the proliferation of citizen groups. Most are temporary coalitions formed to challenge specific decisions, often disbanding once the issue is resolved. But some groups maintain themselves with a core of activists who remain interested in other projects affecting the community and who can mobilize a larger constituency when specific issues arise.

Sustaining these local groups are the large national associations such as the Sierra Club and Friends of the Earth, whose membership has more than tripled since the 1960s. Their relatively stable constituency can be mobilized to intervene in diverse technical areas, and they seek to increase public participation in areas of science and technology that present potential risk. They were, for example, among the active participants in the recombinant DNA dispute.

Consumer protection and public interest science groups have also proliferated: Such organizations as the Center for Science and the Public Interest, the Center for Concerned Engineers, the Coalition for Responsible Genetic Research, the Clearinghouse for Professional Responsibility, and Science for the People call attention to the social impacts and political dimensions of science and technology,[14] and they try to provide citizen groups with the technical expertise necessary to challenge policy decisions.

The interest of these various groups extends beyond specific policies to the broader issue of decisionmaking power. Controversies are pervaded by questions of responsibility and control; "expert accountability" is a central theme. The demands of local community groups often resemble a kind of "mini-nationalism," as they seek to protect themselves by greater local control against the intrusions of technology. Expertise, they claim, should rest with those affected: "We need no experts"; it is an "arrogant assumption" to leave decisions to scientists. Thus citizen groups call for better information, provisions for "counter-expertise" and greater opportunities for participation.

PARTICIPATORY MECHANISMS AND ADMINISTRATIVE REFORMS IN SCIENCE AND TECHNOLOGY POLICY

Demands for greater participation have coincided with congressional qualms about the increasing power of administrative agencies. To sustain the authority of the representative system over decisions in technical areas, the Congress has built up its own technical competence, and it has also imposed legislative requirements on agencies to involve citizens more directly in the formulation and implementation of science and technology policies.

PARTICIPATION AND THE REPRESENTATIVE SYSTEM

Increasingly involved in decisions concerning science and technology, Congress has greatly expanded its technical staff: between 1947 and 1976 the combined staffs of the Senate and House increased from 2,513 to 13,272. Staff appointments have included specialized science-trained consultants who gather technical information and select technical witnesses for public hearings. The many hearings on recombinant DNA, for example, were largely structured by professional staff through the selection of key witnesses. A number of these hearings were specifically organized as a means to brief special congressional committees on various technical and political aspects of this controversial issue.

To improve its technical competence, Congress has also established several research services.[15] The Congressional Research Service was created in 1970 to provide information on special issues. The Office of Technology Assessment (OTA) was established in 1972 to assist Congress in evaluating and planning specific technologies. These technical services allow Congress to re-assert legislative control and oversight over issues that have previously been delegated to administrative agencies.

The primary congressional response to anxiety about democratic representation has been through legislation. The Airport and Airways Development Act, the Federal Water Pollution Control Act, the Coastal Zone Management Act, the Highway Safety Act, the National Environmental Policy Act, and the Energy Re-organization Act all contain requirements for greater direct participation in administrative agency decisions.

ADMINISTRATIVE CHANNELS OF PARTICIPATION

Administrative agencies are a major arena for political action, often replacing legislative bodies in defining policy problems and devising their solutions.[16] Yet agency commitment to rational, efficient decisionmaking usually precludes significant public involvement, except in a few highly politicized proceedings such as the power plant siting hearings of the Nuclear Regulatory Commission.[17] Participatory reforms are mostly intended to expand the information available to the public and to channel information about public preferences to decisionmaking agencies. However, some procedural reforms seek to open the administrative process to negotiation and compromise, allowing public representatives to take part in the development of policies. I will briefly review some of these procedures, suggesting why critics often dismiss them as inadequate and seek participation through other means.

Participatory reforms have provided greater public access to data and reports that underlie the decisionmaking process. The Administrative Procedures Act requires administrative agencies to publish proposed rulemaking in the *Federal Register* and invite public comment. These comments are then taken into account when final regulations are developed. The *Register* contains drafts of controversial proposals; for example, HEW published early drafts of proposed regulatory guidelines for the use of human subjects in research in order to solicit public comments and to test public acceptability of HEW procedures. The National Research Act of 1974 extended this use of the *Federal Register*, as a forum as well as a bulletin board, by creating the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research to advise HEW. It required the Secretary of HEW to publish the Commission's reports and recommendations prior to announcing his proposed rules and to respond if he failed to follow them.

The Freedom of Information Act further extended public access to government reports, as did the requirements for Environmental Impact Statements (EIS). Under the EIS requirements, agencies must publicize the existence of documents and actively disseminate them. This model has generated proposals for "Scientific Impact Statements" that would similarly evaluate the potential effects of research and solicit public review.

These participatory channels can provide useful information and also inform decisionmakers of citizen concerns, but critics contend that considerable initiative, money and access to expertise are necessary to seek out and to utilize useful material. The *Federal Register* is available in libraries and is indexed, but with 60,000 three-column pages per year it is a cumbersome document. Agencies must disclose information on request, but sometimes charge search and copy fees. Citizens unfamiliar with what is available or lacking access to expertise are often unable to request material with sufficient specificity, and crucial material on controversial topics is often withheld as proprietary information.[18] Thus, it is argued, while legislative reforms provide greater legal access to information, they are intended more to reduce conflict than to devolve power. More collaborative modes of participation include having citizen representatives, not simply as informants, but as partners with some power to exercise direct influence.

Advisory boards are a frequent channel for such collaborative participation. In 1975, 45 agencies employed 1,267 advisory committees with 22,246 members. However, a 1976 survey of 16 Federal energy agency advisory boards reported that nearly 50 percent of the board members were industry representatives, while consumer and environmental representation comprised 4 percent and 3 percent respectively.[19] Agencies justified this balance by arguing that citizens' groups were not interested, lacked technical qualifications, or were simply inappropriate; citizens' groups, however, charged agencies with trying to close their deliberations to "outsiders." Their arguments have had increasing effect on increasing citizen representation on advisory boards.

For example, the Federal Energy Administration advisory committees doubled their consumer and environmental representation to 10.8 percent and 7 percent respectively. The FDA has paid consumer representatives elected by consumer groups on its advisory panels. Of the 18 members of the National Council on Health Care Technology appointed to assess the safety, the effectiveness and the social impacts of health technologies, two must be from law, one from the field of ethics and three from "members of the general public who represent the interests of consumers of health care." The National Institutes of Health (NIH) has experimented with consumer representation on advisory panels that are responsible for reviewing research proposals. NIH uses two stages of review: the first by study sections composed of scientists who judge the technical competence of proposals, the second by advisory councils composed of scientists and informed laymen who decide the allocation of funds on the basis of NIH priorities and social considerations.

Disputes over controversial areas of biomedical research have brought expanded public participation in the NIH advisory groups. The 25 people composing NIH's Recombinant DNA Advisory Committee, for example, include several lawyers and ethicists, a scientist from a public interest organization, and several non-scientists who have been actively critical of existing controls over this research. Similarly, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research includes five scientists, three lawyers, two ethicists, and one public representative. It has sought to enhance public involvement through open meetings where public comment was encouraged, through surveys of national opinion to evaluate its own effectiveness as perceived by the public and through the creation of five independent public information centers. In addition, NIH and NSF require that institutions seeking Federal funds form Institutional Review Boards (IRBs) to monitor all research proposals involving human experimentation. These IRBs include lay participants: A recent study found that most members (50 percent) were biomedical scientists, 21 percent were social scientists, and the rest included administrators, lawyers, clergymen and other non-scientists.[20] New policy proposals call for including more lay members on the IRBs and opening IRB meetings to the public.

Some see this form of participation as a desirable way to forestall potentially harmful consequences of research and to ensure scientific work of greater social utility. But others fear that participation will lead to administrative delays, excessive constraint based on short-sighted goals and, ultimately, to dangerous social control over scientific inquiry. Thus, the response to legislative mandates for greater participation proceeds with caution.

OTHER CHANNELS OF PUBLIC INFLUENCE OVER SCIENCE AND TECHNOLOGY POLICY

The structure and responsibilities of government institutions place important constraints on meaningful public participation through administrative procedures. Thus critics of science and technology have developed their own channels of participation through the courts and special referenda, and this in turn has inspired a number of participatory experiments, outside the usual government channels but intended to increase public influence over government policy.

LITIGATION AND REFERENDA

Litigation has been a major means for citizens to restrict and direct technological change.[21] The courts may define responsibilities for damages caused by technology; they can improve injunctions against proposed technologies that may be harmful, and they can galvanize administrative agencies and regulatory commissions into more effective action by re-affirming standards.

The ability of citizens to use the courts has expanded with the extension of the legal doctrine of standing which determines who has a right to be heard. "Standing" to sue varies from state to state, but the rules of standing have generally been liberalized to allow a private citizen without an alleged personal economic injury to present a grievance as an advocate of the public interest. Moreover, in environmental litigation, the courts have accepted as cause for standing not only damage to property rights but some of the more subtle health and aesthetic impacts of technology.

Citizen litigation as a means of policy influence has focused mostly on technological applications, but occasional cases—e.g., the Boston "grave-robbing" trial of medical researchers studying aborted fetuses—focus on science as well. Medical malpractice suits may be extended to clinical research or genetic experimentation. And the growing concern over the adequacy of safety precautions in the laboratory opens possibilities for legal action.

The role of the courts in cases associated with science and technology poses considerable problems. Despite a well-developed system of using expert witnesses, the courts cannot easily deal with esoteric scientific information that involves uncertainty. Many of the problems of science and technology require evaluation of potential risk at some future date. Weighing cancer risks, for example, requires judgments based on statistical evidence. Yet judges are accustomed to ruling on the basis of precedent. Moreover, developments in science and technology raise problems that require re-analysis of legal principles. What, for example, is the appropriate role of law in controlling genetic experimentation? How can the courts deal with new questions that arise in the wake of expanded clinical research without stifling scientific progress? These are some of the issues confronting the courts as they are increasingly used to influence decisions about science and technology.

The referendum is a prominent and growing feature of the United States political landscape and, significantly, technological decisions such as airport expansion, nuclear plant siting and nuclear waste disposal plans are appearing on ballots. The direct citizen vote is a form of participatory democracy within representative governments, and it is intended to preempt administrative authority. While using referenda for technical issues poses problems of representation, adequate information and co-optation, its growing popularity suggests an interest in shifting technical decisions to a more participatory framework.

PARTICIPATORY EXPERIMENTS

The current surge of participatory demands reflects the decline in public confidence in decisionmaking authorities. Concern with this trend has shaped a number of experimental efforts to establish the procedural conditions that would encourage public acceptance of technology and restore trust. For example, agencies search for ways to gain information about actual public preferences. Opinion polls are the most common technique, but in 1970 the U.S. Forest Service employed an instrument called "code-involve" to uncover public attitudes about its proposed use of DDT to protect Douglas Firs.[22] Code-involve is an applied content analysis system designed to transfer large quantities of information from written statements in diverse sources (e.g., editorials, petitions, letters) into a condensed form for policy review. Condensing scattered statements into clusters of opinions, the system offers a wider spectrum of attitudes than may emerge from other procedures such as public hearings, and the frequency of a given view as it appears in written statements provides a measure of its prevalence.

Some "quasi-experiments" attempt to use systematic observations of the actual behavior of individuals as evidence of policy choices. For example, to assess the social impact of power plants, investigators measure actual reactions to siting policies as an indicator of public preferences.[23] Participation in such experiments is random and mediated by the procedure: The only questions answered are those that are raised by the researchers and their sponsors. Nevertheless, social research techniques are increasingly exploited as a surrogate for direct involvement—as a way to increase public influence while avoiding the problems of representation inherent in many participatory procedures.

A number of experiments employ mediation procedures. Modeled after industrial-labor negotiations, they are intended to promote constructive dialogue among contesting parties to a controversy. The process is based on the voluntary participation of groups concerned about a project; they meet to debate the issues with a third party who facilitates arbitration of the dispute. Mediation procedures have helped to reach settlement in several environmental cases.[24]

Another participatory procedure emerged from the recombinant DNA dispute when citizen review boards formed to advise the city councils in several communities on policies for allowing the research to go on. The first of these, in Cambridge, Massachusetts, was organized on the principle that decisions about risks and benefits are questions not for scientists but for laymen, and that laymen are able to face technical matters, educate themselves, and reach just decisions.[25] This review board, composed entirely of non-scientists, discussed complex questions about the risks of DNA research for four months and, with a modification requiring local monitoring, approved the research under the Federal guidelines set down by the National Institutes of Health. A similar citizen review structure has been proposed by a study group at the Oak Ridge National Laboratory to resolve nuclear siting disputes.

SOME EUROPEAN MODELS[26]

Several European experiments are worth noting. In the Netherlands, an elaborate public inquiry system was developed on the principle that the public must be consulted on all decisions affecting the environment. Government plans are preceded by the publication of "policy intentions" dealing with political and philosophical questions: the objectives of growth, the goals of particular technological projects and their likely impacts. These are widely distributed for public comment. Local governments organize discussion groups, information evenings, and photo and sound shows on the plan. Television programs present alternatives and encourage people to send in their written comments. Reactions are analyzed by a representative advisory group, and the appropriate minister must answer criticism and either reformulate or justify his policy. The entire dossier developed through this process serves as a basis for parliamentary decisionmaking.

The Austrian government set up a series of structured public debates between scientists with opposing views on nuclear policy. The purpose was to highlight controversial dimensions of the nuclear program and to clarify the areas of persistent disagreement. This was mainly intended to inform parliament but also to engage public reaction.

In 1974, disturbed by the growing anti-nuclear movement, the Swedish government initiated an experiment in public education. Using an existing system of study groups managed by the principal popular organizations and political parties, the government financed a program to inform broad segments of the public about energy and nuclear power. The program involved some 8,000 study circles, each with about 10 members who met together to discuss those energy-related questions they felt to be most important.

The Burgerdialog in Germany, first organized in 1974, represents a similar effort to involve broad sectors of the public in an information program. Organizations such as churches, unions, and adult education groups have been funded to organize discussion groups and meetings which include speakers both for and against nuclear power. The goal is to strengthen confidence in the democratic process and to restore public confidence in the administrative authorities.

These information efforts are for the most part intended less to ascertain public opinion than to inform citizens about the value of specific technology policies and to convince them that risk is minimal. Informing the public is equated with participation, even though its aim is to create more favorable attitudes. While in most cases opposition persists, such experiments are important to observe for the insights they provide into diverse ways of accommodating greater public involvement.

SPECIAL PROBLEMS OF PARTICIPATION IN SCIENCE AND TECHNOLOGY POLICY

Democratic ideology requires participation in decisionmaking by affected interests. But in the arena of science and technology there are special problems in expanding political choice—for citizens seeking to influence policies involving complex technical material, for bureaucracies responsible for efficient technological development, and for scientists concerned about the progress of research.

PROBLEMS OF CITIZEN INFLUENCE

Citizen involvement in science and technology policy depends on many of the same factors that affect political participation: leadership, community organization, access to the media, and the visibility and urgency of the immediate issue. However, other factors, such as the availability of information and the distribution of expertise, assume special importance in this policy area. In order to have any real influence in the process, citizen groups must cope with complex and often uncertain technical material. And they must have access to experts who can help to analyze the material and maintain control of the information and grasp of the policy choices. Thus access to technical resources is a key problem for citizen groups.

A related problem lies in the complex network of decisions that contribute to technical policies. Citizens seeking to influence nuclear energy policy, for example, must focus their resources on very diverse areas: Federal research policy, licensing procedures, fuel supply, and waste management decisions. But also they must try to influence state regulations on utility rates, permit procedures, Federal and state environmental legislation, and local government land use regulations. Participation on all these fronts is costly, especially since those who intervene must be able to present an informed and technically valid case.

The fuzzy boundaries between the technical and political dimensions of science policy compound the difficulties of expanding public choice. Too often political issues, such as the tradeoffs between risks and benefits, are defined as technical, and technical questions of feasibility become confused with political acceptability.

Finally, administrative ambivalence may limit public choice. Agencies control the agenda of hearings and select the members of committees. The timing of public involvement—often late in the decision process—puts citizens at a disadvantage where they can only react to proposals already accepted and even already underway.

ADMINISTRATIVE CONCERNS

Participation presents administrative authorities with a dilemma. They have likely judged the propriety of their policies according to their contribution to general economic growth, only to discover that their activities are perceived as allied to special interests. They have likely assumed that legitimacy would flow from their technical competence, only to find critics insisting that legitimacy must rest on public consent. Expertise remains central to insuring the efficiency of administrative and legislative procedures, yet deeply rooted democratic values hold that public consent is a "right," and that procedures including more direct citizen involvement are intrinsically "better."

Administrative agencies try to avoid political conflict by following procedures that are limited to those who share common assumptions. Demands for involvement by citizens' groups with different assumptions disrupt procedures, preclude accommodation, and raise political dilemmas. Participatory measures, it is feared, will encourage endless demands from private interests or from self-appointed public representatives who lack the technical knowledge necessary to assess decisions.[27] This will lead, they fear, to impossible expectations for risk free technologies and unresolvable polarization: "environment vs. jobs," "local costs vs. regional benefits," "individual choice vs. public interest." There is concern that such participants will seriously obstruct the decisionmaking process, causing costly delays and precluding the long-range perspectives necessary for effective technical planning. Indeed, those who manage the scientific and technological enterprise feel constant tension between the ideal of participation as a source of legitimacy and its practical consequences.

THE IMPLICATIONS OF PARTICIPATORY POLITICS

> Scientist, regulator, lawyer, and layman must work together to reconcile the sometimes conflicting values that underlie their respective interests, perspectives, and goals.[28]

As governments fumble to accommodate participatory demands in different political and administrative situations, difficult questions remain. What form can lay participation take in complex technical areas? Who should be involved? What are its implications for the representative process? And, of particular concern to this report, what are the implications of greater lay involvement for the application of resources to research, for the strength of the research institutions, for the way scientists choose their topics and conduct their work, and for the general climate—political, social and intellectual—in which research takes place?

Perceptions of the consequences of greater participation reflect quite different political perspectives on the nature of the scientific enterprise and the appropriate relationship between science and the citizen. Some argue that the layman, unable to deal with the complex technical material necessary to evaluate science and technology, must leave responsibility to the scientists. Others claim that the political and moral implications of science and technology are matters of central public concern that require political, not technical, competence: hence the citizen must be involved.

Many scientists have interpreted demands for participation as an indication of anti-science attitudes. They fear that lay involvement would threaten their autonomy and bring external controls that would infringe on freedom of inquiry and virtually paralyze the research process.[29] However, experience suggests that even the critical public does not want to stop either science or technology, although particular projects may be under attack. Lay involvement in the NIH advisory review boards has not destroyed the grant review system; study sections continue to review the scientific merit of individual proposals and control the quality of research. Although the existence of institutional review boards to monitor human experimentation is still viewed by many scientists as an "unwarranted intrusion" that impedes research, lay participation in these boards has not significantly obstructed research. Indeed, non-scientists have been generally less active and less influential than scientists.[30] Even the Cambridge Review Board supported the continuation of recombinant DNA research under the NIH guidelines, clearly accepting the premises of the scientific community about the value of research. And, in the end, the intense public concern over this issue abated, while control remained with the NIH as before.

Attitudes toward participation also remain ambivalent among those with political concerns. Some criticize participatory procedures as simply another means to maintain the existing structure of influence. Participatory procedures are easily co-opted; token participation can simply reinforce the status quo, providing means for informed consent rather than expanding democratic choice. Others suggest that participation provides a means to protect minority interests and to include a wider range of issues in the policy agenda. These arguments develop in all policy areas—and as science and technology assume growing social and political importance, they are no exception. Democratic ideals imply that science policy be subject to greater public scrutiny and political control. Indeed, a fundamental source of public concern about science and technology is their implication for political choice. Accommodating these demands of democracy while maintaining the best of scientific learning is a key and growing challenge that will warrant careful attention over the coming years.

NOTES

1. For a review and analysis of these conflicts, see Dorothy Nelkin, ed., *Controversy: Politics of Technical Decisions*, Beverly Hills: SAGE Publications, 1979.

2. Mendelsohn, Everett, and Weingart, Peter, "The Social Assessment of Science: Issues and Perspectives," In Mendelsohn, E., Nelkin, D., and Weingart, P., eds., *The Social Assessment of Science*. Conference Proceedings, University of Bielefeld, 1978.

3. Opinion Research Corporation, *Attitudes of the U.S. Public Towards Science and Technology*, Study III (September 1976), p. 51.

4. The above data on public attitudes are from the National Science Board, *Science Indicators 1976*, USGPO (1977), pp. 168-182. Similar ambivalence shows up in European surveys. See Commission of the European Communities, *Science and European Public Opinion*, Brussels, 1977. A more recent survey focusing specifically on medical research shows greater confidence in this field than others. Most respondents felt medical research has changed life for the better (very much better—59%, somewhat for the better—34%). But 49% agreed that in deciding what to do, scientists are more concerned with their own research interests than with public benefits; National Commission for the Protection of Human Subjects, *Special Study*, DHEW Publication No. (OS) 78-0015, 1978.

5. Nelson, Richard, *The Moon and the Ghetto*, New York: W.W. Norton, 1977.

6. Gianos, P., "Scientists as Policy Advisors: The Context of Influence," *Western Political Quarterly* (September 1974), pp. 429-456; Benveniste, G., *The Politics of Expertise*, Berkeley: Glendessary Press, 1972.

7. Meynaud, Jacques, *Technocracy*, London: Faber and Faber, 1968; Ellul, Jacques, *The Technological Society*, New York: Alfred A. Knopf, Inc., 1970.

8. Noble, D., *America by Design*, New York: Alfred A. Knopf, 1977; Winner, L., *Autonomous Technology*, Cambridge: MIT Press, 1977; McCrae, D., "Science and the Formation of Policy in a Democracy," *Minerva* (April 1973), pp. 228-242.

9. Primack, J. and von Hippel, F., *Advice and Dissent: Scientists in the Political Arena*, New York: Basic Books, 1974; Benveniste, G., *op. cit.*; Lakoff, S., "Scientists, Technologists and Political Power," In Rosing, I. S., and Price, D., *Science, Technology and Society*, Beverly Hills: SAGE Publications (1977), pp. 355-392.

10. Klaw, Spencer, *The New Brahmins: Scientific Life in America*, New York: Morrow, 1969; Lapp, R., *The New Priesthood*, New York: Harper and Row, 1965.

11. For example, see Nelkin, Dorothy, *Nuclear Power and Its Critics*, Ithaca: Cornell University Press, 1971; Nelkin, Dorothy, *Jetport*, New Brunswick: Transaction Books, 1975. For the propensity to protect when unable to avoid problems, see Hirschman, Albert, *Exit, Voice and Loyalty*, Cambridge: Harvard University Press, 1970.

12. Alford, R., *Health Care Politics*, Chicago: University of Chicago Press, 1975.

13. Mazur, A., "Disputes Between Experts," *Minerva*, XI (1973), pp. 213-262.

14. See, for example, "Ann Arbor Science for the People Editorial Collective," *Biology as a Social Weapon*, Minneapolis: Burgess, 1977.

15. Carpenter, Richard, "Information for Decisions in Environmental Policy," *Science*, 168 (12 June 1970), pp. 1316-22; and Casper, Barry M., "Scientists on the Hill," *Bulletin of the Atomic Scientist* 33, November 1977, pp. 8-15, for discussion of scientific advice to Congress; Fox, Harrison, Jr., and Hammond, Susan W., *The Invisible Force in American Lawmaking*, New York: The Free Press, 1977.

16. For a general discussion of this problem see Lowi, Theodore, "Four Systems of Policy, Politics and Choice," *Public Administration Review* (July/August 1972), pp. 298-309; and La Porte, Todd, "The Study of Public Organizations," In Marini, Frank, *Towards a New Public Administration*, Scranton: Chandler, 1971, p. 20.

17. Scott, William, "Organizational Government: The Prospects for a Truly Participatory System," *Public Administration Review* (January/February 1969), pp. 43-53; Cramton, Roger, "The Why, Where, and How of Broadening Public Participation in the Administrative Process," *Georgetown Law Journal* 60:3 (February 1972), p. 529. See also Gellhorn, Ernest, "Public Participation in Administrative Proceedings," *The Yale Law Journal* 81:3 (January 1972), p. 367.

18. A review of these problems appears in Wade, Nicholas, "Freedom of Information," *Science* 175, 4 February 1972, pp. 493-502; and Kolata, G., "Freedom of Information Act: Problems at the FDA," *Science* 189 (4 July 1975), pp. 32-33.

19. An unpublished report by J. Sullivan, cited in Boasberg, Tersch, "Implications of NSF Assistance to Non-Profit Organizations," *Report to NSF*, Washington, D.C. (1977), p. 64.

20. Gray, Bradford, Cooke, Robert A., and Tannenbaum, Arnold S., "Research Involving Human Subjects," *Science* 201:22 (September 1978), pp. 1094-1104.

21. For a review and bibliography on citizen litigation see Di-Mento, J., "Citizen Environmental Litigation and Administrative Process," *Duke Law Journal* 22 (1977), pp. 409-452. For the difficulty of finding conceptual correspondence between scientific developments and legal concepts, see Carroll, James, "Participatory Technology," *Science* 171 (19 February 1971), pp. 647-653.

22. Clark, Roger, and Stankey, George, "Analyzing Public Input to Resource Decisions: Criteria, Principles, and Case Examples of the Codeinvolve System," *Natural Resources Journal* 16 (January 1976), pp. 220-222.

23. Murray, B. et al., "Peaking Plant Environment Report (EAC, September 1973)," reported in Heberlein, Thomas, "Some Observations on Alternative Mechanisms for Public Involvement: The Hearing, Public Opinion Poll, the Workshop, and the Quasi-Experiment," *Natural Resources Journal* 16 (January 1976), pp. 197-212.

24. For reviews of environmental mediation efforts, see the EIA Review published at MIT.

25. Cambridge Experimentation Review Board, "Guidelines for the Use of Recombinant DNA Molecule Technology in the City of Cambridge," December 21, 1976.

26. For an extended review of these models, see Nelkin, Dorothy, *Technological Decisions and Democracy*, Beverly Hills: SAGE Publications, 1978.

27. See discussion in U.S. Congress, Senate, Committee on Government Operations, *Public Participation in Government Proceedings Act of 1976, Hearings*, Washington, D.C.: GPO, 1976.

28. Bazelon, David L., "Risk and Responsibility," *Science* 205 (20 July 1979), p. 278.

29. See *Daedalus*, March 1978, devoted to the theme of Limits of Scientific Inquiry, and the Nobel Symposium, Volume 44, *Ethics for Science Policy*, Pergamon Press, 1979.

30. A study of IRBs found that, although accepting the legitimacy of the IRBs, 25% of the 940 biomedical scientists and 38% of the 395 behavioral and social scientists interviewed felt the review procedures are an unwarranted intrusion on the investigator's autonomy; 43% of the biomedical scientists and 54% of the behavioral scientists felt the procedure had impeded the progress of research. National Commission for the Protection of Human Subjects, *Report and Recommendations on Institutional Review Boards*, DHEW Publication (OS) 78-0008 (September 1978), p. 75.

2 Information Resources and the New Information Technologies: Implications for Public Policy

*by Donald A. Dunn**

SUMMARY

This paper is concerned with policy issues in the information sector of the world economy. Changes in information and telecommunication technologies have led to growth of the information sector from about a quarter to a half of the U.S. economy in the last three decades.[1,2,3] These technologies are among the most rapidly changing technologies in use today. Television and the computer occupy our time and attention at home and at work in ways and to an extent that would have been difficult to imagine 30 years ago. And the telephone has become much more intensively used as the cost of long distance calls has dropped.

One of the driving forces for change in information technology has been the evolutionary, steady reduction in cost in the computer and communication industries. As the cost of a computation has dropped, new information products, such as the pocket calculator, have come into being and have become almost universal tools in business and the home. Although we are now reaching some fundamental physical limitations in such fields as VLSI (Very Large Scale Integration), cost reductions are expected to continue.

Information specialists expect a wide range of new applications for information technology because the hardware and software costs for these applications are already below the values necessary to make them economically viable. In these new applications, delays in implementation are a result of institutional and regulatory obstacles rather than hardware and software costs. Examples of new industry-oriented systems are EFT (Electronic Funds Transfer), EMS (Electronic Message Systems), POS (Point-of-Sale fund transfer and inventory control systems), and office automation. Systems tailored to the home market also are evolving, parallel with the industry-oriented systems. Home computers are now widely available at prices below $1,000, and devices to permit these computers to link with each other through the telephone network (as business computers have done for many years) are available at prices below $200. Video recording technology, scrambler systems for pay-by-program television, and several new forms of information storage-and-retrieval systems that will use the home television set as a display terminal are available today at modest cost.

New "stand-alone" systems, such as the pocket calculator, raise few new policy issues. The markets that supply these systems in the United States are unregulated and competitive. The United States has produced a large fraction of the world supply of these systems up to this time, but Japan and Western Europe are now gaining market share in this field, and the development of a

* Professor of Engineering-Economic Systems, Stanford University

national competitive strategy for the future has become a significant concern.[4,5]

Information systems combining computer power with communication offer a variety of services that stand-alone systems cannot provide. One category of such systems is person-to-person communication systems that transmit messages, store them in a computer-controlled storage medium, and offer random access to any message from a list of messages arranged and displayed by sender name or some other indexing arrangement. Such electronic message systems can transmit messages through the telephone network or a broad-band network such as a satellite communication system. Input terminals can be conventional computer terminals with alphanumeric keyboards or communicating, w[illegible] processing typewriters. Voice message systems of this type are not yet available commercially, except in the simple form of a voice recording unit that does not offer random access or a displayed list of messages. The other broad category of computer-communication system is storage-and-retrieval systems that make use of common, continuously updated data bases. Examples are airline or hotel reservation systems, EFT, and POS systems. These network-based systems do typically raise national policy issues, partly because the telecommunication networks that they use traditionally have been provided by firms or governmental entities that operate in markets in which the government has imposed limits on entry and controlled prices.

As the use of the telephone network in providing new information services becomes more widespread in the coming decades, its economic significance will increase, and high quality telephone service will become an even more significant national resource than it is today. In many countries of the world, the telephone network is managed by the same governmental entity that manages the postal service. In these countries, the telephone industry often has been used as a source of funds to subsidize postal service, and the quality of telephone service has deteriorated. The United States is therefore fortunate that it has subsidized its postal service from general revenues rather than telephone revenues. The United States is also fortunate that it enters the coming decades with a strong position in the world market for information services and information products, because every indication is that this will be a major growth sector of the world economy in the 1980s and 1990s, just as it was in the 1960s and 1970s.

Two broad classes of policy issues arise from the increasingly widespread use of new information technologies. One has to do primarily with business interest in using such information services as office automation, POS, and EFT in the production of other goods and services. The other has to do with consumer interests, both in the use of information services such as EFT that are an adjunct to the consumption of various noninformation services, and in the use of home information services such as EMS. The policy issues in managing new information services for business use primarily involve establishing national management systems that will allow industry to grasp new opportunities in this field. Policy concerns with respect to consumer interests involve both new opportunities and protection for consumers against the potential misuse of information and new information services.

In order to see these issues clearly, it will be helpful to examine the structure and operation of the information industries, starting with the creation of new information, proceeding through communication and the storage and retrieval of information in libraries and related institutions, and then considering the use of stored information in creating new products and services supplied through the marketplace. Note that this process of information creation, communication, storage and retrieval, and use in further creation is an integrated process in which all the steps are closely linked to each other. It is therefore essential to an understanding of information policy to examine this entire process, even though new information technology has greater direct impact in some points of the process than others.

Many information policy issues arise out of the peculiar nature of information, when it is viewed as an economic good. Information has many of the properties of a "public good," i.e., a good that one user can consume without diminishing its availability or usefulness to another user. Information products or services typically involve a high fixed cost of creation and a low marginal cost to make the service available to each additional user. Because it is possible to copy information at a much lower cost than to create it, investors who might hope to profit from the sale of information will tend to underinvest in this type of activity, relative to their investment in other activities not subject to copying. It is this economic argument that underlies patent and copyright policy. Patents and copyrights create property rights in information and ideas with the objective of making investments in the creation of intellectual property more attractive, relative to investment in goods not subject to copying. Patents and copyrights not only protect the creator of new information from "piracy" by others seeking to make copies, but also create transferable pieces of property that can be sold in the market, thus enhancing the value of new information by making it more like other goods.

The same property rights in new information that provide incentives to create it also operate to encourage the wide distribution of this information. Publishers and manufacturers seeking to maximize profits will attempt to sell their products widely, although they will set prices higher and hence sell a smaller quantity than organizations that maximize sales. Property rights also play a role in privacy and access to information. An individual's rights to protect himself from intrusion and from misuse of personal information are appropriately viewed as property rights, and the scope and extent of these rights are the basic policy issues in privacy and access.

It is useful to view the people and institutions dealing with the creation, distribution, storage and retrieval, and application of information as making up the national information infrastructure. It is important to note that this infrastructure is mostly people who can use information effectively and only secondarily is it a set of retrievable facts. To take an example, U.S. investment in the exploration of space has yielded two things: (1) an enormous set of data, reports, and facts that relate to space and space technology; and (2) a set of people and facilities in universities, private firms, and NASA spaceflight centers who are capable of applying space technology to new problems. This combination of people, facilities, knowhow, and retrievable facts makes up the national information infrastructure in the space technology area. The retrievable facts represent one of the tools that the people working in this area make use of in designing new systems. These facts also can be valuable to people in other fields. It should be emphasized, however, that these facts and the information storage-and-retrieval systems in which they are stored represent only a small fragment of the total information and capability that is available in this field. A comprehensive approach to the production and use of knowledge in space-related fields would be concerned with the management of the entire set of people and facilities.[6]

The focus here is much more limited. We first deal with some of the policy issues that arise in connection with the creation of new information, primarily scientific and technical information. Then, rather than dealing with the national information infrastructure as a whole, we deal with the narrower policy issues stemming from communications and information storage-and-retrieval services and their use. A broader, integrative view of this field could help develop a national information policy for the future. This integrative work remains to be done. Some suggestions for future work in this area are given in the concluding section on a national information policy. Viewed as a national resource, the information infrastructure is clearly the most important resource we have. However, managing it is more like managing the entire economy than managing a more conventional resource such as oil. Therefore, it remains to be seen whether a national information policy can be put together and, if it can, whether it will be a helpful way of thinking about this broad set of issues.

THE CREATION OF NEW INFORMATION

Policy options regarding the creation of new information are discussed here in relation to: (1) patents and copyrights; (2) public investment in artistic activities, basic research, and applied research and development; (3) encouragement of industrial investment in these activities through tax incentives and other approaches; and (4) consumer interests in the rate and direction of research and development. While a wide variety of cultural factors affect a society's orientation toward art, science, and technological change, economic factors play a central role in determining the rate and direction of change. The following discussion will consider only economic factors, emphasizing changes in science and technology.

Patents and Copyrights

When we think of the individual inventor or creator of a new work of art, it is easy to see the economic effects of granting a patent or copyright to this individual. The patented invention or copyrighted work is protected against copying for some period of years and is thus made more valuable and more readily sold. This increased value creates an incentive for further investment in innovation and invention.

There is an apparent tension between the policy objectives of obtaining a high national level of creativity and the policy objective of obtaining rapid dissemination of the results of the creative process. The policy instruments, such as copyright laws, that have been used to encourage creativity do so by creating barriers to copying and apparently act as obstacles to rapid dissemination. However, the tension is primarily a tension between short- and long-run objectives. In the short run, an innovation can perhaps be most rapidly disseminated by allowing free access to it. But in the long run, it is necessary to be concerned not only with dissemination of known ideas, but also with the continued creation of new ideas, so there will be something to disseminate. Patents and copyrights encourage both innovation and the disclosure of innovation. The alternative, allowing free dissemination, results in almost total secrecy about innovations, which obviously does not promote dissemination. Even under a property right system many innovations such as computer software are not protected, and innovators often go to considerable lengths to keep their ideas secret.[7,8]

The effects of patent laws on the operation of a modern, competitive industrial market can be rather different from the effects on individual inventors. In modern industry, the invention process has been commercialized. Inventors are hired and organized to create new ideas that will be most beneficial to the firms that employ them. In some markets the innovation process has been accelerated to a very high pace. The computer industry, for example, has a rapid development cycle, typically less than five years for a major innovation. A rapid obsolescence of products naturally accompanies this rapid introduction of new products. Five-year-old computers may work very well, but their value is only a small fraction of their purchase price.

An important distinction needs to be made between the invention process that may be involved in creating a new product and the innovation process that is concerned with selecting the specific characteristics and technology of the new product and bringing it to the marketplace. Many innovations are not patentable, but innovation is protected by trade secret law and by the length of time it takes to copy a new product. In a high technology field, the time to copy may be more than a year. A firm that is a year or two behind its competitors may find that its competitors have written off the costs of creation by the time its product reaches the market, so it does not gain a price advantage through copying. In such a market, copying would not be a successful strategy. The role of patents in such a market is unclear. Patents on basic inventions that will be used in several cycles of innovation have long-term value. Patents on obsolete products are obviously not of value. The usual argument that firms will underinvest in innovation does not seem to apply to rapidly changing, high technology markets. Firms in these markets must innovate in order to survive.

Firms can effectively nullify the effects of patents by entering into cross-licensing agreements. Firms, in effect, give up the potential rewards from occasional basic patents in order to avoid the risk of competitors' inventions blocking their access to the market. There is a risk that cross-licensing and patent pools can violate the antitrust laws,[9] but if all new entrants to an industry can join the licensing agreements, the effects are not anticompetitive.

The economics of invention and innovation in markets with rapidly changing technology appears to be an important field for research.[10] Neither the operation of such markets without government intervention nor the effects of patents and cross-licensing agreements in such markets is now well understood.

Public Investment in the Creation of New Information

As an alternative to creating property rights in new information through patents and copyrights, direct public investment can be made in the creation of new information and works of art In areas in which the government has a mission responsibility, as in defense and space, it can be expected to support the research that it believes will be most beneficial to its missions in the long run. In areas in which the private sector is responsible for providing products and services to consumers, there is also a potential role for government-supported research, especially basic research. The economic argument that firms will underinvest in research that leads to inventions subject to copying is even more applicable to basic research aimed at understanding nature, because patents do not cover theories or laws of nature. Thus, the discoveries that come from basic research will benefit a firm's competitors as much as the firm itself (except for public relations benefits), so the amount of basic research done in the private sector will tend to be less than is socially optimal.[11,12] Some form of governmental intervention in the market to create increased incentives for carrying out basic research is therefore appropriate, and direct government funding is a straightforward way to support basic research.

Once government funding of research is adopted as a national policy, a question arises with respect to the ownership of patents and copyrights on innovations made in this research. Presumably, the national interest is best served by a government patent policy that will maximize innovation. Government ownership of patents results in disclosure, but it does not create incentives for firms to make the necessary investments to bring these patented innovations to the market. Granting exclusive rights to firms that do make such investments would enhance the incentives to develop these innovations, much as homestead rights have been used to encourage the development of government land.

Another important policy issue in this area is that of the allocation of funds. What areas of research should receive funding, and at what levels? A balance of many diverse interests is somehow achieved in the present system. However, there may be opportunities for improving the present system: for example, by creating more independent sources of research funding likely to support research leading in new directions. Both industry and mission-oriented agencies could strengthen their positions in the long term by supporting basic research projects of special interest to them, rather than relying on others to provide this support.

Industrial Investment in the Creation of New Information

Industrial investment in research can be increased through tax incentives. However, there is the risk that the amount of new research may be small in relation to the amount of tax subsidy, because firms have an incentive to reclassify existing activities to qualify for favorable tax treatment as well as to initiate new research.

Also important is the possibility of more industry-sponsored research, on an industry-wide basis, in universities, industrial research labs, or research institutes. There are likely to be more cases like VLSI arising in the future, in which it is important for an entire industry to develop a new set of techniques that will be used throughout the industry in the future. Projects to develop these techniques could appropriately be funded and managed by the concerned industries themselves, without governmental intervention. Industry cooperation in such research programs could, however, have antitrust implications, and it is possible that new legislation would be helpful in encouraging this type of industry-wide research activity. Most of the precedents for industry-wide research occur in regulated industries, and even then the precedents are mixed. For example, Bell Laboratories serves the operating companies of the Bell System, but it does not directly serve independent telephone companies. The Electric Power Research Institute (EPRI) is an industry-wide organization in the electric utility industry. Until recently, the electric utilities relied primarily on General Electric and Westinghouse to do the research and development in electric power, with only a small industry-wide research program. Now EPRI receives 1 percent of the revenues of the electric utilities to use for research.

The most important feature of industry sponsorship of research, in my opinion, is that industry would inject its own ideas and interests into research that it sponsors. Development and design engineers from industry have a significant contribution to make to the direction of basic research that could most effectively be expressed in an organizational context in which industry is paying the bill.

Consumer Interests in the Creation of New Information

The principal consumer interest in research and development is selection of projects that will lead to products and services of the greatest value to the consumer. When the market is functioning properly, no special attention to this problem is required. But the market often fails to function properly, as a result of governmental intervention or some other cause. In such cases, research and development can come to emphasize the interests of producers rather than consumers. Examples can be found in nuclear energy, military hardware, and medical technology, all of which rely on nonmarket forces to determine project selection.

Two approaches to this distortion of the market are possible. First, an attempt can be made to increase competition and to allow producers a wider latitude in their choices of the approach to be taken in meeting consumer needs. Second, increased participation of consumers in the project selection process can be encouraged. For example, consumers can influence future project selection by evaluating the results of previous projects or existing services. In order to be effective in representing consumer interests, an evaluation would have to be carried out entirely under consumer control.[13,14]

Current moves toward deregulation in many industries offer significant opportunities for benefits to consumers from the more rapid introduction of new products and services. Regulation often has had the effect of slowing technological change, sometimes with the intention of benefiting consumers through reduced prices. But by keeping prices low, regulators have limited the rate at which new technologies could be introduced. The telephone industry provides a good example of this effect. The rate of introduction of digital switching technology into the U.S. telephone plant is still extremely slow.[15] It is not clear that the consumer benefits more from this policy than would be the case if prices today were higher and the newer technology were made available sooner. Similar effects have been caused by restrictions on entry into regulated markets by firms that would force a more rapid pace of technological change.[16,17] Again, the telephone industry provides a good example. Since 1968, the interconnection of customer-provided terminals, PBXs, and on-premises wiring to the telephone network has been allowed under FCC and state rules.[18] The pace of innovation in this market has dramatically increased since this policy was adopted in 1968, and American business now has a much wider range of options for telephone service open to it than could reasonably have been expected if pre-1968 policy had been continued. Many of the innovations in this field have been introduced by the new, unregulated interconnect industry, but many innovations are a result of telephone company response to competition.

It is not difficult to find examples of large-scale systems that seem to be evolving in ways that are not beneficial to consumers. The suggestion from economic theory is that the firms responsible for these systems tend to adopt "inappropriate" technologies because their economic incentives are distorted by the influence of nonmarket forces. For example, the choice of new technology in hospitals is not based on consideration of economic efficiency. Cost-reimbursement systems of payment combined with a lack of effective competition result not only in high costs but in inappropriate choices of technology.[19] The study of the economics of innovation in such systems could help in the development of deregulation plans and other approaches to the management of these systems that could bring about a closer alignment of producer and consumer interests.

COMMUNICATION SERVICES

Communication, information storage and retrieval, and computation can be regarded as the three main functions of computer-based information services. A typical transaction service, such as credit card verification, will employ all three functions. Here we discuss communication both as a component function of other systems and as a type of end service. The most significant innovations in communication services in recent years have been in person-to-person services rather than broadcasting, and many of these innovative services are data communication services: i.e., the data being transmitted are characters rather than voice or video signals. Of course, any type of signal, whether characters, voice, or video, can be put into digital form and transmitted as digital data.

Most of the data communication services now available, and most of the future computer-based information services, such as EFT, POS, and electronic message service (EMS), will use the telephone network as part of their communication networks, because the telephone network already interconnects most homes and offices in the United States. Because the telephone companies are regulated in the United States, which implies restrictions on entry of new firms into their markets, a number of policy issues arise regarding which firms can or should be allowed to offer which services. Difficulties arise principally because telephone service is not priced in proportion to its costs. Therefore, telephone companies object to competitors that might compete only in services that the telephone companies now price well above cost. The competitors object to telephone company competition where the telephone companies are pricing the service below cost.

There are several additional issues concerning the use of transmission and switching systems that are not part of the telephone network, but that could compete with the telephone companies for business, especially voice traffic. The telephone network can be thought of in simplified form as consisting of two parts: (1) local distribution, including switching; and (2) interexchange, or long haul plant. Communication satellites, coaxial cable, and microwave relay systems are the present long haul technologies. Optical fibers are beginning to be used. The cost of local distribution has changed very little in the last 30 years, while the cost of long haul plant has decreased steadily over the same period and the prospects are good for a continuation of this trend. These cost trends have been reflected to some extent in reductions in the price of long-distance calls, but local calls are heavily subsidized from long-distance revenues.

Competition exists in the provisions of interexchange service between AT&T's Long Lines Department and a number of competing firms, none of which yet has a significant market share. Local distribution is now exclusively by wire pair, and there is no significant competition in the local distribution market. However, the possibility exists of using radio communication for local distribution. Proposals to use portions of the FM radio band and portions of the band near 10 GHz for digital radio transmission are now being considered.[20] Bands in the 2 to 3 GHz region have already been allocated to Multipoint Distribution Service (MDS) that could be used for local distribution of digital radio signals in competition with the local telephone plant. Current trends in U.S. communication policy favor the introduction of competition and will probably result in a growing number of choices for users, both among technologies for local and interexchange service and among firms capable of providing basic communication services. These competing firms are now called specialized carriers, but in the longer term, it may well be that they are allowed to operate on an essentially unregulated basis because they will be operating in a competitive market. Legislation now before the Senate would essentially deregulate these interexchange carriers, although they would still be called carriers.[21]

Electronic Message Service

AT&T's Long Lines Department and the specialized carriers can be viewed as "underlying" carriers, because they offer the basic transmission facilities that must underlie any data communication system. Packet-switched data networks, such as those provided by Tymnet and Telenet, make use of lines provided by underlying carriers and connect these lines to computers at the network nodes that do the switching. Users can connect directly to such a network at a node or go through a local telephone line to the nearest node. Access to a variety of information services, such as time-shared computing or bibliographic search service, can be obtained through data networks of this type.

A particular class of information service, electronic message service (EMS), is of particular interest because of the policy issues it raises. EMS is a form of message service in which a user types a message on his terminal, edits it, and transmits it to one or more addressees. The message is then stored in a computer. When an addressee is ready to receive messages, he indicates this to the computer and a list of messages waiting for him is provided. The texts of individual messages may be requested, read, and filed if desired for future reference. EMS has been available as an adjunct to time-shared computing service for a number of years, on both Tymnet and Telenet, and also on Arpanet. As much as half of the traffic on these networks can be attributed to EMS. The user group that uses these networks is a relatively homogeneous group with many shared interests in computer-related topics. Such users have found that being able to send each other short, typed messages is a valuable adjunct to the telephone and regular mail service. It now seems likely that this type of service will grow very rapidly in the next few years and that businesses that are not otherwise computer users will use it widely. Studies of offices suggest that the most important opportunities in office automation lie in improving executive communication through such techniques as EMS and facsimile.[22]

The policy issues raised by EMS stem from the form of the 1934 Communications Act and its interpretation to the effect that communication services must be regulated, while noncommunication services must not be regulated.[23,24] Thus, the issue arises whether EMS is a communication service within the meaning of the act or whether it is some form of data processing. If it is a communication service, all providers of this type of service must be regulated and existing carriers will be able to offer it. This result is not desirable, because it requires regulation of services for which the usual justifications for regulation are missing. Except for the threat of regulation, this type of service would be available today from a wide range of unregulated competitive firms. New legislation may be necessary to deal with this issue, and pending legislation before the House and Senate would deregulate this class of service (see note 21).

This class of EMS message services, EFT, POS, and facsimile together provide a capability to transact much of the world's business electronically, on a terminal to terminal basis, without the need for the use of any hand-carried mail. For some years to come it will be more economical to use hand-carried mail for transmitting long documents rather than facsimile. But it is evident that over the next decade or two we will see a shift from the use of hand-carried mail to electronic mail for a large fraction of mail carrying short, transaction-oriented messages. The need for a postal service, as we now know it, will undoubtedly diminish. However, as has been the case with many other new media, the coming of electronic mail will undoubtedly only partially replace hand-carried mail. Some hand-carried mail will be originated as electronic mail. Mail can be sent by electronic means to the post office nearest the recipient and hand-delivered from that office to recipients that do not have appropriate electronic mail terminals. Mailgram is an existing service of this type.

The policy issues suggested by this shift from hand-carried to electronic mail will strongly affect the future of the U.S. Postal Service.[25] The private express statutes give the postal service a monopoly over some kinds of mail service, but presumably not electronic mail. Clearly the postal service is free to use electronic mail internally. It is also free to offer various forms of hybrid service, analogous to Mailgram, in which messages are transmitted to the postal service electronically or in the form of computer tape and converted to hand-delivered letters. However, the postal service could also, potentially, establish a new national electronic communication network and seek to provide EMS on both a terminal-to-terminal basis and otherwise. The EMS industry views this possibility with alarm, because the post office could easily engage in a type of cross-subsidy of which the telephone companies have been accused, and offer EMS at prices below cost.[26]

TELECONFERENCING AND THE OFFICE OF THE FUTURE

Teleconferencing can be defined as any telecommunication system that allows three or more persons to communicate. Computer teleconferencing is a straightforward extension of EMS that allows a group of persons working on a common project to send messages to each other and to the group that can be read at the convenience of each user rather than in real time. There have been several studies of this type of conferencing, and it appears that there often are situations in which this type of conferencing can be extremely useful.[27]

The usual forms of teleconferencing occur in real time, with some of the participants in remote locations. Telephone teleconferences are the cheapest and simplest type. A simple extension is audio teleconferencing, in which a conference room is equipped with microphones and loudspeakers that allow the group of conferees in the conference room to speak naturally, yet have their voices carried to remote users and allow remote users to speak to the conferees in the conference room. A useful addition to audio teleconferencing is high-speed facsimile, so documents can be exchanged during the conference. A further addition that remains to be invented is an electronic blackboard that would allow conferees at two or more locations to share the same writing space. Various devices are now in existence that approximate a shared writing space, but they are still poor substitutes for what is needed.

The extension of audio to video teleconferencing is straightforward but expensive in terms of long-distance line costs. It is doubtful that it is worth the cost. If an adequate way of exchanging still visuals can be made available to audio conferees, audio teleconferencing may turn out to be adequate for most applications. One estimate from the research is that audio and computer teleconferencing can substitute for about 45 percent of all business trips to meetings, whereas video adds only about 8 percent more substitution possibilities.[28]

Whatever forms of teleconferencing eventually become accepted, it is evident that teleconferencing—when combined with EMS, facsimile, and improved computer-based information storage-and-retrieval systems—will provide a major increase in the communication capabilities available to the modern office worker. Some jobs probably can be done perfectly well by workers in decentralized locations. For example, workers d[illegible]ing with the public probably can do just as well in locations easy for the public to reach as in large central city office buildings. The possibility that workers will go to work in small office complexes within biking distance of their homes or will work at home has been widely discussed and studied in the context of the office of the future.[29] Improved telecommunications could, therefore, result in new settlement patterns in which the central city plays a reduced role as a means for facilitating person-to-person communication.

CROSS-SUBSIDIES AND BARRIERS TO ENTRY

The basic idea of subsidizing some types of services, either through cross-subsidies from other services or through the erection of barriers to entry to certain classes of protected businesses, is a familiar element of national telecommunications and information policy, worldwide. The economic consequences of this type of policy rarely have been examined in detail. It is clear that there are costs, and that the costs are borne by the nation's businesses, and ultimately by consumers generally, in the form of increased communications costs. But the effects also can appear in the form of service quality deterioration. In some countries telephones are not available except on payment of a substantial sum, comparable to the full capital cost of a line (about $1,000), and then only after a long wait. Such countries are likely to view electronic mail as a new luxury item to be taxed, rather than as a new service that can make their businesses better able to compete in world markets. Generally speaking, consumer and business interests (other than the interest of the subsidized industry) coincide and favor an open market in information services with as few cross-subsidies as possible.[30]

One type of cross-subsidy in communication services can be beneficial to most users. This subsidy is one from heavy users to light users of the service. It is beneficial to all users to have as many users on the system as possible, in order to increase the number of users with whom communication is possible. In order to induce marginal users to join the system, the fixed cost of joining may appropriately be set somewhat below cost and the difference made up by charging higher prices per unit of usage. This policy helps to achieve universal service, often stated as an objective of the U.S. telephone industry. This pricing policy can be advantageous to providers of service, when the new marginal subscribers use additional optional services, such as long-distance service, and make a net positive contribution to costs. Countries that set the price of obtaining service initially at a price higher than cost are losing this benefit.

INTERCONNECTION

Another major issue in communication systems policy is interconnection. Interconnection can arise between networks, such as competing EMS networks or local telephone networks in adjoining areas, and it can also arise with respect to terminal equipment. Interconnection of networks enhances the value of service for users on both networks, so the consumer's interest is ordinarily enhanced by interconnection. At first glance, it appears that interconnection also would benefit network providers, who would be able to extract some of the benefits of interconnection in the form of higher prices. However, network providers often are engaged in a form of competition in which each hopes to become the dominant network provider. In such cases, providers will resist interconnection. Providers who are already dominant can use refusal to interconnect as a weapon against new entrants. Historical studies of the evolution of the U.S. telephone network provide examples of both types of behavior.[31,32]

Within the next few years the interconnection issue can be expected to arise with respect to EMS. It will be in the users' interest to have a basic capability to send messages to users in any EMS network from any other EMS network. It may be possible to provide a minimum form of interconnection that will allow this capability while allowing each provider of service to offer special features that add to the basic capability. New, small EMS providers will favor interconnection, and large EMS providers seeking to achieve a dominant market position will resist it.

INFORMATION STORAGE AND RETRIEVAL

Some information can be expressed in print form; other information requires photographs, motion video, audio, or the physical presence of a particular person. If information can be expressed in a medium subject to storage and retrieval, it can be made accessible to users at times and places of their own choosing.

A reduction in the cost of videorecording technology has made television into a medium, similar to books, that can be used on demand. Although the cost of professional television programming is so high that a program must be shared widely in order to reduce the cost per user to an acceptable level, this sharing need not take place through broadcasting, now that videorecording is available. A shift from broadcast viewing to individual viewing on demand, and various intermediate forms of "narrowcasting," is now beginning.

In the print media, the use of copying machines to produce copies of books and journals on demand has or may soon become sufficiently low in cost to permit a new approach to libraries. Rather than being a source of items to be loaned, a library is likely to become a source of items to be copied. Various storage media, such as microfilm and the optical disk, are adaptable to the automated production of low-cost copies. Once the cost of production of a single copy of a book or article on demand becomes sufficiently low, it will be more efficient to use the library as a storage and production center rather than as a storage and loan center. When this occurs, a reorganization of the publishing industry can be expected, especially in connection with the publication of low circulation items such as professional journals. From the standpoint of the user, such a change is likely to result in greatly enhanced access to the print media, increased use of the literature, and substantially increased expenditure on printed media and library services. Particularly interesting is that the critical cost at which it will become economically efficient to produce documents at the library is not just the library's cost of doing so. Rather, it is the user's cost: i.e., the sum of the price charged by the library plus the user's time cost. Obviously, the user will be willing to pay something to get a copy of a document immediately rather than later. Therefore, the library's price for producing a copy on demand can be substantially higher than the price that could be charged for future delivery from a publisher's stock of mass-produced copies that might cost less per copy to produce.

Perhaps the most significant changes are taking place as a result of reduction in the cost of computer memory. Computer memory costs are being reduced both (1) in very rapid access time systems that use magnetic and semiconductor memory and are suitable for heavily used files; and (2) in longer access time systems that are much less costly and will use the random access optical disk for less often used information. The optical disk appears to be the technology that this field has been looking for for many years.[33]

The storage and retrieval of information from computerized data bases is now one of the most pervasive forms of computer usage. For many purposes, it is extremely useful to have a single, continuously updated, master file of information. This file need not be located in a single location. Pieces of it can be distributed among a number of locations, but it is essential that the combined set of pieces of information form a single data base. This type of system is widely used in two rather different applications: (1) files of information on transactions that typically include data on individual users, such as the amounts of their purchases, their bank balances, etc.; (2) files of less heavily used information on subject matters of importance, such as company correspondence or published information of the sort found in libraries.

LIBRARIES OF THE FUTURE

Once information has been put into a medium subject to storage and retrieval, it can be stored in some form of library, and bibliographic entries that describe it in terms of its author and subject matter can be entered into one of the computer-based bibliographic search systems now widely available. The library of the future probably will include: (1) a single bibliographic search system that can be used to search nearly the entire collection of the world's literature; (2) immediate access to the full text of all heavily used references on the same computer system that is used to store bibliographic data; (3) access within less than an hour, by facsimile, to the full text of most other references that are indexed in the bibliographic data base. Certainly a user would appreciate these features. The technology for such systems is available now, but full text systems are still expensive. The present limitation on bibliographic search systems is the cost of manual indexing and data base maintenance, not the cost of hardware or software.

A single, standardized, bibliographic search system will probably not occur without governmental intervention in the market. Three approaches are possible: (1) subsidizing the process of reclassifying data entries not now in standard format or in standardized search systems; (2) making sure that, starting sometime soon, all new entries are made in a standardized system, leaving older entries in whatever nonstandard classification system was used when they were entered; (3) developing new computer techniques that will allow a user, in effect, to translate nonstandard search systems into a standard language.

The consumer stands to benefit from moves toward standardization, both of the format used for entries and of the software used to gain access to computer-based library systems.[34] There is a real possibility that a standardized system could be developed that would allow supplementary features to be added by specialized service providers, but that could be used without these features for general searches. Such a system would permit growth and improvement to take place, while maintaining simplified access to the bulk of the world's literature.

It is, of course, evident that once a library system of this type becomes available, it will reduce the incentive for individuals to have subscriptions to journals. At this point, the present system of publication of scientific papers may become obsolete. Users will prefer a system that allows them to seek out specific papers on specific topics of interest rather than browsing through journals, which mostly contain papers of little immediate interest. Of course, computer-based systems also will allow users to browse through the most recent papers in any specified field of interest when this is desired.

The mechanism for financing the library of the future is built into its structure. Computer-based bibliographic search obviously can be charged for in proportion to its use. Document retrieval can be charged for in proportion to the cost of locating the document and of providing a facsimile copy. Presumably there will be a national or international network of libraries that store documents and offer document retrieval service.

This type of library is protected against copying and reselling of both its full text documents and its bibliographic search file on a large scale by copyright law. Small-scale copying and sharing cannot reasonably be prevented. But such sharing will not be worth the trouble in most cases, because few users (other than students) will want copies of documents that their friends have at the same time as their friends do. In the long run, the cost per document will probably be too low to make sharing worthwhile.

In addition to these core services, most occasional library users will want personalized librarian service to provide assistance in using the computerized bibliographic search service. And a few users will want access to persons with specialized knowledge of the field being searched. A good library should be able to provide these additional services to users willing to pay their costs.

PRESTEL

One version of the library of the future is available in London today as Prestel.[35] Prestel is a system that uses a television set as a display terminal and uses a low-cost interface box that transforms digital signals from the telephone network into display signals. The user can use any one of several types of keyboards to transmit signals through the telephone network to a central computer that has a wide range of information available for use in its memory. The user is led through various index pages to the information of interest. Current news, stock and commodity market information, updated transportation schedules, and regularly updated white and yellow pages will probably be among the offerings. The importance of Prestel and other similar systems being planned and tested in the U.S. and Europe lies in their potentially very rapid growth and high ultimate market penetration. Because they are based on the use of the television set and either the television broadcasting network or the telephone network, they can be made available in most homes and businesses in the U.S. at very modest cost, and consequently in a relatively short time. The systems like Prestel that use the telephone network will be able to offer a very large number of different pages of information to their users. Prestel plans to offer 170,000 pages.[36] Broadcast systems like Ceefax (a derivative of See-Facts), also an English system, are limited to a much smaller number of pages, such as 100. Experiments using the broadcast network are in various stages of development in the United States: in Salt Lake City, St. Louis, and Philadelphia. The U.S. Department of Agriculture is sponsoring an experiment in Kentucky that will provide farmers access to weather, crop, and market information.

The crucial question in all of these systems is the extent to which people will want these services and be willing to pay for them. Only fairly large-scale market trials that allow time for producers to adapt to user interests will give reliable answers. The most interesting feature of Prestel is its adaptive, open market approach to the question of what information should be included. Instead of trying to decide this question by administrative decision, it is left to suppliers of information, who pay so much a page to have their information stored. Suppliers receive payment on a per page per use basis, and can set their own price per use, but they must pay the Prestel system a minimum price per page per use. Suppliers anticipating relatively low usage of their information by users willing to pay a substantial price would set their price at perhaps 20 cents per page per use. Higher usage pages might be priced at 2 cents per page per use.

One policy issue raised by such technologies is, are there regulatory obstacles to the offering of such services in the United States? If so, what action is necessary to permit systems of these types to develop? The pure television systems would require the cooperation and interest of the broadcasting industry plus some minor regulatory changes. Systems that use the telephone network, like Prestel, could be designed to use the local telephone network only or, if it were desired to establish a single national system, they could use packet-switched networks, such as Tymnet or Telenet. In any case, no special regulatory issues are involved in a strictly storage-and-retrieval system. However, once a system of this sort is available, a very low-cost additional option is to allow users to send typed messages to each other which are stored in the central computer and made available to the recipient on request. This electronic message service (EMS) does raise regulatory issues discussed previously. It is probably unrealistic to think of a Prestel-like system without an EMS capability, and in fact Prestel is already being used for communication among information providers in England.

Other policy issues arise in connection with the liability of information providers for false or inaccurate information. Libel laws will cover this situation, but the potential for very small firms or individuals incurring very large liabilities appears to exist in a way that is not possible in television broadcasting.

THE MARKET IN INFORMATION STORAGE-AND-RETRIEVAL SERVICES

There is a basic national interest in having high quality information storage-and-retrieval services available. Governments can contribute to the achievement of this objective either through direct government action or through policies that encourage the operation of the market. For example, government actions that might lead to standardization of bibliographic search systems would improve the operation of the market in bibliographic search services. The government directly affects this market by its approach to the funding and operation of libraries and similar institutions. Libraries are not like research. There is no intrinsic difficulty in charging users for library services, so the private sector will not underinvest in library services, and private sector underinvestment is therefore not a reason for governments to subsidize or operate libraries. However, historically, governments have been the principal funding sources and operators of libraries, partly on equity grounds and partly because libraries, like schools, have been seen as providing beneficial externalities: i.e., beneficial effects received by members of the community other than library or school users as a result of the use of the library or schools.

The main difficulty with government operation of libraries is the pricing policy that has usually been adopted. Users pay nothing for service, and both users and nonusers pay for the libraries through taxation. Users are thus subsidized by nonusers, but going with this subsidy is the inability of users to influence the services they receive, either by refusing to pay or by otherwise expressing their preferences. When legislatures make grants to libraries or schools, the libraries and schools naturally seek to satisfy the legislatures rather than the users. When the consumer pays directly, assuming there is no governmentally induced monopoly, the consumer can exercise a choice and patronize the library or school that offers the best service.

If there is a concern that some low income users may receive no service under a user-pay system of libraries, low income users can be given the subsidies rather than giving the subsidy to the libraries. Users could be given aid in the form of "library stamps" or scholarships.[37] Low income and other nonusers would obviously be better off with a user-pay system of libraries than with a tax-supported system that provides only services they don't want, because their taxes would be lower.

A further difficulty with tax-supported libraries is that they drive most private-sector library services out of the market, because the price to users is zero. Only private library services offering quite different services can coexist with a public library offering zero-price service. The scientific and technical user, as well as most other users, probably would be best served by libraries that charged prices to users proportional to cost. Government policies that provide payments to libraries through users and that encourage the operation of a market in library services would be beneficial to users, because a greater diversity of service offerings would be expected. Experimental programs designed to test these ideas could be implemented at very low cost.

PRIVACY

Many of the most significant information policy issues arise in connection with data bases that contain information on specific individuals. The principal dangers in this area involve misuse of data on an individual either by an individual or by an organization that is either not authorized to have access to the data or that may have a right of access to the data for one purpose but uses the data for another purpose.

Various legal safeguards are already in existence in the United States to protect individuals' files from unauthorized access by others and to provide individuals access to data concerning themselves, and further legislation is under consideration.[38] A number of foreign governments recently have passed new privacy legislation. Some of this legislation is concerned with data bases containing personal data on a nation's citizens that may be located in a foreign country.[39] Transborder data flows associated with such data bases are subjected to constraints intended to force international data service providers to maintain data bases in each country where they have customers. It is by no means clear that the result of such legislation is increased privacy for the customers, since their own governments will have easier access to their files.

Service providers have not yet offered many innovative services intended to increase privacy. However, as more financial services are automated, the market can be expected to present a range of privacy options to its users. The principal obstacle to the offering of privacy in the market will probably be governmental regulations, some of which probably will have been enacted to protect consumers' privacy. An important policy issue will soon arise as to the right of an individual to conduct his or her financial affairs in secret. A related issue will be the right of an individual to communicate by means of encrypted data flows. The issue will not be acute, as between individuals, but rather as between an individual and the government. The way this issue can be expected to arise became clear in connection with discussions concerning the adoption of a U.S. data encryption standard recently.[40] The proposed standard is believed to yield encrypted data that can be decoded with a computer that would cost in the tens of millions of dollars in 1980 and that would be readily available only to very large organizations. If system providers ignore this standard and offer encryption that cannot be decoded with even the largest computers available today, users will be able to achieve more privacy than some governments may wish. Similar secure data base services will also probably soon be available. Thus, this policy issue may soon be whether or not the market in privacy will be allowed to function.

A somewhat different privacy issue arises in relation to advertisers who wish to use various information media to reach potential customers. Some users of these media may wish to use the capability of the media to reject all advertisements. For example, the post office could offer a service to customers that wished it to dispose of all junk mail rather than delivering it. Similar services could be made available to users of electronic mail services of the future. The most critical area today for consumers is the telephone, because its use for advertising is potentially so much more intrusive and disturbing than the use of mail systems or even television. Telephone users would now almost certainly be subjected to electronically automated advertising messages on a large scale, if it were not for the intervention before Federal and state utility commissions of a few consumer-oriented individuals. Even though the issue is unresolved, a number of firms are already transmitting automated messages to homes.[41] The policy issue presented by automated telephone advertising is whether or not users have a right to choose whether they will receive such ads or not.

Privacy policy can be viewed as a piece-by-piece building up of the definition of an individual's property rights in his or her own personal data, communications, and connections to the national communication networks. A continuing examination of this set of issues will be required, as the new media become more widespread.

ACCESS

The other side of the privacy issue is access. Under what conditions shall the government, a corporation, or an individual have the right to unauthorized access to stored information about another person, corporation, or governmental activity? This issue is quite clearly delineated when it arises within the purview of a court, as a part of a legal proceeding. For example, the rights of the parties in an antitrust suit to gain access to each other's internal memos and other private documents have been litigated many times and are fairly well defined. Similarly, the government's right to access to private information when it obtains a search warrant is fairly clear. However, when the government collects information without a warrant and without notice to the individual, such as the telephone numbers called from a particular telephone or the names of payees and the amounts of checks written by an individual, the issue is less clear.[42]

A rather different kind of access issue arises in connection with access to new information created through investments in research and development. Such information may, of course, be classified in a military sense, but assuming it is not, should restrictions be put on access to such information by foreign nationals? In the United States, the transfer of some unclassified defense-oriented technology and information is made more difficult through the operation of the Trading with the Enemy Act. However, if information does not fall within the scope of this act and is unclassified, should it be made available to foreign competitors on the same basis that it is made available to U.S. industry? Should information only be made available to nations that adopt reciprocal policies?

Three alternative policies could be followed in this area: (1) attempt to impede the flow of new U.S.-originated information in science and technology to foreign nationals while not impeding the flow of information to U.S. industry; (2) seek to give U.S. industry special help in making use of new information of this type, wherever it originates; or (3) make no distinctions between users on the basis of their citizenship or their home country.

The first option is likely to work poorly in preventing determined seekers of information from getting access to whatever is available to U.S. industry. U.S. industry is now highly internationalized, both at home and abroad, and it is difficult to distinguish, in many cases, between U.S. and foreign companies. The first option will almost certainly impede the flow of information to U.S. industry, even though it would be intended to avoid this result. The costs associated with this option could turn out to be greater than the benefits.

The second option is worth exploring, and there may be some practical actions that could be taken. As noted, however, often the special skills and knowhow of industrial groups are more valuable than the facts filed in libraries. Recognizing this fact, foreign firms have acquired blocks of stock in U.S. high technology firms in recent years, at least partly in order to increase their access to this more subtle type of information.[43,44,45,46] Research and the development of specific proposals to enhance the quality of science- and technology-related information services for U.S. firms could be undertaken.[47] In addition, studies of foreign firms and their approaches to the use of research information might yield some new insights and ways to improve U.S. industry's use of information.

A further factor in preferring the second or third option to the first is that U.S. relations with countries that are not now competitive with the United States might be set back by the first option. The second option could be expanded to encompass the facilitation of transfers of information to developing countries as well as transfers to engineers within the United States.

In summary, the transfer and exchange of science and technology information on an international basis is a subject that requires a national choice. The President's recent science and technology policy message indicates a concern for this issue and an interest in providing assistance to developing countries.[48] Further research and analysis of these issues would be helpful in arriving at a national policy.

COMMERCIAL LAW

As we move away from the use of checks and other pieces of "commercial paper" and move toward electronically maintained financial records, a whole new set of legal rules will have to be adopted with respect to the responsibility of the parties for losses incurred in a transaction.[49] An important element of computer data bases is that all parties to a transaction will be able to gain access to the same data at the same time. Transactions will be completed without any waiting period during which pieces of paper are moved physically from place to place.

A wide range of new business opportunities will be created when users in most homes and offices have the ability to commit funds from terminals located in the home or office. Many transactions that are now simply too cumbersome will become easy. In particular, the possibility of a great many more auction sales will exist. Some new opportunities for the sale of perishable commodities will arise. For example, hotels with empty rooms may be willing to rent some of them at greatly reduced prices, and auctions of rooms at hotels throughout the world may be held on a continuous basis. The opportunity to make or lose a great deal of money in a short time will undoubtedly become more readily available.

The policy issues in this area will not be greatly different from those in other areas of business law, but it will be necessary to adopt new rules to deal with the new ways transactions will be made.

NATIONAL INFORMATION POLICY

The scope of national information policy has been suggested by the topics treated above, but many related topics also could have been included: education, radio and television broadcasting, newspapers, mobile radio, freedom of information, international treatment of the news, and consumer information. The proper scope of national information policy has not yet been delineated. However, there are good reasons for drawing the boundaries around at least the topics treated here. The creation of new information is closely connected with the use of existing information and the ways that existing information is stored and retrieved. And storage and retrieval are closely connected with business transactions and personal communication, because much of the same hardware and software is used by the same people in all of these types of service. Once the field has been defined to include these areas, we see that we have included what we know as a people, what we are doing to learn more, and the tools that we use to conduct individual transactions and to communicate on a person-to-person basis.

There is a value in bringing together the ideas and issues involved in this set of national activities, because doing so calls to our attention their interrelatedness and importance in our lives. There are opportunities for improving the operation of the systems that provide information services in ways that will lower prices and increase the number of users served, the quality of service, and the privacy and security offered to users. Many of these opportunities will be enhanced by taking an integrated view of this area. Only a beginning has been made in this report toward providing this type of integrated view.

The perspective that has been emphasized has been economic. The policy option that has been suggested in most situations is to limit governmental action to the definition of property rights designed to encourage the operation of the market. This same approach has been suggested in patent and copyright, privacy, interconnection, and the management of communication systems generally. Certain situations in which government regulation or government provision of services traditionally have been operative offer opportunities for improved service to consumers through deregulation or the adoption of user-pay pricing policies. In some cases, the optimum policy is unclear, and research is needed to formulate a new policy. The principal types of new legislation suggested are laws aimed at deregulation of markets. These suggestions are in the same spirit as the legislation now pending before the House and Senate subcommittees on communication.

Three broad policy objectives for national information policy that probably would achieve wide acceptance are: (1) to enable public and private organizations to make full use of available information resources; (2) to encourage innovation in the application of technological information resources to the solution of national problems; and (3) to encourage establishment of demonstration user networks: The intention of this paper has been to call attention to the possibility of achieving these objectives through increased reliance on the private sector and the operation of the market. In most situations, research and analysis are required before definitive policy proposals can be put forward. I believe that it is important that any such analysis give adequate consideration to the possibilities suggested here, in addition to the more traditional approaches that have been taken in the United States, and most other countries to date.

NOTES

1. Machlup, Fritz, *The Production and Distribution of Knowledge in the United States*, Princeton, N.J.: Princeton University Press, 1962.

2. Bell, Daniel, *The Coming of Post-Industrial Society*, N.Y.: Basic Books, 1973.

3. Porat, Marc U., "Communication policy in an information society," In Robinson, Glen O., ed., *Communications for Tomorrow*, N.Y.: Praeger Publishers, 1978.

4. Garwin, Richard, Testimony before the Subcommittee on Science, Technology and Space, Senate Committee on Commerce, Science, and Transportation, Feb. 10 and April 26, 1978, Part 2, Serial No. 95-77, 1978.

5. Baruch, Jordan J., *Symposium on Domestic Policy Review of Industrial Innovation*, Washington, D.C.: U.S. Dept. of Commerce, Jan. 15, 1979.

6. The Space Industrialization Act of 1978, H.R. 14297, Bill introduced by Mr. Fuqua, 95th Congress, Oct. 12, 1978.

7. Nycum, Susan H., "Legal protection for computer programs," *Computer/Law J.* 1 (1978), pp. 1-83.

8. Shapley, Deborah, "Electronics industry takes to 'potting' its products for market," *Science* 202 (Nov. 24, 1978), pp. 848-849.

9. Turner, Donald F., "Patents, antitrust, and innovation," *U. of Pittsburgh Law Review* 28 (1966), pp. 151-160.

10. Phillips, Almarin, "Patents, potential competition, and technical progress," *Am. Econ. Rev.* 56 (May 1966), pp. 301-310.

11. Nordhaus, William D., *Invention, Growth, and Welfare: A Theoretical Treatment of Technology Change*, Cambridge, Mass.: MIT Press, 1969.

12. Mansfield, Edwin, "Research and development, productivity change, and public policy," *NSF Colloquim on R&D and Productivity*, Washington, D.C., Nov. 1977.

13. Maynes, E. Scott, Morgan, James N., Vivian, Weston, Duncan, Greg J., "The local consumer information system: an institution to be?" *J. Consumer Affairs* 11 (Summer, 1977), pp. 17-33.

14. Dunn, Donald A. and Ray, Michael L., "Local Consumer Information Services," Report No. 16, Program in Information Policy, Engineering-Economic Systems Dept., Stanford, Calif.: Stanford University, April 1979.

15. U.S. v. AT&T, Plaintiff's first statement of contentions and proof, Civil Action 74-1698, D. Ct., Washington, D.C. (1978), pp. 448-451.

16. Posner, Richard A., "Taxation by regulation," *Bell J. Econ. and Mgmt. Sci.* 2 (Spring, 1971), pp. 22-50.

17. Capron William M., ed., *Technological Change in Regulated Industries*, Washington, D.C.: Brookings Institution, 1971.

18. U.S. F.C.C., Use of the Carterfone Device in Message Toll Telephone Service, 13 FCC 2d 420 (1968), Reconsid., 21 FCC 2d 153 (1970).

19. Fuchs, Victor R., *Who Shall Live?* New York: Basic Books, 1974.

20. Before the FCC, Xerox Corp., Petition for Rulemaking in re Amendment of Parts 2, 21, 87, 89, and 91 of the Rules for the establishment of a new common carrier Electronic Message Service (EMS) in the band 10.55-10.68 GHz, Nov. 16, 1978.

21. Communications Act Amendments of 1979, S. 611, Bill introduced by Mr. Hollings, Mr. Cannon, and Mr. Stevens, 96th Congress, 1st Session, Mar. 12, 1979.

22. Bair, J. H., "Communication in the office of the future: where the real payoff may be," Proc. 4th Intl. Conf. on Computer Communication, Kyoto, 26-29 Sept. 1978, pp. 733-739.

23. Dunn, Donald A., "Limitations on the growth of computer communication services," *Telecommunications Policy* 2 (June 1978), pp. 106-116.

24. U.S.F.C.C., Amendment of Section 64.702 of the Commission's Rules and Regulations, Docket 20828, Notice of Inquiry, 61 FCC 2d 103 (1976), Tent. Decision and Further Notice of Inquiry and Rulemaking, FCC 79-307; July 2, 1979.

25. U.S. Commission on the Postal Service, Report of the Commission on the Postal Service, Washington, D.C.: The Commission, April 1977.

26. Nyborg, Philip S., "Regulatory inhibitions on the development of electronic message systems in the USA," *Telecommunications Policy* 2 (Dec. 1978), pp. 316-326.

27. Johansen, Robert, "Social evaluations of teleconferencing," *Telecommunications Policy* 1 (Dec. 1977), pp. 395-419.

28. Harkness, Richard C., et al., *Technology Assessment of Telecommunications/Transportation Interactions*, Report to NSF on Contract NSF-C1025, Menlo Park, Calif.: Stanford Research Institute, September 1976.

29. Ibid.

30. Dunn, Donald A., "National and international policy issues in computer communications," Proc. 4th Intl. Conf. on Computer Communication, Kyoto, Sept. 26-29, 1978, pp. 209-214.

31. Gabel, Richard, "The early competitive era in telephone communication," *Law and Contemporary Problems* 34 (Spring, 1969), pp. 340-359.

32. U.S. v. AT&T, Plaintiff's statement of contentions and proof, Civil Action 74-1698, D. Ct., D.C. (1978), pp. 118-178.

33. Kenney, George C., et al., "An optical disk replaces 25 mag tapes," *IEEE Spectrum* 16 (Feb. 1979), pp. 33-38.

34. Schierer, Helen F., "Bibliographic Standards," In Cuadra, C. A., Luke, A. W., and Harris, J. L., eds., *Ann. Rev. of Info. Sci. and Tech.* 10, Washington, D.C.: Am. Soc. for Info. Sci. (1975), pp. 105-138.

35. Reid, Alex, "Flying fast but safe," Special Survey: V deotex, *Intermedia*. London: Intl. Inst. of Comm. (May 1979), pp. 22-25.

36. Ibid.

37. Friedman, Milton, *Capitalism and Freedom*, Chicago, Ill.: Univ. of Chicago Press, 1962.

38. *Personal Privacy in an Information Society*, Report of the Privacy Protection Study Commission, Washington, D.C.: U.S. Govt. Printing Office, July 1977.

39. Wilk, Charles K., ed., *Selected Foreign National Data Protection Laws and Bills*, OT Special Publication 78-1, Office of Telecommunications, U.S. Dept. of Commerce, March 1978.

40. Diffie W. and Hellman, M.E., "Exhaustive Cryptanalysis of the NBS Data Encryption Standard," *Computer* 10, No. 6 (June 1977), pp. 74-84.

41. Crock, Stan, "If the phone rings and its Zsa Zsa, don't be surprised," *Wall Street J.*, p. 1, Aug. 15, 1979.

42. Reporters Committee for Freedom of the Press v. AT&T, 593 F. 2d 1030 (D.C. Cir. 1978), Cert. Den. 99 S.Ct. 1431 (1979).

43. Amdahl, Gene, Testimony concerning financing of Amdahl Corp. by means of Japanese venture capital, Field Hearings, Senate Committee on Commerce, Science and Transportation, San Francisco, Calif., Oct. 30, 1978.

44. "A. G. Siemens completes formation of microcomputer company with Advanced Micro Devices, Inc., purchases 400,000 shares of Advanced Micro Devices for $18 million," *Wall Street J.* (Jan. 20, 1978), p. 17.

45. "A. G. Siemens, owner of 80 percent of common stock of Litronix, Inc. plans to offer $8 million for remaining shares," *Wall Street J.* (Nov. 8, 1978), p. 36.

46. "Northern Telecom, Inc., U.S. subsidiary of Northern Telecom, Ltd., Canada, plans to boost its stake in Intersil, Inc. to 20 percent from current 11 percent," *Wall Street J.* (May 31, 1977), p. 36

47. The National Technology Innovation Act of 1979, S. 1250, Bill introduced by Mr. Stevenson, Mr. Hollings, Mr. Inouye, Mr. Ford, Mr. Riegle, Mr. Moynihan, and Mr. Schmitt, 95th Congress, May 24, 1979.

48. Message from the President of the United States transmitting a Science and Technology Policy for the Future, House Document No. 96-81, 96th Congress, 1st Session, U.S. Govt. Printing Office, Mar. 27, 1979.

49. Baxter, William F., Cootner, Paul H., and Scott, Kenneth E., *Retail Banking in the Electronic Age: The Law and Economics of Electronic Funds Transfer*, Montclair, N.J.: Allanheld Osmun & Co., 1977.

3 Communication of Scientific and Technical Information: Implications for Federal Policies and Research

*by Melvin Kranzberg**

SUMMARY

The explosive growth of scientific publication following World War II led to information "overload"; scientists and engineers complained of difficulty in keeping up with the latest developments. But they also complained of information "underload," the inability to find the specific information they needed.

To meet this "overload-underload" paradox, the field of information science emerged. Attention became focused on the literature-machine interface and the man-machine interface as new technologies, especially computers, were introduced. But an informal information network exists outside of, while supplementing and complementing, the formal network. Interpersonal communication—through "invisible colleges" and "technological gatekeepers"—remains a major channel for dissemination of information.

To meet the above and other problems, it has been proposed that a national coordinating body be created to develop and implement policy in information services. That proposal is neither feasible, desirable, nor necessary. Technical elements of the information system have advanced rapidly in the private sector, so there is little need for governmental direction or subsidy.

Various measures might be taken to extend and improve the formal communications system, such as: basic research on modelling the information system; strengthening of storage-and-retrieval systems (formation of disaggregated information subsystems); development of a comprehensive catalog of available data bases; and government funding of translation of abstracts.

Also, utilization of NTIS by all Federal agencies; research to assess the impact on information flow of secrecy versus patenting decisions; and the requirement of an "Information Assessment" statement in connection with any proposed changes in patent policies should be considered.

Further, government research should be encouraged on how such computer-communications combinations will affect the informational system as should regulatory agency review of impacts on science-technology of corporate actions on information-communication developments.

Meanwhile, progress in informational technology opens new capabilities for scientific-technical research. There should be further investigation of (1) real-time information systems which allow a user to engage the data directly; (2) the implications for scientific research and information systems of computerized communication devices for solving scientific problems directly; (3) increased utilization of informational capabilities in serv-

* Callaway Professor of the History of Technology, Georgia Institute of Technology.

ics; and (4) wider usage and participation in information systems by social scientists and the public.

THE INFORMATION "OVERLOAD" AND "UNDERLOAD" PARADOX

"Water, water everywhere, but not a drop to drink." The plaint of Coleridge's Ancient Mariner expresses the information supply-and-demand dilemma which faced scientists and engineers during the quarter century following World War II. The exponential growth of publications so deluged practicing scientists and engineers that even a well-informed scientist or engineer experienced great difficulty in keeping up with the latest developments.

Paradoxically, in the midst of this surfeit of information, the working scientist or engineer complained of difficulty in obtaining the specific information needed for his research. In brief, he suffered from "information overload."

THE IMPLICATIONS

The publication flood which inundated the scientific-technical community became a matter of vital national concern with the recognition that scientific and technological advances depended, in large measure, on the flow and dissemination of useful information. Empirical evidence demonstrated a positive relationship between the innovativeness of firms and their openness to outside information.[1] Furthermore, those fields exhibiting the greatest degree of innovativeness—such as pharmaceuticals, chemicals, instrumentation, electronics—were precisely those which exploited the new information technologies most successfully.

THE RESPONSES

There were three responses to the paradox of too much information but not the right information at the right time or place. One was the development of a new field: information science. This field is to be distinguished from science information, which is the product of every scientific discipline; information science is concerned with how that information is produced, organized, communicated, and used.

A second response was the development of secondary information devices: abstracting and indexing services to reduce the vast number of publications to digestible form. Third was the computerization of information systems. This new hardware technology proved marvelously efficient and rapid in storing, locating, and retrieving information.

However, these new elements added two new problem areas to the information-flow process: the literature-machine interface, and the man-machine interface. Although more sophisticated methods of abstracting and coding lubricated the literature-machine interface, scientists still complained of difficulties in the information flow. Perhaps that was because these improvements responded to the needs of the storage-and-retrieval specialists but not to those of the information users. This last point was crucial, because the end of an information system is not simply the accumulation of information, but its use.

THE SYSTEMS APPROACH

In their zeal to cope with the publication flood, information scientists sometimes forgot that the problem is not so much scientific publication as scientific communication. Scientific communication is a systems problem, involving both the generators of the information and its users.

Scientific communication is not confined solely to the printed word (or formula); informal, person-to-person communication networks allow for easier and quicker communication. These formal and informal networks complement one another; for example, the scientist/engineer seeking information first informally consults another individual, who refers him to a bibliographical reference that brings him into direct contact with the most useful part of the formal system.

Failure to adopt a systems approach led to neglect of the informal network and its relationship with the formal system. As a result, inadequate attention was given to systemic elements external to the problem of incorporating information in electronic devices where it would be quickly accessible if the *seeker knew where and how to search for it*.

The last proviso hints at a number of systemic elements that militate against the most effective use of current information systems. First, the sheer bulk of the information intimidates potential users. Second, there is the "translation" problem; that is, structuring the information along lines familiar and meaningful to the users. Other external problems concern proprietary information and its security when it is encoded in a computerized network.

Still another shortcoming of the existing information system derives from the fractionated structure of 20th-century science. Specialists seldom are familiar with the literature of neighboring fields, and their vocabulary often is incomprehensible to them. These disciplinary "blinders" hinder the flow of useful knowledge between fields.

Another factor is that much of the scientific research reported in foreign-language periodicals remains closed to American scientists and engineers.

These shortcomings represent only a few of the systemic elements affecting the generation, transmission, and use of scientific-technical information. Nevertheless, informational systems have made great strides during the past two decades.

RECENT PROGRESS—AND PROBLEMS

TECHNICAL INNOVATIONS

Only two decades ago, an electronic computer cost $200,000, weighed three tons, had some 2,000 vacuum tubes, and required ten tons of air conditioning to keep it operable. Today the same work can be done by a three-chip, nine-ounce, programmable calculator, costing less than $300. The number of memory elements per semiconductor chip has doubled every year for the past few years, and the semiconductor industry expects to continue that rate of improvement. Such rapid progress in microelectronics has not only enlarged the capacities of information-processing devices but also has lowered their cost. It is now within the realm of probability to conceive of computer terminals so inexpensive that one could be located at every desk—leading to the possibility of a totally machine-readable information technology. Indeed, individual on-line computer terminals for information applications grew from 500,000 units in 1972 to more than 1,500,000 by 1977—a 33 percent annual increase.[2]

Secondary information services, such as abstracting and indexing, also have become increasingly sophisticated. The electronic hardware enables these services to handle larger quantities of scientific literature and to access, organize, identify, and locate literature very quickly.

In some cases, professional societies (e.g., the American Chemical Society) are systematically organizing knowledge in their domains, offering abstracting and other information services to members. Private industry also has entered into the information business (e.g., Lockheed's DIALOG), and government agencies have developed large-scale information systems (e.g., NTIS, SSIE, NASA, and the National Library of Medicine). In addition, some universities and firms maintain information services and are tied into national data systems.

There has been technical progress along other lines, too. Telecommunications widens the scope for disseminating and networking science information; micrographics allow the printed record to be duplicated in compact form and distributed very inexpensively. New information technologies, such as View-data and Teletext systems, are also emerging.

Such rapid technical progress should make furthering the dissemination of scientific-technical information simple. It does—but another major element must be considered in the system: the human factor.

THE HUMAN LINK

All studies of the flow of information in scientific research and technological innovation highlight the critical importance of person-to-person communication. Within the scientific community this task is performed by "invisible colleges"; that is, small groups of highly productive scientists train students, set research priorities, and monitor the rapidly changing knowledge in their fields. Not surprisingly, these scientists are aware of the latest advances long before printed publications reach those scientists who have not yet "matriculated" at their "college."

Interpersonal communication also remains foremost in technology, but somewhat different arrangements prevail. "Technological gatekeepers" act as channels for bringing external knowledge into the industrial research and development (R&D) laboratory. Other informal means by which individuals acquire and transfer essential information for innovation are the movement of people from firm to firm, from government to academia, to industrial laboratory, and back and forth; and attendance at professional meetings and trade shows.

These informal networks do not function separately from the formal information networks of publication and storage-and-retrieval systems; rather, they are interdependent, with the informal system frequently diffusing the formal system's information. How does one effect optimal interaction of the systems?

The question is difficult because elements of both the formal and the informal systems are being rapidly transformed. For example, Garrett Hardin ("Will Xerox Kill Gutenberg?" *Science*, December 2, 1977) predicted that copying devices would put an end to the five-century domination of printed publications. Of course, the copying machine is not the only villain. Computer-linked reproduction systems, microfilms, teletype systems, and the like also are contributing to the demise of the book and journal. Though journals and books may disappear, the printed word, even though it may take computerized or microfiche form, will remain the essential vehicle for information transmission for a long time to come. Partly this is because the reward system of the scientific community measures a scientist's contribution in terms of published research papers. Thus, the obsolescence of the book and journal does not mean the end of the printed publication, but only its transformation into electronic modes.

The withering away of the traditional printed mode of information dissemination is offset by the flowering of other means. Despite its shortcomings, the machine-man interface is growing stronger, as indicated by increasing use of computerized systems. Lower cost contributes to this increased use, but it is also a result of the increasing familiarity of the scientific and technical community with the new information technology. Younger scientists are more accustomed to the computer as an everyday tool than are their older counterparts. Thus one of the human barriers to the use of computerized scientific and technical information services will soon disappear.

Most attention has been focused on improvements in the hardware elements of the information network, and little thought has been given to the informal person-to-person network. Can it be made more effective in disseminating information?

Here the news is both good and bad. There has been great growth in the number of meetings, conferences, trade shows, and other gatherings which facilitate face-to-face communication. However, this phenomenon might be short-lived. Inflation and transportation (energy) difficulties might discourage attendance at such gatherings. Because of financial problems besetting universities, academicians would be the first affected, but the mobility of those in industry and government also would be constricted.

To keep the "invisible college" in session and the "technological gate" open, there would have to be more Federal travel grants for attendance at scientific and technical meetings, both domestic and foreign.

Electronic substitutes for face-to-face communication offer some hope for maintaining the informal network. Still in the demonstration stage is teleconferencing (also called "computerized conferencing"). Using computer terminals to provide written discussions for meetings, it also includes conferencing by means of television. Although demonstration projects have shown teleconferencing to be less expensive than mail or long-distance telephone communications,[3] its development has lagged, largely because human beings apparently prefer face-to-face meetings. Teleconferencing simply does not offer the amenities and background of companionship provided by informal arrangements at professional meetings. Nevertheless, rising transportation costs and time savings might ultimately force wider use of this adjunct to the informal network.

Although it is technologically feasible, teleconferencing thus faces the problem of user reluctance. If that were overcome—and increased familiarity might help—teleconferencing would have other consequences besides facilitating information exchange. It could have dramatic impact on the degree of centralization and decentralization in organizations; allow handicapped and disadvantaged individuals to participate more fully; and facilitate the formation of new information networks without large expenditures of time, money, and energy. Continued funding of demonstration projects would familiarize individuals with this new communications technology and the requisite behavioral changes preparatory to its extended future use.

INSTITUTIONAL LINKAGES

Other linkage problems affecting the flow of scientific/technical information involve the tri-cornered relationships among academia, government, and industry. That between academia and industry rests upon a division of labor: the university carries on long-term, curiosity-oriented basic research, and industry applies the knowledge thus derived.

But innovation is not a unilinear progression from basic research through development to application, as the classical model would have it. Instead, it is an ecological system wherein basic research often is an integral part of development and application.[4]

The changing role of R&D in innovation dictates changes in industry-university relations, with concomitant changes in their information exchange. One suggestion for more fruitful interaction is that government encourage industry to support longer term, continuing research in universities, covering all states of research short of actual proprietary commercialization. Different universities would be devoted to different industrial sectors, facilitating information exchange among scientists and technologists concerned with the same problems, and revolutionizing institutional relationships.[5]

Government-industry linkages are equally important. Sticky problems arise concerning the ownership—and hence the dissemination—of information produced in the private sector under government contract, as well as questions about the role of patents in encouraging or inhibiting information flow.[6]

Disseminating the results of R&D carried on in government laboratories also poses problems. Government agencies provide little useful information in generating or applying innovations, perhaps because Federal R&D is usually so mission-oriented that it ignores possible uses outside the original purpose. The lukewarm success, at least in its early stages, of NASA's technology utilization program has been attributed to the fact that the needs of those not involved in the aerospace industry were not considered in developing the innovations.[7] On the other hand, the model of the National Advisory Committee on Aeronautics and of the agricultural extension services indicates that when the information is user-oriented (and often user-generated), technology can be transferred successfully.

This list by no means exhausts the institutional linkage problems. There are communication difficulties within industrial laboratories[8] and even among scientists and engineers in different departments within the same university. Interestingly enough, computerized information systems sometimes can surmount these institutional and personal hurdles, allowing the formal network to be more effective than the informal communication system.

POLICY OPTIONS

Many different government policies, even those not specifically related to science and technology, affect the flow of scientific and technical information. They include such items as support for scientific research and education, patent laws, regulations regarding advertising and disclosure of product information, programs for information exchange, antitrust and tax laws, and the like. Obviously it is impossible to analyze every government action, present or projected, in terms of its implications for scientific and technical communications. Attention therefore must be focused on those policy options that impinge most directly on information flow.

The impact of government policies on information flow differs from one industrial sector to another, depending on such items as the technological level ("high" or "low"), the extent and nature of past government regulations, and the size of the business units involved.[9] As a result, the policy options discussed below are those tending to have the broadest bearing upon scientific and technical communications, irrespective of the specific industries involved.

Finally, the National Science Foundation (NSF) itself had commissioned two reports dealing with policy options: Joseph Becker, "A National Approach to Scientific and Technical Information in the United States" (NSF Contract C963); and Bruce G. Whalen (Mitre Corporation), "Scientific and Technical Information: Options for National Action" (NSF Contract 3-76-SP-1002). The options and recommendations embodied in those documents also must be considered.

The policy options and recommendations discussed below are those (1) expected to have the most powerful influence on scientific and technical communication; (2) covering the broadest spectrum of economic, political, and cultural interests; and/or (3) specifically recommended for science/technology communication policies.

National Coordination

The Becker and Whalen reports, submitted to the National Science Foundation in 1976, recommended the creation of a national policymaking body for science information. Its duties would include coordination of national science/technology information activities, developing policy for the disposal of scientific/technical information developed by the government, developing strategy for international information exchanges, setting priorities for science information programs, and establishing funding levels.

In the few years since these reports were prepared, it has been asked whether a new organizational mechanism for establishing and coordinating Federal policy regarding science/technology information is necessary, feasible, or desirable. The public mood is opposed to the creation of still another Federal body. In addition, the Office of Science and Technology Policy (OSTP) has taken on the task of guiding the scientific/technical enterprise to meet national objectives of innovation and productivity, including responsibility for scientific and technical information policy. It is questionable whether an additional bureaucratic apparatus is necessary for the informational subset of scientific activities.

Also, the technical components of the information system—speedy storage-and-retrieval mechanisms, computer terminals, national networks, and the like—have been progressing apace under the aegis of the private sector, without direction or subsidy from the government. In other words, the technical means for rationalization of the system are falling into place without the formation of a national policy or coordinating body.

Within the NSF, the Division of Information Science and Technology encourages research in the field of information science itself, and the Division of Policy Research and Analysis (created since the Becker and Whalen reports) is available to analyze its policy implications. Although not aimed directly at all aspects of the information system, these and other government programs seem adequate to cope with the problems without creating still another coordinating panel or policy board.

Finally, there is the danger of "overcoordination" which might lead to the exclusion of promising alternatives. In such a fast-moving field, which offers so many exciting and different potentials (as does scientific/technological communications), there is need for pluralistic approaches and even some redundancy. As Martin Landau has stated, "If there is no duplication, if there is no overlap, if there is no ambiguity, an organization will neither be able to suppress error nor generate alternative routes of action."[10] Coordination is desirable to prevent unnecessary and costly duplication, but not at the expense of a healthy pluralism.

Extending Formal Systems

Computerization of information storage-and-retrieval mechanisms has led to "faster and faster librarians." Now attention must be paid to the entire information system, including the man-machine linkages and the behavioral problems involved therein, as well as the producer-user linkages.

Employing a systems perspective, Russell Ackoff and his colleagues at the Wharton School developed a planning methodology in SCATT (Scientific Communication and Technology Transfer Systems).[11] Identifying different "stakeholders" in the science/technology information community, SCATT places them within the context of a national system. This approach marks a first step in a systems analysis of the informational organism.

Because of the many variables in the science/technology information system and because information processing and dissemination are evolving, it is essential to detect changes in the system and to chart possible weaknesses in the linkages. Hence the NSF should continue to support basic and applied research on dynamic models of the science/technology information system.

Any such research should give particular attention to the behavioral aspects, for the "human node" in the information transfer process has been somewhat neglected. Although there is a body of knowledge regarding the processes of human cognition, memory, and learning, we need information directed at specific problems encountered at the interface with machine information systems. Models of the system can indicate where further research and action might be necessary.

STORAGE-AND-RETRIEVAL SYSTEMS

Technical progress in storage-and-retrieval and abstracting and indexing services has outstripped advances in the informal network. The problem then becomes one of matching the capabilities of the formal system with the needs of the informal network, thereby making the information system more effective.

As in all other fields, scientific and technical communication is stimulated when those possessing common interests get together to share their information. Such informal networks can be strengthened if they are organized to take advantage of integrated computer-communication networks.

However, information needs differ from field to field, and even at different steps in the research and innovation processes. So instead of "coordinating" information systems to provide "standard" information reports, it is important to make the system correspond to the different structures of scientific fields and also to the different mind-sets and informational habits of their practitioners. Overzealous standardization might impose the mind-set of information science specialists on scientific and technical fields where such habits of thought might not be applicable.

The sophisticated hardware of today's computerized information networks allows great flexibility. All electrical networks, especially computer/information networks, can be a force for unifying control and/or dispersal for little systems comprehended within bigger systems. Instead of demanding uniformity, the hardware allows diversity with reasonable consistency. These diversities should be fostered because they correspond more to the varying habits of work and thought of different scientific and technical communities.

The government should encourage segments of the scientific/technical community that have not yet taken advantage of the latest technical devices and research in science information to develop the structure, forms, and means for more effective information dissemination within their respective communities. This superstructure could then be linked to larger systems in order to take advantage of useful external information, to effect economies of scale, and to employ the most appropriate technology for indexing, storage and retrieval, and dissemination on a national scale.

SOFTWARE PROBLEMS

By definition, the software system is "mushier" than that of the hardware. It is also costlier. Handling the increased volume of information has escalated information costs[12] despite decreases in hardware costs. Although the cost picture could change as we move toward a machine-readable system, there still remain problems in improving software operations.

One of these problems is the lack of coordinated information regarding *data bases*. Once a scientist knows which body of literature to consult, it is relatively easy to conduct a search; however, the problem is first to find which data base to use. Although the National Technical Information Service (NTIS) makes available data bases of Federal scientific and technical information sources, and the commercial services have catalogs available to their subscribers, no complete catalog of data bases exists at present. In order to reduce duplicated efforts and save time, a national catalog of available data bases is necessary for on-line machine-readable systems. The Federal Government should fund development of such a data base catalog. It is important to note, however, that the mere existence of a national data base catalog guarantees neither accessibility nor use; concomitant adjustments must be made in the hardware, software, and behavioral elements in the system.

There is not only the man-machine problem, but there is also the problem of transferring information from one field of science-engineering to another. *Cross-disciplinary references* are important because studies of the innovation process indicate that the "invasion of ideas" stimulates creative research. This problem is partially correctable through the expansion of indexing and cross-referencing, making certain that, in the process, the information from one field is cataloged in terms employed by other fields. Software programs must henceforth take fuller account of cross-field referencing.

Categorization problems also occur within fields. For example, searching for literature about a specific chemical presents difficulties because the same substance can be known under different names. The drug chlorpryomazine goes under about 100 designations. *Chemical Abstracts* resolves this problem by listing a substance under one entry and cross-referencing other names to this one agreed-upon name. Other fields can be encouraged—by modest government funding—to establish *standardized nomenclature* and cross-referencing for computerized indexing systems.

VOICE COMPUTERIZATION

The ability to communicate with computers by voice responds to a genuine user need. Speech is a more natural and efficient means of communication than an artificial computer language punched on a keyboard mechanism.

Advances in microelectronics make possible speech recognition by computers and limited man-machine voice communication. Although human-computer voice-to-voice communication still seems far away in terms of technical capability, it is not too early to begin studying its possible impact on the science information system. Modest funding of research in this area, including the development of software systems, would be worthwhile for charting future potential.

TRANSLATING ENGLISH

A subset of the man-machine communication task involves easier access to data bases through framing questions to the computer in standard English rather than in code. Such a system would be particularly useful for those who might be deterred from using the retrieval system if they could do so only by learning a special computer language. Philips Research (The Netherlands) already has an experimental system which attempts to answer questions posed in uncoded language.

Because of the "market-pull" for such a device, the private sector should not require government subsidies for demonstration projects and commercial application. However, the NSF is funding basic and applied research on cognitive problems, symbol recognition, and the like which would assist in the technical development of this user aid.

FOREIGN TRANSLATIONS

In addition to problems in "translating" the literature into computerized forms so that users can recognize it, there is also a translation problem in the more traditional sense: translating foreign literature into English. Until recently, American scientific and technological primacy was so overwhelming that foreign scientists and engineers eagerly sought American publication: there seemed little need to keep up with the literature from abroad. This is no longer the case. Scientists in other countries have always done respectable research—which we neglect at our peril. They are now making outstanding contributions to fields that once were the "property" of American science and technology. In 1960, roughly 34 percent of the world's scientific and technical periodicals were published in this country; by 1975, the figure was only 17 percent.[13] Obviously, we can no longer afford to ignore the foreign literature.

Translation of foreign scientific and technical periodicals and reports would be valuable, not only for our country's scientific/technical enterprise but also for international and military strategic purposes. For the private sector to undertake the translation of foreign works would be too expensive; the audience for such material is limited. Thus, major governmental support is needed, so that our scientists and engineers can keep abreast of the leading edge of worldwide research. At the least, abstracts of foreign items translated into English should be incorporated into our data systems.

DISSEMINATING FEDERAL R&D

Only a few years ago the National Technical Information Service (NTIS) was heavily criticized for its shortcomings in disseminating scientific/technical information emanating from the government. Since then, there has been remarkable improvement. The NTIS now sends out about 20,000 items daily from its collection of more than a million titles, and it supplies about 4 million documents or microforms annually. NTIS adds about 70,000 new summaries or reports, available in paper or microfiche, to its on-line computer research service (NTISearch) every year.

Furthermore, NTIS has been aggressively merchandising its services by publishing about 26 weekly newsletters, giving current summaries of new research reports and other specialized information, and through a biweekly journal directed at librarians and technical information specialists.

Although NTIS is improving its dissemination operation, it still does not incorporate materials from all government agencies. Legislation or executive guidance could require all Federal agencies to use NTIS for collecting, cataloging, packaging, and disseminating information.[14] But such a requirement would raise a number of questions.

Should NTIS widen its net to include technical materials other than those from government agencies? Certainly it should not restrict itself to such materials, but it is doubtful if it should become the omnibus collector and disseminator of *all* technical information. Although a single all-embracing system might seem desirable to information scientists, such a system would pose problems of government competition with the private sector, possibly stifle the formation of smaller specialized networks, and detract from the benefits of healthy pluralistic endeavors. And even if NTIS restricted itself solely to government materials, should existing Office of Management and Budget (OMB) Circulars on cost recovery apply, or could it be demonstrated that the dissemination of scientific and technical information is so valuable to the commonweal that the general public should support it?

THE PATENT QUESTION

The growing number of U.S. patents granted to foreign applicants has given rise to fears that America's technological primacy is endangered by foreign competition. Of the 70,292 patents issued by the U.S. Patent Office last year, some 37 percent were granted to persons living abroad. This number has been rising in recent years, and there is argument as to the cause:[15] Is foreign work outstripping American innovativeness, or does the increase in foreign patents simply reflect the need of foreign firms for patent protection as they seek to invade the vast American market? And does the number of patents tell us anything about technical quality or economic value? Not surprisingly, the discussion has called into question the role of patents in the innovative process.

The chief justification for patents is to stimulate innovation by providing legal protection for inventors, and to provide for disclosure of inventions. Because patenting involves disclosure of proprietary information, there is some ambivalence in the business community between the desire for legal protection and the benefits of secrecy. There are also questions involving the ownership and patenting of inventions made through government contracts—that is, with public funds. Without a proprietary position, few firms are willing to invest the sums necessary for successful commercialization.

The U.S. Patent Office contains a wealth of technological information. That information is indexed and cross-referenced into a comprehensive retrieval system, although there are frequent demands that a more efficient computerized information system be employed. These demands arise in connection with searches regarding proprietary rights, rather than from those seeking information for scientific and technological research. To industry, the main problem involving patents is their reliability—and that is a legal question. It would seem that the present search system, even though antiquated in some respects, is scarcely a barrier to information flow for innovative purposes.

In order to make patent literature contribute to further innovation it has been recommended that the patent "abstract" (made by the patentee) contain more information describing potential applications of the patent. But it is not clear that the abstract would provide information that might stimulate others to explore different applications of the patent.

Because patents remain a means for public dissemination of the results of technology, changes in the patent system are bound to affect scientific and technical information. Thus patent policies which cause industrial corporations to rely upon secrecy to protect their technical advances rather than on the legal protection afforded by a patent demonstrate that decisions made on legal-economic grounds can affect information transfer.

This situation would call (1) for research to assess the impact of secrecy as opposed to patenting decisions on the flow of information among industrial firms and among different sectors of the economy, and (2) for an information assessment in connection with any proposed changes in the patent laws.

COMMUNICATION-INFORMATION SYSTEMS

Another problem involving government policy is developing as the communication and information functions join, especially through computerized information carried over long distances for communication purposes. An already complicated situation in regard to communications regulatory policies will become even more complicated as three giants of American business—IBM, AT&T and Xerox—get more deeply committed to activities combining communication and information functions. Government policies regarding monopolies and public utilities regulation are involved, as well as the legal status of common carriers.

Although legal, financial and communication policies will occupy the forefront of public discussion on these developments, the fact is that this combination of powerful technologies, especially when employed by powerful corporations, will have major impact on the communication of scientific and technical information. Yet little thought has been given to these implications.

Two policy recommendations are in order: (1) support for research on how the computer-communication combination will affect scientific/technical information; and (2) the Department of Justice, Federal Communications Commission, Securities and Exchange Commission, Federal Trade Commission, and other regulatory agencies should be required to inform and consult with the Office of Science and Technology Policy regarding the possible impacts on science and technology of their proposed actions on computer-communications developments.

Extending Informal Systems

Although we have suggested means whereby the human "nodes" in information systems can be made more effective through improvement in the man-machine interface and through user-oriented software, how can we make the informal network more effective?

ENCOURAGEMENT OF PERSONAL CONTACTS

In *Managing the Flow of Technology*, Thomas J. Allen suggested that the government and industry pay more attention to investment in person-to-person communication.[16] Such an investment could stimulate university-government-industry interaction as well as communication among scientists within those institutions.

The Japanese experience reveals the importance of personal communication. Although they have access to the published literature, including American patents, without ever leaving home, the Japanese do not rely solely upon formal informational means.[17] Instead, they send delegations to attend American scientific meetings and tour our industrial establishment. The Japanese success in developing their scientific and technical enterprise demonstrates that person-to-person communication remains the most effective means for information dissemination and transfer.

Because our government plays so important a role in America's scientific and technological enterprise, government scientific/technical managers represent personal embodiments of information systems. They should be encouraged to attend professional meetings where their information potential can be augmented, just as academicians should be funded for that purpose.

A Federal program should send teams of American scientists and engineers abroad to gather information and to participate in interpersonal information networks. To this end, language programs would have to be instituted so that the Americans could communicate with their foreign colleagues.

ELECTRONIC "TRAVEL"

The interpersonal network can now be augmented by "teleconferencing." Although demonstration projects have shown the technical feasibility of teleconferences, we have not yet made them truly operational by incorporating them into institutional frameworks. Perhaps this is because teleconferencing has derived chiefly from science-technology "push" rather than "need-pull." The imaginative scientist has the "wouldn't-it-be-great" syndrome: Wouldn't it be great if we established an electronic network so that people could exchange ideas without having to leave home.

There might be some hidden dangers, which behavioral scientists should investigate. For example, the introduction of full telecommunications within a national corporation might encourage its scientists and engineers to work in isolation, obtaining data from the information system and feeding new data into it without ever leaving their laboratories. But because the high-density communication group is the individual and his immediate work group, the dispersion allowed by telecommunications might ultimately stifle the creativity of the individual who benefits from close contact with colleagues.

NEW AREAS FOR SCIENCE/TECHNOLOGY INFORMATION SYSTEMS

Heretofore we have dealt mainly with the communication of traditional types and forms of scientific and technical information for customary uses by the established scientific/technical community. However, the science and technology of communications are evolving rapidly, providing the possibility of applying them to new uses within science and technology themselves. At the same time, shifts are occurring in the social, economic, and political patterns of the nation which also offer the opportunity of using scientific and technical information devices and systems for broader purposes and for a larger public.

Problem Solving

We are not yet taking full advantage of the capabilities of the newer information technologies. Information processing devices can perform many functions beyond storage and retrieval and computations; they can, for example, synthesize the knowledge in the data base and present it to the researcher instead of a printout of separate research findings. The next step would be for the computer to take this information synthesis and work out the problem which the researcher is trying to solve.

Computers are already commonly programmed to answer questions by processing the data in their memories in accordance with specific programs. Now, however, real-time information sytems are available that allow a user, working at a computer terminal, to engage the data directly.

This combination of communications with computerized technology has been particularly useful in satellites employed for scientific purposes, such as NASA's earth resources satellites (Landsat), as well as for monitoring disarmament agreements between the United States and Russia. Real-time information systems do not require abstracting and indexing services or storage-and-retrieval mechanisms; instead, the work is directly with the data itself, which can be stored in digital or pictorial form for subsequent analysis and retrieval.

Combining data communication with scientific research itself might be the wave of the future. We can imagine a computer named "Norbert" being developed which would unite the information capacity of the computer with its problem-solving capabilities. The difficulties in developing "Norbert" might take several decades to resolve, but funding for "his" technical development should be made available, as well as research on possible implications for scientific research and information systems.

Information Technology in the Service Sector

The shift in importance from the manufacturing to the service sector of the American economy has been mirrored to only a small degree in information technologies. The chief reflection has been the automation of the office through computerized billing services and inventory-control systems. The question must then be asked: Are there no other ways in which the functional combination of computer and communications can enhance service productivity?

At a time when a great deal of attention is being paid to increasing productivity, the productivity of the ever-growing service sector languishes. Perhaps the dispersed nature and small scale of service-sector units militate against efficiency. Science/technology education and the processing of science information form part of the service sector, yet little attention is paid to applying science and advanced technology to services.

Although we might not be able to improve the productivity of, say, the barber, we might be able to use informational technologies for managing the delivery of services more effectively. The field of health services has already begun to develop more effective management techniques, employing some modern informational technologies (e.g., multiphasic screening), but more might be done. If nothing else, some thought should be given to the possible "spinoffs" from hardware and software advances in science/technology information systems to the service sector.

Social Problem Solving

Once there was a clear dividing line—or at least it was perceived as such—between scientific and technical matters and socioeconomic-political concerns. But the public is gradually beginning to realize that many (most?) social, economic, and political problems facing us have a scientific and technical component. As a result, people are turning to scientists and engineers for the amelioration of perplexing and complex problems, including such concerns as environmental protection, crime, urban blight, health services, consumer protection, and, indeed, many other items usually subsumed under the rubric of "quality of life."

Although the public has a tendency to look to a "technological fix" for easy resolution of these problems, there is increasing awareness that science and technology are enmeshed within a social context. Hence, for effective problem solving, there must be further study of the interface between science/technology and society. The rising interest in technology assessment and risk/benefit analysis reflects this reality.

There needs to be greater cooperation among the scientific disciplines toward ameliorating problems of a societal nature, and social science information must enter the scientific/technological information system. That information network will no longer be used only by the scientific/technical community, but must also accommodate the needs of social scientists, including applied social scientists—government officials, administrators, and, of course, the public. As a result, a larger group must be trained and educated in the use of informational hardware and software.

Three major challenges emerge: (1) developing and strengthening the social sciences informational system as a necessary component of the science/technology information system; (2) making scientists and engineers aware of the social, human, environmental, and economic context of their work; and (3) educating a new and wider public of information users, many of whom do not have statistical, mathematical, and computer knowledge.

The first task can be met by applying some of the same measures to further social science information systems that were used to bring the science/technology information system to its present level. The last two problems require cooperation of educational groups and institutions in a Sisyphean task. A good case can be made for continuing funding of research and educational ventures in such areas as the science-technology-society-values interface and in the public understanding of science.

Public Understanding of Science

There is a marked trend toward public participation in science policy decisions, as evidenced by the development of diagnostic and prognostic agencies in the government affecting science and technology (Occupational Safety and Health Administration, Office of Technology Assessment, Environmental Protective Agency, Nuclear Regulatory Commission, and the like). If the public is to participate in science policy decisions, it must be scientifically literate.

Educating the public poses a new problem: How can information technologies assist in rejuvenating education in basic science and in invigorating the public understanding of science and technology?

This is part of a larger educational question, of course. A decade ago there was great progress in developing informational technologies for programmed instruction and various types of audiovisual aids. Yet the sophistication of the technology was not matched by sophistication in use—another example of the failure to take account of the "human node" in informational systems—and it can scarcely be said that today's average American is sufficiently schooled in basic scientific and technical knowledge to make intelligent assessments of public issues involving science and technology.

Nevertheless, the science/technology information system represents a public asset. We must now begin anew to figure out how this system can be employed for wider service to the public.

NOTES

1. See Myers, Sumner, and Marquis, D. C., *Successful Industrial Innovations*, Washington, D.C.: NSF-69-71, 1969; Allen, Thomas J. and Cohen, Stephen I., "Information Flow in Research and Development Laboratories," *Administrative Science Quarterly* 14 (1969): 12-19; Langrish, J. et al., *Wealth from Knowledge*, London, 1972; Johnston, Ron, and Gibbon, Michael, "Characteristics of Information Usage in Technological Innovation," *IEEE Transactions on Engineering Management* EM 22, 1975; Carter, C. F., and Williams, B. R., *Industry and Technical Progress*, Oxford, 1957; Science Policy Research Unit, University of Sussex, *Success and Failure in Industrial Innovation*, London, 1972; Rosenbloom, R., and Wolek, F., *Technology and Information Transfer*, Cambridge, Mass., 1970.

2. Salzman, Ray M., and Niskanen, Anthony, "Data Entry Systems," *Modern Data* (Feb. 1973), pp. 28-43.

3. The advantages are argued in Turoff, Murray, "Human Communication via Data Networks," *Data Exchange*, Feb. 1974; "Computerized Conferencing: A Conversation with Murray Turoff," *Data Exchange*, Oct. 1974; Turoff, Murray, "The Costs and Revenues of Computerized Conferencing," *Proceedings of the Third International Conference on Computer Communication*, pp. 214-221. For a more balanced statement of advantages and disadvantages, see Johansen, Robert, Vallee, Jacques, and Spangler, Kathleen, *Electronic Meetings: Technical Alternatives and Social Choices*, Reading, Mass., 1979, esp. chs. 5 and 6.

4. Kelly, Patrick, and Kranzberg, Melvin, eds., *Technological Innovation: A Critical Review of Current Knowledge*, San Francisco, 1978, Part I.

5. For example, in the spring of 1979, the Monsanto Co. entered into a joint project with the Harvard Medical School. In the summer of 1979 Rep. George Brown (California) introduced a bill (H.R. 4672) which would encourage such industry-academic cooperation by establishing centers for industrial technology which would be affiliated with universities.

6. In 1979 Senators Harrison Schmitt (New Mexico), Howard W. Cannon (Nevada), and Adlai E. Stevenson (Illinois) introduced the Science and Technology Research and Development Utilization Policy Act (S. 1215) concerning the rights to government-sponsored inventors, in order to encourage private industry participation in Federal R&D programs and commercial use of their results.

7. See Doctors, Samuel I., *The NASA Technology Transfer Program: An Evaluation of the Dissemination System*, Washington, D.C.: 1971; and Sayles, L. R., "The Innovative Process: An Organizational Analysis," *Journal of Management Studies* 2 (Oct. 1974), pp. 190-204.

8. In *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R&D Organization*, Cambridge, Mass., 1977, Thomas J. Allen stresses the role of the "gatekeeper" who provides the linkage of the R&D laboratory to external scientific-technical information; Allen also shows how human and organizational systems can be restructured (including the spatial arrangements of offices) to bring about better person-to-person contact.

9. See Knight, Kenneth E., Kozmetsky, George, and Baca, Helen R., *Industry Views of the Role of the Federal Government in Innovation*, Austin, Texas, 1976.

10. Landau, Martin L., "Redundancy, Rationality and the Problem of Duplication Overlap," *Public Administration Review* 29, July-August 1969.

11. Ackoff, Russell, et al., *Scientific Communication and Technology Transfer Systems*, 2 vols., Wharton School, University of Pennsylvania, 1975, NSF-GN-41883.

12. "From 1960 to 1974 the GNP has grown 177 percent, R&D funding has increased 136 percent and S&T communication resource expenditures have increased much more by an estimated 323 percent." King, D. W., *Statistical Indicators of Scientific and Technical Communication*, 1960-1980 Vol. 1, Rockville, Md., King Research, Inc., Center for Quantitative Sciences (1976), p. 15.

13. King Research, Inc., *A Chart Book of Indicators of Scientific and Technical Communication in the United States*, Rockville, Md. (December 1977), p. 11.

14. It should be noted that the Smithsonian Science Information Exchange has recently been moved to the Department of Commerce, and there is the possibility that it will eventually be combined with NTIS.

15. The statistics are from the Commissioner of Patents and Trademark *Annual Report* for Fiscal Year 1978. An analysis is to be found in Schiffel, Dennis, and Kitti, Carole, "Rates of Invention: International Patent Comparison," *Research Policy*, 1979. A popularized account of the patent-innovation problem appears in "Innovation: Has America Lost Its Edge?" *Newsweek* (June 4, 1979), pp. 63-68.

16. See Allen, op. cit. Similar conclusions are to be found in Pavitt, Keith, *The Conditions for Success in Technological Innovation*, Paris: OECD, 1971; and Kelly, Patrick, and Kranzberg, Melvin, *The Flow of Scientific and Technical Information in the Innovation Process*, NSF Grant No. DSI 74-13045-A04, Atlanta: Georgia Tech., 1977

17. Examples are to be found in Ozawa, Terutomo, "Japan's Technological Challenge to the West, 1950-1974," *Motivation and Accomplishment*, Cambridge, Mass., 1975 and Tsurumi, Yoshi, "Japanese Multinational Firms," *Journal of World Trade Law*, Jan./Feb. 1973.

4 Privacy: Impact of New Technologies

*by James B. Rule**

SUMMARY

Characteristic of modern, "advanced" societies is the development of vast systems of dossiers, computer records and other files on individuals. These systems are created and used by organizations—bureaucracies, both governmental and private—to guide their dealings with the people depicted in the files. Computing and other modern technologies have greatly accelerated the growth of such systems, and multiplied their effects on individuals. In the last fifteen years, both the existence of the systems and their uses have become the subjects of sharp public controversy.

In the wake of these controversies, an official approach to the protection of privacy has begun to emerge in legislation and policy. This approach seeks to make personal data systems, both computerized and conventional, conform to rules of due process in the handling of people's data. Specifically, systems are to be open in their workings, receptive to individuals' attempts to ascertain and correct their own files, responsible for the accuracy and confidentiality of data held in them and limited to "legitimate" uses of the files. These principles have formed the bulk of recent privacy protection activity in America and abroad.

Clearly these principles, if forcefully applied, can make a considerable difference in ensuring fairness to individuals in the uses of their data. But some commentators, including the Privacy Protection Study Commission, have noted significant limits to this approach. For one thing, the due process mechanisms created here may be so complicated and time-consuming as to vitiate their usefulness to most people. For another, this approach does virtually nothing to limit the development of record systems which could, if political conditions permitted, unduly constrain the exercise of constitutional rights.

Needed are innovations in privacy policy that seek ways of circumventing the need for mammoth, sophisticated personal record systems. A search for such less "information-intensive" alternatives to the endless growth of bureaucratic recordkeeping on persons may be the only alternative, in the long run, to continued erosion of privacy. Such alternatives do exist, but considerable research and reflection are required before they can be put into effect. Only when the investments in making do with less personal information approximate those now made in finding new uses for such data will the most compelling needs for privacy protection be met.

INTRODUCTION

English is unusual in having a single word for all the things covered by the term "privacy." For a homeowner, "privacy" may mean a hedge around the backyard to block the looks of neighbors and passersby. For a politician, it may mean the chance to think aloud, without having remarks retained "on the record." For a job-seeker, it may mean not having to divulge an ep-

* Professor of Sociology, State University of New York, Stony Brook.

isode from one's past to prospective employers. For a celebrity, it may mean the chance to go out in public unrecognized. For a pregnant woman, it may mean the opportunity to secure an abortion anonymously and without intrusive questioning.

Within the last 15 years in America, and more recently in other "advanced" societies in the West, privacy has become a political issue. But the controversies which have created the issue have not involved all these different senses of privacy. The "privacy issue" as it has arisen in the last few years has focused on *personal documentation*—dossiers, computer files, and other personal data as collected and used by organizations. The debates have focused on the following questions: Who has the right to collect information on members of the public? What uses of such information should be considered legitimate? How can fairness be ensured in the uses of personal data? When is the transfer of information collected by one agency to other users justified? How much say should any individual have in the treatment of his or her data—and how should such a voice be implemented? And how much personal data keeping *in general* is good for the kind of society in which we wish to live?

ORIGINS OF THE ISSUE

Any understanding of the privacy issue demands some interpretation of its origin. How have the uses of personal information which are so controversial today been singled out for public attention? What roles have science and technology played in these changes, and how have scientific and technological influences meshed with political and social forces?

For most casual observers, the answer to these questions is simple: The cause of today's privacy controversies is technological change. Sophisticated microphones, unavailable only a few decades ago, can record conversations previously considered inviolate. Wiretaps can monitor conversations for "key words" without anyone present to listen to the whole conversation. Special cameras can penetrate darkness to record scenes whose privacy would otherwise be absolute. Psychological tests can plumb personal feelings and capacities once considered beyond the reach of investigation—and to influence those tested through the experience of such monitoring. Electronic analysis of the human voice can fix the identity of a speaker nearly as reliably, it is said, as the matching of fingerprints.

But among all technologies, computers take first place among the iconography of modern anxieties about privacy invasion. People seem to believe that the computer, more than anything else, invites the collection and use of personal information about individuals. Inscrutable, authoritative, all-remembering, the computer confers unimagined powers upon its users—or so the popular image seems to have it. And the further perfection of computer techniques, unfolding with the quiet inevitability of scientific "progress," promises to press the demands of technology on personal privacy further and further.

There is considerable truth in this vision, but in the final analysis it is both exaggerated and oversimplified—both with regard to computing and to technology in general.

Technologies do not develop untouched by human needs and social exigencies. Many fascinating and feasible technologies remain indefinitely in the realm of theoretical fancy, simply because no influential social interest stands ready to underwrite their development. We have rather sophisticated systems for retailers to run credit checks on would-be customers, for example. But we have no comparable intelligence systems to enable customers to check on retailers or their products. For new technologies to be born, both scientific understanding and social sponsorship are required. To understand the origins of the privacy controversy, we must examine both the technologies which have made it possible to collect and use more personal data *and* the new social conditions which have nurtured the development of these technologies.

It hardly takes a sociologist to identify the more obvious of the social conditions. Present-day privacy controversies have centered on the data management activities of *organizations*, that is, bureaucracies, public or private. In modern societies virtually every basic social relationship requires direct or indirect participation by organizations. Key life junctures—from birth, christening and immunization through education, medical care, marriage, home and automobile ownership, insurance and, ultimately, death—require certification by organizations. The regular need for such certification in turn leads to creation of massive data files on the persons concerned. These records may be manual or computerized, centralized or dispersed; their effects on the lives of those they depict may be helpful or coercive; but they cannot but matter to the people depicted in them.

At the turn of the century, many present-day record-keeping bureaucracies simply did not exist. Income taxation, social security, credit reporting, most law enforcement records systems, and many other of today's major personal records systems were yet to be founded. Elsewhere—as with birth and death registration, passport issuance, educational records and welfare payments—some systems existed, but in very abbreviated forms. The recourse to record' eping bureaucracies is such a fundamental part of our lives today that we have all but forgotten how recently they developed.

What functions do detailed personal records have in the workings of organizations? Why are bureaucracies prepared to pay the considerable costs of developing and using such data?

Data files are the key means—and often the *only* means—by which organizations can determine how to act toward individuals. Records enable organizations to deal with people precisely, according to exact details of their past histories or present circumstances. The data so recorded, of course, may not be those which the people concerned would consider important; instead they are data which the organization holds essential for framing bureaucratic action toward the people to whom they refer. Sometimes such action entails holding people responsible for past misdeeds—as in criminal recordkeeping or in certain aspects of credit reporting. Sometimes it means documenting circumstances of people's lives which bear on their responsibilities to organizations—as in income taxation or conscription. Sometimes it means recording information bearing on organizations' responsibilities to the people depicted—as in social security, medical records and certain aspects of insurance.

Whatever the nature of the relationship, however, personal data systems clearly represent a new social link between institutions and the public. It is a link that mixes the anonymity and impersonality of bureaucratic action with the intimacy and discrimination of personal acquaintance. Such *fine-grained concern* by organizations toward individuals, as it has been called, represents a distinctive social phenomenon of modern times.[1]

Thus modern organizations build up data files on people as a kind of "information capital"—as important to their operations as capital in the usual sense. Without such data resources, organizations would be unable to discriminate in their dealings with the vast numbers of otherwise anonymous persons whom they daily confront.

Individuals experience this dependence just as strongly from the opposite direction. One's life is continually shaped by the contents of one's record: the manner in which such records are generated, stored, interpreted, and transmitted affect one's life chances in countless ways. Not only do we all *have* more personal records about us written down here and there, these records also *matter* more in the way in which we are treated. The enormous social impact of the recorded word—both for individuals and the organizations that deal with them—has provided the ingredients of most of today's privacy controversies. As the Privacy Protection Study Commission recently noted,

> in American society today records mediate relationships between individuals and organizations and thus affect an individual more easily, more broadly, and often more unfairly than was possible in the past. This is true in spite of almost a decade of effort to frame the objectives of a national policy to protect personal privacy in an information-dependent society.[2]

Nearly all writers on the subject have echoed this recognition of the sheer importance of records in affecting people's lives.

Computing often plays a central part in managing these record systems. No less important, however, are other relatively modern technologies which, since they have been with us for some time, we tend to take for granted—e.g., telecommunications, a modern postal service, modern filing methods, and the like. Indeed, modern management techniques are essential for the orderly operation of large personal record systems.

Among these technological and social conditions for the emergence of bureaucratic personal recordkeeping, computing may actually be the least urgently required. Some very large record systems are in fact maintained by hand, both in America and abroad. In this country, for example, many of the most important personal records used in insurance are still manually maintained. Similarly, at least until recently, much of the central recordkeeping for industrial security clearances was carried out on literally millions of blue index cards; the visitor to the site where they were stored saw them stretching off into the distance like some inland sea.

Strictly speaking, then, computing is neither a necessary nor a sufficient condition for sophisticated, large-scale personal record systems, or for the public controversies they generate. But if computing does not *create* the bureaucratic appetite for data on people, it certainly makes it possible to satisfy such cravings to an extent otherwise unthinkable. Computerized storage and transmission of large amounts of data—including personal data—are among a very few major organizational activities which have actually become radically cheaper in the last several decades. Under these conditions, it is little wonder that organizations have developed the discriminating, fine-grained relationships with their publics which the use of large amounts of personal information makes possible.

Indeed, these radical changes in the cost and feasibility of storing and using personal data demand the reshaping of our customary notions of "private" and "public" information. A major protection of privacy has been the unavoidable "waste" of much personal information. We do not ordinarily consider the times and frequency of our entering and leaving our places of residence "private," in that we do not seek to conceal our comings and goings from whoever may be passing at the time. But if we find that someone is systematically noting such information over an extended period, we feel that our privacy has been infringed upon. Such "public" information is ordinarily "wasted," like an untapped resource generated in the manufacture of some other product. We rely on such "waste," and we do not generally welcome ingenious attempts to "recycle" wasted data on ourselves, whether by a nosy neighbor or a bureaucracy.

Privacy, as most people seem to use the term, means the unavailability to others of information about one's self. Most people probably do not seek *absolute* privacy; a neighborhood where no one notices one's comings and goings may seem rather cold, after all. But the loss of privacy quickly becomes objectionable when information about one's self is not only known, but also recorded as a basis for unspecified future actions or judgments. Thus the well-being that most of us seem to feel on knowing that "our" data, though not actively concealed, are normally "wasted." Many of the satisfactions of privacy seem to come from not having to work too hard to safeguard it. One likes the idea that information about one's self, though in some sense "public," will not be indefinitely available for others' use.

Computing, however, helps organizations wear away at privacy in this sense, by making it possible to "save" information which otherwise would be "wasted." To be sure, the effects of such erosion of privacy on individuals are mixed. Some such effects are almost wholly beneficial—one wants one's physician, after all, to be able to assemble the most comprehensive medical dossier possible on one's self. Similarly, a "good" credit history does the subject no good unless it is recorded—something impossible unless collection systems are capable of counteracting the "waste" of useful credit information. Elsewhere, of course, effects on individuals are unpleasant, as in encounters with law enforcement agencies or IRS auditors. But however one may reckon these advantages and disadvantages in particular situations, this decrease in "wasted" information certainly has to be reckoned a loss for privacy in the usual sense of that word. From certain points of view, this loss may be warranted in light of compensating advantages—say, in the strength of government or in access to convenient services, such as credit. But the loss is felt nonetheless, and one consequence is public controversy over, and demands for, the protection of privacy.

Personal data systems, then, make demands on people's privacy interests. But what, exactly, are these interests? In what sense are people injured by systematic collection of data about themselves—or are such injuries purely the fancies of the people concerned?

One kind of privacy interest, obviously, is the interest in avoiding the mishandling of one's own data files. One does not want one's dossier to become the basis for "unfair" discrimination—whatever one may mean by this. Similarly, one wants to prevent arbitrary, inaccurate or uninformed judgments from being made on the basis of, or in the name of, one's own "file." Thus we all have strong interests in preventing bad effects from arising from the use of records on ourselves—just as we do, for that matter, in avoiding the effects of derogatory word-of-mouth reputation.

In addition, most people probably also experience what one might call an "aesthetic" interest in their own privacy. People shield certain information about themselves—such as the sight of excretion, sexual activity or profound emotion—simply because there is something inherently uncomfortable about others' having this information. Here the concern is not with the more distant *consequences* of having such information used against one's interests at some later point. Rather, this kind of privacy interest is a desire to withhold information as an end in itself.

A third category of privacy interest is more abstract, but perhaps in the long run even more important for policy considerations. This is what one might call people's *collective* or *holistic* interest in the effects of data systems on social relations throughout society. People may suffer bad effects from data systems, after all, even if no records of their own are involved. If data systems are used to create a climate of intimidation and fear, everyone suffers, because public discussion and private innovation are stifled. If the feeling becomes widespread that anything we say or do *might* be held against us, the quality of public life is diminished, even if these feelings are exaggerated. Similarly, the *potential* use of a given data system as an instrument of repression, even if its present purposes are nothing of the kind, poses risks that all members of a society implicitly share.

Before today's privacy controversies, however, these and other privacy interests were often dimly perceived, and often not noted at all. To the extent that they are recognized today, they have emerged only gradually from nearly fifteen years of public reflection and controversy.

CONTROVERSY

"All historical events are inevitable," it has been observed, "but some are more inevitable than others." In retrospect, it is hard not to regard the controversies that swirled up around these changes as more inevitable than most. Beginning in the 1960s, the new demands by organizations for personal data set off intense public debate. How much data on private persons did private and government bureaucracies really have? How did the use of such data affect the lives of those depicted in the files? What rightful interests could ordinary people be considered to have in the treatment of "their" data? How do trends of change in personal data management affect such interests? And perhaps most importantly, what are the proper avenues for asserting such interests?

To be sure, these questions did not arise fully formulated. Instead, what appears in retrospect as the emergence of the "privacy issue" came in a series of skirmishes over specific data-management schemes. One of the first of these public controversies appeared in 1966 and 1967 over a proposed National Data Center. The idea was to centralize personal data from many Federal agencies, including the IRS, the Census, Social Security and a number of others, in a single statistical clearinghouse. The data were supposedly to be held in strict confidence and used for statistical purposes only—not for decisionmaking on the individuals involved. But the idea aroused intense disapproval in Congress, prompting remarks like the following from one particularly critical representative:

> Good computermen know that one of the most practical of our present safeguards of privacy is the fragmented nature of present information. It is scattered in little bits and pieces across geography and years of our life. Retrieval is impractical and often impossible. A central data bank removes completely this safeguard.

Reactions like this caused the Data Center idea to be dropped from consideration in 1967.

Since then, privacy-related controversies have been a staple of congressional concern. Specific practices to come under congressional scrutiny have run the gamut from credit reporting and psychological testing to political surveillance and misuse of IRS files. The Watergate period, of course, greatly sharpened these concerns, but one must not forget that the issue was well established before then. A partial listing of congressional hearings and reports relating to privacy since 1967 shows more than fifty such events, more than half of them occurring before the Watergate period. Despite the wide range of settings, the themes of these controversies were remarkably consistent. They most often entailed disputes over the propriety of the use of personal data files by powerful organizations. The cumulative effect was a search for principles which could guide and make publicly acceptable the further development of such systems.

Some statements to the Congress envisaged sweeping approaches to these dilemmas. Note the remarks of Jerome Wiesner, then president-elect of MIT, to a congressional committee in 1971:

> I doubt that anyone is aware of the full extent of the surveillance and information-collection activities that go on in this nation. Many people, myself included, have long operated on the assumption that our activities are being monitored.

Then he continued,

> We should be prepared to accept the cost of considerable inefficiency in our various social and governmental processes to safeguard our privacy, and, as I judge it, our freedom, dignity, happiness and self-respect. By costs, I mean both the financial costs and the loss of a degree of control that the state might otherwise have over genuinely threatening individuals. . . .

The strong words by Wiesner were by no means unique among the statements of those who participated in congressional hearings in this period.

Nevertheless, the approach implied in Weisner's statement has not predominated in the emergent policy response to these controversies. Instead of seeking to limit the growth of personal recordkeeping, the predominant policy effort has been to find procedures to help people live more comfortably with their records.

This response has developed slowly. Its first manifestation was the Fair Credit Reporting Act, which went into effect in 1971, regulating the collection, storage and dissemination of personal information in consumer credit and certain aspects of insurance and employment. Further development of these principles came in Records, Computers, and the Rights of Citizens, a report commissioned by the Secretary of Health, Education, and Welfare in 1973. The principles set down in this report, in turn, guided the second major Federal legislation on the subject, the Privacy Act of 1974, which regulates the treatment of personal data by Federal agencies. Finally, the Privacy Protection Study Commission, created by the Privacy Act of 1974, issued its report in 1977 setting forth an additional 162 recommendations for legislation and policy affecting the treatment of personal data by both government and private organizations.

To be sure, these four efforts hardly exhaust the list of landmarks in the evolution of privacy protection policy. Other important contributions can be found, for example, among the extensive body of state legislation and in the guidelines offered by President Carter in 1979. But one can view these four elements as central in establishing an approach to privacy protection which now enjoys wide acceptance. While the details vary from case to case, the approach does embody a core key of principles:

> *First*, that the existence and workings of personal data systems ought to be open and publicly known;

Second, that recordkeeping organizations ought to be responsible for maintaining personal data in an accurate, timely and complete form;

Third, that data-keeping organizations be responsible for ensuring that the contents of files ought to be disseminated only for "legitimate" purposes;

Fourth, that means ought to be available to individuals to ascertain and, if necessary, correct the contents of their files and the uses made of them.

The various legislation and policy applications of these principles differ considerably in their force and effectiveness. The Fair Credit Reporting Act, for example, has been widely criticized as relatively weak in its strictures upon credit bureaus. It prohibits virtually none of the reporting practices current in the industry at the time of its adoption, and its provisions for access to and correction of files are cumbersome and disadvantageous to consumers. Similarly, the recommendations of the HEW report leave the recordkeeping practices of the CIA and many law enforcement agencies virtually untouched. The Privacy Act of 1974 is likewise relatively weak in its strictures upon these same recordkeepers; moreover, disclosure of filed information is largely limited only to purposes "compatible with the purpose for which it was collected"—a vague and scarcely stringent restriction. Finally, even the recommendations of the Privacy Protection Study Commission, though they seek to strengthen the provisions of the Privacy Act at a number of points, implicitly accept some highly privacy-invading practices, such as the disclosure of IRS and Social Security information to locate parents who abscond leaving children on public support. Needless to say, such disclosures fail even the vague test of compatibility with the purposes for which the information was originally provided.

Much could be written about the successes and failures of particular laws and recommendations in meeting the promise of the four principles outlined above. But for present purposes, the principles themselves hold greater interest than specific applications. What promise do these principles hold for protecting individuals' privacy interests, and what are their limitations? How fittingly does this approach to privacy protection respond to the public anxieties and demands that brought it about? These questions bear serious and reflective consideration.

FAIRNESS IN PERSONAL DATA MANAGEMENT: THE PROCEDURAL APPROACH

One might characterize the four principles noted above as a *procedural* approach to privacy protection. The aim is not to prevent collection and use of personal data, but instead to create a body of standard, equitable procedures—"rules of the game"—through which individuals can influence the handling of their own data. The report of the Privacy Protection Study Commission, for example, characterizes its intent as that of establishing for individuals "assertible interests" in their own files. The language differs in other statements on the subject, but the underlying idea is quite similar. The goal is to create mechanisms for people to participate in the creation and use of data files on themselves, and thereby to secure fair, even-handed treatment.

Perhaps the most important question about these procedural measures is whether people actually use them. Do people manage to assert the "assertible interests" which they are accorded? Evidence on these points is fragmentary, but worth considering. Since the Fair Credit Reporting Act went into effect, one notes, credit bureaus have reported a steady stream of inquiries from consumers about their files. Similar indications are to be found in the results of my study of some 291 Long Island, New York, residents; some 18 percent of those interviewed who reported having had difficulty with credit also reported having gained access to their own credit files.[3] This figure is almost certainly much greater than it would have been before the Fair Credit Reporting Act, when credit reporting firms generally sought to keep their activities out of public attention. Thus, although we have nothing like an exhaustive survey of the effects of this law, it appears that consumers have used it to a significant extent to discover, and presumably to correct, the contents of their credit files.

We have similar evidence of the use of the Privacy Act of 1974. Appendix Four of the report of the Privacy Commission notes the existence of some 6,700 systems of personal records covered by the act throughout the Federal Government as of 1976.[4] When the act went into effect, some Federal agencies experienced great influxes of requests from people seeking access to their own records. The Justice Department, for example, recorded some 35,000 such requests during 1976. People's access to their own files has frequently resulted in requests for changes in the files; most often these seem to have met with compliance. The Department of Health, Education, and Welfare, for example, reported receiving some 19,000 requests for changes in personal records in 1976, of which only 79 were denied. Similarly, the Department of Defense received some 11,000 such requests, of which only about 150 were not fully granted.[5] Thus, the act seems to have occasioned a flood of assertions of the newly created assertible interests, and general agency acquiescence with many of these assertions.

At the same time, these routine corrections may have come about only where agencies felt no important interests of their own threatened by the sought-after changes. While the Privacy Commission reports some 70 ongoing legal actions by individuals to change their files as of 1976, they note that "No case seeking damages has yet been decided against the government. . . ."[6]

Moreover, the seemingly large numbers of successful requests for amendments must be seen in relation to the literally billions of data files covered by the Privacy Act. In this context, the rate of requests cited above is less impressive. As the Privacy Commission authors themselves remark,

> In part, this less than expected utilization of the Act can be attributed to the difficulty of finding out how to use the *Federal Register*, and of wending one's way through the maze of Privacy Act procedures.[7]

In other words, though the procedures established in the act to afford access to one's own file may be relatively straightforward, considerable effort and inconvenience may be necessary for people to establish the existence of record systems which could contain files on them.

Such difficulties point to a fundamental limitation of the procedural approach to privacy protection. At best, it would seem, law and policy deriving from this approach can create significant openings for individuals to monitor and influence the content and uses of their records. But this requires active involvement from the individual and may entail considerable time and effort. As the number of "assertible interests" which people enjoy in various record systems multiplies, one's ability to monitor any one relationship naturally decreases.

Thus, as personal recordkeeping becomes more common, the creation of assertible interests through procedural reform clearly has its limits. People can keep up on some assertible interests all of the time, it would seem, or all of them some of the time. But we have passed the point where most people can realistically monitor all data uses about themselves all of the time—any more than they can analyze all the available product data about every item they buy at the supermarket.

Procedural guarantees to provide people some say in the uses made of their own files are indispensable. But procedural measures of the sort discussed above, it appears, scarcely offer a comprehensive answer to the need to protect privacy interests. For one thing, as we have seen, people simply may not have the time and other resources to exercise meaningfully all the rights accorded them by such guarantees. For another, the procedural approach provides no guidance for determining what forms of personal recordkeeping ought or ought not to exist. This issue cannot readily be considered purely in terms of the individual interest in fairness addressed by the procedural approach. It requires that we consider other privacy interests.

THE INTERESTS OF PRIVACY PROTECTION

The question of how legal and other mechanisms can assure fairness and discourage arbitrariness in the treatment of individuals' data files is, for most people, what the privacy debate has been all about. Yet this approach is restrictive, and ultimately unsatisfactory if taken alone. For one thing, people may simply be unable to master the task of keeping abreast of all their "assertible interests." For another, the public's interests in the workings of personal data systems are not simply the aggregate of individual interests in seeing one's own records handled "correctly."

In addition to these individual interests, I have noted *collective* or *holistic* interests in the effects of personal data systems on social life *in general*. If the cumulative effect of burgeoning personal data systems is to increase the overall power of organizations vis a vis individuals, then no one really escapes these effects. We live in a world where increasingly sophisticated personal data management is relentlessly reducing the "distance" between large organizations and individuals' private lives. Whether or not our own names happen to be in the files, we cannot avoid experiencing the enhanced power of government and private organizations to influence people.

This distinction has been lost in most discussions of the privacy issue. One reason for this may have been the pervasive influence of the idea of "balancing" as a guiding concept for privacy protection. Nearly every major statement on the subject has characterized procedural reform as a means of "redressing the balance" between individual claims of privacy, on the one hand, and organizations, on the other. Thus in *Privacy and Freedom*, probably the most influential of all books on the subject, Alan Westin wrote,

> If privacy is to receive its proper weight in any process of balancing . . . what is needed is a structured and rational weighing process, with definite criteria that public and private authorities can apply in comparing the claims for disclosure or surveillance through new devices with the claims to privacy.[8]

Similarly, in the opening paragraph of the Privacy Protection Study Commission report, one reads that the invasion of privacy will continue "unless steps are taken soon to strike a proper balance between the individual's personal privacy interests and society's information needs." One finds the same metaphor—and, more importantly, the same logic—in virtually every major statement on the subject over the past fifteen years.

But the trouble with good metaphors is that they beguile us into forgetting that they *are* metaphors. We are right to seek procedural measures against arbitrariness and unfairness wherever organizations use personal data systems. But the development of these systems itself fundamentally alters relationships between powerful institutions and the public. Due process in the handling of personal records can hardly restore these relationships to their state before recordkeeping developed. Organizations now can collect and use much more detailed personal information than ever before. As a result, they have the potential power to act much more forcefully and discriminatingly in their dealings with people—for purposes either benign or coercive. These facts are not fundamentally altered by procedural measures to ensure fairness and due process in information use.

A list of names and current addresses, a record of people's recent expenditures, a log of their travels or a list of their associates—these things, mobilized by the capabilities of modern information technologies, confer upon organizations considerable power to influence people, or indeed to coerce them. Procedural restraints prevail only so long as the political and social conditions endure in which the restraints are established. The Nixon years, after all, saw the misuse of many of the Federal Government's personal data resources despite earnest efforts of officials in agencies such as the IRS to restrict such breaches of procedure. Once the capabilities of systematic personal recordkeeping are in place, the possibility of their misuse can never be completely eliminated.

The risks posed by such possibilities might be more acceptable, if only the situation were static. But this is not the case. Indeed, if people agree on anything about personal data systems, it is that such systems are in a headlong process of change. The social and technological forces which have brought these systems into existence continue to magnify their importance. Organizations continue to find more and more ingenious ways of obtaining and using personal data; they readily use these methods to sharpen the discriminating decisionmaking characterizing their relations with their publics. Public opinion, in turn, often demands the exercise of such bureaucratic discrimination—discrimination between the creditworthy and the deadbeats; discrimination between tax cheaters and those justly entitled to reduced tax liability; or discrimination between those entitled to welfare or other benefits and the unentitled. As the ability of organizations to observe and act on such fine differences grows, the power of organizations over people's lives grows commensurately. And as the intensity of information use rises, privacy, as most people know it, is inevitably eroded.

The holistic issues raised by these considerations are disturbing—no doubt because they appear much more difficult to resolve than individual issues of fairness and due process. It is not comforting to consider that the treatment of personal information may be more open, less arbitrary and fairer, yet at the same time more all-encompassing and potentially more threatening. Yet these risks cannot be denied. Procedural approaches to privacy protection, if applied forcefully, do offer the chance to the individual to deal more effectively with his or her data. But they do not help us to decide when a given personal data system must be counted too intrusive or too potentially repressive to be worth the risk.

Perhaps the most astute acknowledgement of these points is found in the report of the Privacy Protection Study Commission. These remarks about the Privacy Act of 1974 apply as well to other procedural approaches to privacy protection:

> It was passed partially as a protection against premeditated abuses of Federal agency records but, more importantly, in recognition of the fact that even normal uses of a record about an individual can have harmful consequences for him and that this potential harm can be greatly magnified by the use of emerging computer and telecommunications technology. Despite these antecedents, however, there is little in the Privacy Act to prevent premeditated abuses of power through the misuse of recorded information, particularly where internal agency uses are concerned.[9]

In other words, creating "rules of the game" for fair and equitable use of personal information may not help much if the data-keeping party loses its incentive to adhere to the rules. Nor, as the Commission authors note a few paragraphs later, does such procedural reform have much effect on the ever-shifting "imbalance" in prerogatives in the direction of data-keeping organizations:

> Actual or potential information abuses are much more likely to result from continuing growth in the government's appetite for information about individuals and in the use of that information for growing numbers and types of purposes. *The real danger is the gradual erosion of individual liberties through the automation, integration, and interconnection of many small, separate record-keeping systems, each of which alone may seem innocuous, even benevolent, and wholly justifiable.*[10]

The interests damaged by such trends are holistic ones. That is, whether we are the victims of misuse of our own records or not, what is damaged is a climate of public life that we all share. For the protection of this kind of interest, we must look beyond the procedural approach.

ANOTHER APPROACH TO PRIVACY PROTECTION

No one really likes the idea of a world where bureaucratic data-keeping on private individuals proliferates without end—even if such activities were carried out fairly and with scrupulous regard for due process. Yet seeking alternatives to the ever-increasing growth of personal data systems requires us to question fundamental assumptions about modern technologies and their social setting.

One of the most influential of these unexamined assumptions has been that of the "need" of organizations for more and more personal recordkeeping. Nearly every major writing on the subject has depicted such needs as virtually a fixed and immutable feature of modern societies. The Privacy Protection Study Commission authors no more than echoed a well sung theme when they wrote, "We live, inescapably, in an 'information society' and few of us have the option of avoiding relationships with record-keeping organizations."[11]

But of course, some things are more "inescapable" than others. The steady growth of organizational reliance on personal data systems is not the simple product of a fixed "need," but rather the result of organizations' search for new ways to gather, maintain, and use such data. An equally earnest search for alternatives could well show that the bureaucratic need for personal information is not as absolute as it has been considered to be.

The principle that might provide such alternatives is hardly exotic. Organizations develop sophisticated personal information systems, it has been argued, to maintain discriminating decisionmaking and action vis a vis the people concerned. Credit grantors seek detailed personal data in order to extend precisely the "appropriate" credit privileges to each applicant. Insurance firms seek background data on their applicants in order to adjust rates for insurance precisely to the risks of insuring a specific person. The IRS collects data to document the precise tax obligations of each taxpayer, and to locate that taxpayer if there should be a dispute about such obligations. It is easy to see how sharpened discriminations like these can be advantageous for the organizations involved—and, from many points of view, for the public as well. But if we consider deliberate minimization of such discriminations as an investment in protecting important privacy interests, we can grasp imaginative alternatives to the endless collection and use of more and more personal data by organizations. We need to take a hard look at what some analysts term "less information-intensive" alternatives to reliance on ever-increasing personal data management.[12]

To develop such alternatives, organizations would be encouraged to invest in ways of doing *without* detailed personal information, instead of investing in more intensive use of such data. The form taken by such efforts would obviously vary according to institutional setting. In credit, it might entail simply limiting the amount of personal information which credit bureaus could collect and use on individual applicants. In income taxation, it might entail reducing the number of circumstances bearing on tax liability—in short, simplifying the determination of tax owing in any one case, so that fewer circumstances had to be documented. For welfare and other benefits, a less information-intensive alternative would be to establish a minimum floor of eligibility for all, thus reducing the necessity of intrusive investigations and detailed recordkeeping to distinguish fine degrees of need and to exclude the ineligible. For insurance, it might mean curtailing the amount of personal information—including particularly sensitive data about community repute—allowed to bear on deciding whether to insure an applicant and at what rate.

The new direction being sought in these examples hardly demands outright proscription of the use of personal data in settings like these. That would be as crude and heavy-handed as the approach which now prevails—i.e., the assumption that any personal data which can be shown to be useful to organizations must necessarily be legitimate for them to collect. What I am suggesting is simply the serious pursuit of ways for organizations to use less personal information. Enormous resources are currently invested in developing ways for public and private organizations to observe finer and finer distinctions in their treatment of members of the public—distinctions that can only be supported by collecting more and more personal data. Only by seeking a new emphasis in organizational planning can we hope to reverse the steady trend toward erosion of privacy.

For all its simplicity, this idea inevitably brings a strong riposte from spokesmen for data-using institutions. Any slackening in the use of personal data, they argue, would ill serve a public that demands the services provided by these data systems. It is popular demand, they claim, which requires exact determination of eligibility for welfare, and hence the severe demands on the privacy of welfare applicants. Popular demand is likewise said to underlie the collection of information on applicants for credit, or Social Security benefits, or insurance coverage, or security clearances. The public wants these services to be provided, it is said, and the only feasible way of doing so is to distinguish between the eligible and the ineligible.

There is a kernel of truth in these observations. Public opinion may applaud the use of personal data systems against "welfare chiselers" or tax evaders—particularly if people view the groups so targeted as standing at some social distance from themselves. The public may also acquiesce, albeit ambivalently, to demands for personal data required for provision of services that they themselves need. But this hardly establishes that personal data systems which today enjoy such acquiescence somehow represent a response to "popular demand." The whole truth on the matter is not nearly so clear-cut.

It would be hard to show any pattern of popular demand for the institution of major personal data systems in America prior to their inception. The immediate source of these innovations has ordinarily been organizations themselves. Indeed, initial public reactions to many requirements for personal data have often run from uneasiness to outright hostility. Events surrounding the inception of income taxation and Social Security, as well as the responses to the periodic extensions of the U.S. Census inquiries, are cases in point. Public opinion may have become inured to the demands of these systems with the passing of time. But such acquiescence hardly attests to a deep resolution of the issues they pose. The attitudes of Americans toward personal data collection remain ambivalent and contradictory.

Since these issues first gained salience in the mid-1960s, there have been a number of studies of public opinion on privacy matters. One of the most recent and most comprehensive is the study commissioned by Sentry Insurance, and designed and carried out by Professor Alan Westin and the Louis Harris organization in 1978. Published in early 1979, this entailed interviews with a representative selection of some 1,500 members of the general public, as well as some 600 top-level officials in public and private organizations. The study holds most interest here for the mixture of support and disapproval it demonstrates among the general public toward prevailing personal data practices.

Regarding inquiries made by employers of job applicants, for example, the great majority of respondents approved of such standard inquiries as those into educational and employment background. Also widely approved were employer access to results of tests of job skills and reports on applicants' past medical histories. Similarly, large majorities of the public approve insurance companies' inquiries into applicants' age, health and medical history in screening applications for life or medical insurance. Nearly two-thirds of the respondents felt that physicians and hospitals were "doing enough" to maintain the confidentiality of personal information.

On the other hand, other results of this same study show public disapproval of common uses of personal data. Some 62 percent of the respondents from the general public, for example, disapproved of employers' inquiries into job applicants' record of arrests where there have been no convictions. Yet such inquiries are by no means rare. About half the sample held it improper to ask job applicants to take psychological tests. Some 24 percent reported that they had been asked to provide information which they considered improper when applying for consumer credit. About three-quarters of the applicants held it improper for insurance companies to inquire into the "lifestyle" or "moral character" of applicants for life or medical insurance—though such inquiries are routine among many, if not most, insurance companies. Similarly, some 81 percent of these respondents disapproved of police access to personal bank account information without a court order; yet many banks provide such information to police, FBI and public prosecutors virtually on demand.

Perhaps the most striking findings of the Sentry-Harris-Westin study relate to public attitudes about privacy in general, more than about specific practices. Certainly the results betray serious, though diffuse, distrust and discontent with the overall state of privacy in America. More than 80 percent of the general public sampled, for example, agreed with the statement, "Some people are prevented from getting fair treatment because of past mistakes kept too long on their record." Some 63 percent agreed with the item, "If privacy is to be preserved, the use of computers must be sharply restricted in the future." And only 27 percent agreed that, "In general, the privacy of personal information in computers is adequately safeguarded."

Perhaps most remarkable of all in these respects is the comparison of some of these responses with those in earlier polls. The Sentry study included one item which the Harris organization had used several times before in similar national samplings. The shift in opinion over time is marked, as the following table shows:

Item		1974 %	1976 %	1977 %	1978 %
Americans begin surrendering their personal privacy the day they open their first charge account, take out a loan, buy something on the installment plan, or apply for a credit card.	Agree	48	47	67	76
	Disagree	43	47	24	21
	Not sure	9	6	9	3

Clearly, Americans believe that privacy is on the wane in their country, diverse as their ideas may be on what to do about it.

In the final analysis, it would be unreasonable to look to public opinion for a formula for resolving the dilemmas of privacy policy. The fact is that Americans espouse conflicting and even contradictory views on these matters—as of course they do on many matters. People deplore the loss of privacy, but they want all the facts in their favor marshalled to support their credit applications. They feel that computing should be restricted in the interests of privacy, but they expect quick processing of their income tax refunds and tough government action against welfare cheaters. Or, as the results of my own study of Long Island residents show, people readily cite common items of personal data which they feel they ought not to be required to provide to organizations; but they rarely refuse to provide such data when actually seeking services from such organizations.

The one point on which my own analysis of these issues is unambiguous is the following: Continued organizational and technological "progress" along present lines can only increase the pressures against personal privacy which have fueled the controversy thus far. Thus what I have termed the procedural approach to privacy protection can hardly provide a comprehensive solution. For that approach seeks to protect privacy by limiting collection and use of personal data to that which is "necessary" for organizational purposes. And modern organizations, both public and private, promise to discover more and more efficient and hence "necessary" uses for personal information. The only hope of avoiding endless growth in pressure against personal privacy lies in the active pursuit of less information-intensive alternatives to present bureaucratic practices.

One sometimes hears that reducing the personal information available to organizations would cripple their operations. But this generalization is so sweeping as to be *prima facie* unconvincing. Doing without personal information would impose *costs*, to be sure, both on the organizations involved and on the public itself. The nature and extent of these costs, as well as their acceptability, would vary with the setting.

In selling auto insurance, for example, firms have traditionally investigated the moral repute of insurance applicants. The reason is that drivers considered disreputable by their peers are believed to fare poorly if involved in litigation over claims, regardless of the merits of their case. Thus insurance companies have sought either to avoid insuring such persons, or to impose higher rates in these cases. But these distasteful and privacy-invading investigations could be avoided if consideration of these "reputational risks" were simply proscribed in writing insurance. The effect would be to transfer the costs of community prejudice from the victims of such prejudice to the community of policyholders as a whole. It is encouraging to note that some insurance companies are indeed de-emphasizing use of these forms of personal data.

Elsewhere the costs of reducing the amount of personal data used by organizations would take different forms. In consumer credit, meaningful reductions in the amount and kinds of personal data allowed to bear on credit decisions would clearly reduce the efficiency of credit operations, and raise the costs. Inevitably, such reductions would mean that some persons who, rightfully or wrongly, would otherwise be denied credit would receive it. The costs of raising the numbers of "bad credit risks" accorded credit would presumably have to be borne both by credit grantors, in the form of higher operating costs, and by other credit users, in the form of more expensive and less convenient credit.

Another setting in which requirements for personal information are now particularly heavy is income taxation. Here the less information-intensive alternative would be to simplify tax regulations, so that fewer detailed personal data would need to be taken into account to determine liability. The costs in this case would not necessarily be felt in higher operating expenditures; indeed, these might be reduced by such simplifications. Rather, the costs—if they are to be reckoned as such—would be in the form of reduced ability of the taxation system to render precise justice to taxpayers by taking fine variations in circumstances into account.

People will obviously differ in their views of the acceptability of such alternatives. Some, for example, might accept restriction of the kinds of data bearing on credit or insurance applications, while opposing any restriction on the personal data available to criminal recordkeeping agencies. Or, others might be willing to consider a "baseline" of entitlement to public support for low-income people as a less information-intensive alternative to welfare investigations, yet be unwilling to consider simplifying income tax laws to reduce the requirement for personal data there. These contending values can only be weighed in the thinking of informed, reflective citizens.

Nor, in any of these settings, do we face a stark either-or choice between use of personal data and total renunciation. What we face, rather, is a dialectical tension between demanding more information in the interests of organizational efficiency and reducing such demands to reduce risks to privacy and individual autonomy.[13] The more relevant information credit-granting organizations can develop on individuals, the more profitable will be their efforts to discriminate among credit applicants. Reducing information available for such discrimination will make credit processes more "forgiving" to those who, for good reasons or bad, have failed to meet those obligations in the past. It will also reduce the collection of personal information in ways which could, under certain social conditions, make such data usable for repressive purposes. The worth of such advantages must be considered, in each case, against the resulting costs.

CONCLUSION

In confronting these choices, and indeed all the choices posed by privacy protection policy, good research can play a significant role. Regarding procedural reforms of personal data systems, we need studies to show how readily people can become informed of the "assertible interests" which legislation and policy may afford them. Further studies are needed to help us understand who actually tries to take advantage of such possibilities, who fails to do so, and how the attempts turn out.

Concerning less information-intensive alternatives to privacy-invading practices, the need for research is even more urgent. We need detailed studies of the costs and benefits of reducing the collection, storage and use of personal data by large bureaucracies. This means both the dollars-and-cents costs to data-keeping organizations, as in the losses that would result from relaxed discrimination in credit. But it would also mean symbolic costs, as in the public indignation which might result if relaxed discrimination in welfare appeared to increase "welfare chiseling." Along with such assessments, we need shrewd, imaginative ideas for alternative bureaucratic techniques to provide needed services without such heavy reliance on personal data. Only when the investments in exploring such alternatives begin to approximate the investments now being made in using *more* personal data will we have an accurate grasp of the possibilities for privacy protection.

But it would be wrong to think that research alone can dictate "solutions" to any of these issues. Research can help assess the gains and losses to various privacy interests through application of different policies. It can help us understand how effective procedural measures really are in defending interests of fairness, or what the costs of doing with less personal information might be in a particular setting. But the insights of such studies become meaningful only in terms of the values of reflective participants in public debate. How much importance do we really ascribe to various privacy interests? What sacrifices are we prepared to make for such interests when they are opposed, let us say, to those of efficiency? Research can inform choice, but only informed individuals can choose. And the only real certainty we have in making these choices is that the values of efficiency and technological progress will sometimes run counter to those of privacy.

NOTES

1. See, for example, James Rule, *Private Lives and Public Surveillance*, New York, Schocken Books, 1974, Ch. 1.
2. *Personal Privacy in an Information Society*, p. 3.
3. Supported by the National Science Foundation.
4. *Appendix Four: The Privacy Act of 1974: An Assessment*, p. 11.
5. Ibid., pp. 32-33.
6. Ibid., p. 34.
7. Ibid., p. 33.
8. *Privacy and Freedom*, p. 370.
9. *Appendix Four: The Privacy Act of 1974: An Assessment*, p. 107.
10. Ibid., p. 108; emphasis in original.
11. *Personal Privacy in an Information Society*, p. 5.
12. See, on this concept, *The Politics of Privacy*, James Rule, Douglas McAdam, Linda Stearns and David Uglow, New York, New American Library, 1980, Ch. 3.
13. This seems to be the sort of trade-off Jerome Wiesner was proposing in the congressional testimony quoted herein.

5 Information Privacy: A Legal and Policy Analysis

*by Robert R. Belair**

SUMMARY

The paper presents a legal and policy analysis of information privacy. Information privacy is defined as those standards that give the data subject control over his personal information so that such data are collected, maintained, used and disseminated in a fair manner.

The paper assumes that information privacy contributes to the democratic functioning of a society and promotes personal freedom. The paper suggests that at least five major developments affect the preservation and development of information privacy rights: (1) increased societal complexity and attendant increases in the collection, maintenance and use of personal information; (2) growth of government; (3) increased importance of personal credentials; (4) increased need for careful recordkeeping processes in a society where institutional authority is suspect; and (5) revolutionary advances in surveillance and computer and telecommunications technology.

The paper describes historical as well as recent efforts by the courts, the Congress and the executive branch to fashion law and policy for information privacy. The paper concludes that constitutional protections for information privacy are vague and that common law protections also have inadequacies. The paper examines in detail the content and implementation of recent statutory protections such as those found in the Federal Privacy Act of 1974.

The last section of the paper examines the adequacy of existing information privacy concepts in light of societal and technological changes. This section also analyzes alternate privacy protection models. The paper emphasizes that the nation has not yet and may never develop a consensus as to the optimum level of information privacy. The paper concludes that policymakers in the 1980s will have to look closely at several information privacy protection concepts including: the role of the data subject; confidentiality protections; restrictions on the use of data; standards for record system management; strategies for monitoring future advances in surveillance and computer information technologies; and strategies for implementing and enforcing information privacy standards.

INTRODUCTION

This paper presents a legal and policy analysis of information privacy by identifying the personal and societal interests served by the information privacy doctrine; summarizing the effect of recent cultural, economic, technological and legal developments upon information privacy; and identifying information privacy mechanisms and concepts that require further development or refinement in the next few years.

The paper comprises four parts: an identification of

* Attorney at Law, Hill, Christopher, and Phillips, P. C., Washington, D.C.

the interests served by information privacy; a historical discussion of the emergence of information privacy as an issue of public importance with particular emphasis on technological developments; an analysis of information privacy law; and a critique of policy mechanisms or strategies that can be used to safeguard information privacy.

The term "information privacy" does not have a universally accepted, exact definition. However, the term is used in this paper (and is most commonly used) to refer to the standards for the collection, maintenance, use and disclosure of information about individuals. In a normative sense information privacy means the use of personal information in a manner that gives the individual maximum control over his information so that it is used in a manner that promotes fairness.

The framework for a comprehensive information privacy code was first expressed in a 1973 study conducted by the Department of Health, Education, and Welfare's Advisory Committee on Automated Personal Data Systems. Its report identified five information principles (dubbed "fair information practice principles") intended to insure that information about individuals is used fairly and in a way that protects confidentiality:

- There must be no personal data recordkeeping systems whose very existence is secret.
- There must be a way for an individual to find out what information about him is in a record and how it is used.
- There must be a way for an individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent (confidentiality).
- There must be a way for an individual to correct or amend a record of identifiable information about him.
- Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuse of the data.[1]

Information privacy should be distinguished from that type of privacy that refers to the individual's right to engage in certain types of privacy behavior free from surveillance, intrusion, or regulation by third parties (usually the government). This type of privacy interest is sometimes labeled "behavioral" or "volitional" privacy.[2]

Privacy law is relatively insensitive to the truth or falsehood of the subject information. This paper does not develop concepts of the use of false information, often covered by the term "defamation." Nor does the paper concern itself with the many kinds of information disclosure that are considered highly beneficial to society. Information about individuals may lead to voting redistribution, design of more equitable public service delivery, affirmative action to compensate for historical inequities, better health care treatment, expanded finance and credit opportunities for reliable lower-income persons, the grounding of airplane pilots who have poor responses or performance, and a host of other benefits.

The analysis in this paper begins with a basic assumption that information privacy is essential to a society's ability to function in a democratic and fair manner and, in addition, contributes to the attainment of personal freedom and autonomy.

According to many observers the recent rapid growth in our society's size and complexity, coupled with revolutionary advances in information and communication technology as well as surveillance technology, has eroded information privacy and created a need for new safeguards on the collection, management, use and disclosure of personal information.

By the mid-1970s, a new set of safeguards modeled after the Health, Education, and Welfare Code of fair information practice had been developed and partially implemented. The paper examines the effectiveness and implications of these safeguards. It suggests that a consensus has yet to emerge regarding the optimum level of information privacy. Furthermore, the paper suggests that existing mechanisms to safeguard information privacy may not prove adequate or may have undesirable costs. It examines other potential mechanisms or policies that could be used to safeguard information privacy.

INTERESTS SERVED BY THE INFORMATION PRIVACY DOCTRINE

It is possible to conceptualize and categorize the interests served by the information privacy doctrine in many different ways. However, for the sake of conceptual clarity this paper posits that the information privacy doctrine protects and serves two fundamental political interests: (1) an interest in insuring that our society functions in a democratic manner; and (2) an interest in insuring that personal freedom is maximized.

The way in which personal information is used often has a critical impact upon an individual's ability to obtain employment, credit, insurance, licenses and other valued benefits. In other words, decisionmakers, both public and private, use personal information to make decisions about entitlement to benefits and other valued resources. If a decisionmaker has obtained inaccurate or inappropriate information; or if sensitive, stigmatizing information is improperly or indiscriminately disclosed; or if a subject is not aware of the existence of a record about himself or cannot obtain the record to challenge its accuracy, timeliness or completeness, then the subject individual may be harmed significantly.

Thus, in a very real sense the interest protected by the information privacy concept is the interest in insuring that society uses a fair decisionmaking process. It is an interest in due process and equitable procedures, and as such is closely related to the democratic ideal of a fair and open society.[3]

Second, regardless of the substantive harm that can be done to an individual who is the victim of unfair or improper information practices, it is also posited that the information privacy doctrine safeguards the individual's interest in personal freedom.[4] It has been argued that the act of collecting, retaining and disclosing personal data by institutional recordkeepers has a "chilling" or compromising effect on the individual subject's sense of freedom and liberty. Some writers assert that when an individual knows that information about him has been collected or used he may feel that he has lost a degree of control or discretion about his life and a concomitant degree of freedom.[5] As one writer has put it, information about oneself is an extension of the self. Thus, to lose control of that information is to lose control of the self—to lose freedom and dignity.[6]

Furthermore, the method of collecting personal information, even when it does not result in tangible harm to the subject, may be seen by the subject as having an intrusive or *in terrorem* effect. Use of the polygraph, wiretapping and eavesdropping devices or hidden cameras to obtain personal information falls into this category. At the other end of the information handling process, the use of personal information by third parties to make personal contact with the subject, even though no tangible harm is done, can also have an intrusive, chilling effect. The individual who is contacted by an unknown researcher for follow-up information about the individual's cancer operation is a victim of this kind of information practice.

THE DEVELOPMENT OF INFORMATION PRIVACY AS A POLITICAL ISSUE

The history of American jurisprudence and politics indicates that our society has always had some concern with protecting the use and disclosure of personal information.[7] However, by the mid-1960s information privacy had become an issue of public importance and the threat to information privacy a topic of popular and scholarly attention.[8]

The period since 1965 has been characterized by a growing concern for the protection of information privacy. A comprehensive national opinion survey of attitudes toward information privacy conducted for Sentry Insurance Company by Louis Harris and Associates, Inc., in 1978 concluded that "public concern about privacy—more specifically, the potential abuse or misuse of personal information by business and by government—has increased steadily throughout the seventies. . . ."[9] The survey indicates that three out of four Americans now believe the "right to privacy" should be akin to the inalienable right to life, liberty and the pursuit of happiness. One out of every three Americans, according to the survey, believes that we are very close to the society portrayed in George Orwell's *1984* in which "virtually all personal privacy had been lost and the government knew almost everything that everyone was doing."[10]

The reasons for the increasing public concern about information privacy are not hard to identify. The Harris survey, for example, indicates that almost all Americans believe that today greater amounts of personal information and more detailed and sensitive personal information are requested and obtained by institutional decision-makers. Americans also believe that their ability to monitor or control the use of their personal information has correspondingly declined.

Although a definitive empirical study has never been done to document this pervasive notion, several studies contain statistics that corroborate this view.[11] It is a logical and deducible result of several recent societal developments. The exact nature and number of the developments or forces that are germane to the information privacy problem is naturally subject to debate. However, for purposes of analysis this paper defines and discusses five such developments.

Growing Complexity of Society

Individuals' transactions with private sector decisionmakers, be they employers, credit grantors, banks, insurers, doctors or other institutional dispensers of statuses and benefits, have become considerably more complex. In an earlier era, financial transactions, for example, were more dependent on cash, were simpler, and if they generated any records at all, those records were likely to be maintained by the individual. Cash purchases create few, if any, records. Similarly, savings held under the pillow leave no auditable information trail. By contrast, financial records in today's environment are inevitably held by second parties: banking and other financial institutions. As a consequence, the friendly neighborhood banker knows an enormous amount about his customers' lifestyle, spending habits, and financial history, and has in hand a considerable amount of biographic history.

The development of consumer credit transactions is but one example of an aspect of modern living that works to create recorded personal information and to separate individuals from ownership of their personal information. Each time an individual uses a credit source the creditor collects and maintains (and often disseminates) personal information. In the period since World War II the amount of consumer credit increased tenfold and by 1977 amounted to over 182 billion dollars. It is estimated that half of the adults in the nation now have at least one credit card.[12]

The increase in the amount of recorded information is not only a response to the sheer growth and complexity of society but, as well, seems to reflect what one writer has called the desire for an "information cushion."[13] Today, individuals and institutions expect that events that have a significant effect on individuals will be carefully recorded for the protection of the individual.

> In the simpler days, a fall from a horse might result in a visit by the town doctor, or it might not. Today a car accident triggers a police report, an insurance report, an investigation of the claim, ambulance and hospital services, diagnostic tests, medical insurance claims, and so on *ad nauseum*.[14]

In short, the whole pattern of our economic and social life has changed; banking and credit are but two prominent and obvious examples. This change appears to have contributed to an equally basic change in the type and amount of information that is collected and traded and in the ownership and control of that information.

Growth of Government Activity

A second development that appears to have fueled the information privacy issue is the growth in the extent and character of government activity. In the period conveniently marked by the New Deal, the government, particularly the Federal Government, has assumed an increasingly important role in the lives of its citizens. Local, state and Federal agencies perform all of the classic governmental functions (police, fire, public health, national defense and tax collection), each of which requires government agencies to collect extensive amounts of personal information. What's different is that government now does so much more. For example, it makes loans, it provides a baffling array of public assistance, and it makes decisions about licenses and permits. Each social service program, whatever its substantive merits, requires the dispensing agency to collect, maintain, update and usually disseminate detailed and intimate personal data about the benefit recipients.[15]

Development of the "Meritocracy"

A third phenomenon that seems to have altered information practice is the gradual development of a whole group of social mores that for shorthand might best be labeled as the "meritocracy."

Americans largely accept the notion that an individual's intelligence, personality, emotional stamina and character can be characterized, quantified and ultimately captured in record form. We believe that this type of record evaluation is relevant to almost all decisions that need to be made about an individual. Therefore, individual credentials are compiled, memorialized and disseminated on a scale that 19th century Americans could simply not have imagined.

It may be that much of the force that propels the drive to collect, maintain and disseminate personal information ultimately derives from this set of beliefs. By the time most of us are ready to enter the work force, we have been tested, measured and analyzed, and all of this information has been carefully (or, worse yet, not so carefully) memorialized in records that will trail us throughout life and, in fact, outlive most of us by a good many years.[16]

Diversity and Confrontation

A fourth development that appears to have contributed to the emergence of information privacy as a public issue has been the erosion in the last fifteen years of some of the American political consensus and its replacement by an era of confrontation politics and distrust of authority. The Watergate scandal and the Vietnam War are the most dramatic examples of events that heightened public distrust of government and institutional authority.

In this sort of climate it is natural that citizens insist that institutions be subject to detailed and presumably rational rules for the distribution of resources, benefits and penalties. It follows that in order to achieve precise, accountable decisionmaking, detailed records and careful recordkeeping covering an individual's background, performance and status are required.[17]

Surveillance and Information Technologies

A fifth development of critical importance is the revolution in surveillance and information and communications technologies. In the last fifteen years four surveillance and information technologies have matured.

ELECTRONIC LISTENING AND WATCHING DEVICES

First, advanced electronic listening and watching devices have been developed. The listening devices make it possible to record telephone conversations and in person conversations without the need to physically place a microphone or other device on the wire or in the room under surveillance.[18] Indeed, according to popular press reports, it is now technologically feasible to monitor virtually all overseas and domestic calls made over telephone circuits, cable lines or microwave transmissions. Microwave transmissions—high frequency radio signals bounced between towers spaced 50 or so miles apart—are especially easy to intercept.[19]

A recent development that promises to vastly increase the feasibility of wide-scale wiretapping and eavesdropping is the use of computers to scan intercepted conversations. When the computer "hears" a trigger word the conversation is taped or identified. This capability makes it feasible to monitor thousands of conversations.

It may even be technologically possible in the near future to use wiretapping technology to detect the feeble electronic emissions from remote electric typewriters and translate these emissions so as to "intercept" the typewritten information.

Senator Frank Church, chairman of the Senate Intelligence Committee, has publicly expressed what is probably a common sentiment about this kind of technological surveillance capability. He has said that bugging technology:

> at any time could be turned around on the American people, and no American would have any privacy left, such is the capability to monitor everything—telephone conversations, telegrams, it doesn't matter. . . . I know the capacity that is there to make tyranny total in America.[20]

At the same time, equally rapid advances have been made in visual surveillance technology. That technology now permits covert as well as overt observation of individuals, at night, or behind windows and somewhat opaque surfaces; and it permits individuals to be magnified and to be videotaped.[21] Police in many cities are using permanently mounted cameras to conduct 24-hour surveillance of downtown streets. Many of the systems are equipped with telescopic night light and videotape capabilities.[22] High resolution cameras can now be made in a small enough size to accommodate James Bond-style covert use. One popular report indicated that cameras had been concealed in a pocket cigarette lighter.[23] Infrared miniature cameras, though still too large to be individually carried, are also available today. They permit users to "see through" many types of surfaces.[24]

New, non-visual types of watching devices are also available. For example, a Harvard professor recently conducted an experiment using 16 volunteers who wore tracking devices that identified their geographic position as well as permitted the monitor and the subject to communicate via beeper signals. According to Professor Ralph Schwitzgebel, the pilot project clearly demonstrated that "we have the technology to monitor people if we want to."[25] Apparently the National Aeronautics and Space Administration has developed a personal tracking and communications system that can be housed in a wristwatch.[26]

TRUTH DETECTION DEVICES

Significant technological advances have also been made in recent years in "truth detection" devices. The polygraph is easily the most common truth detection device. It is estimated, for example, that about 300,000 job-connected lie detector tests are given each year.[27] Standard polygraph machines attach a cuff or other device to the subject and attempt to measure physiological changes in respiration, blood pressure and galvanic skin response. The machine's performance is based on the thesis that lying will produce fear and anxiety, which in turn will produce measurable physiological changes.[28]

In the seventies, truth detection technology has developed to include various mechanisms that purportedly can be used to test veracity without the need to touch the subject or inform him of the surveillance. The voice stress analyzer picks up fluctuations or modulations ("microtremors") in the voice that vary according to the degree of stress. A skilled examiner purportedly can interpret the stress readings to determine whether the subject is lying. Recently, small versions of stress analyzers have been developed that can be attached to a phone. Blinking lights on the machine's console supposedly indicate whether there is stress in the speaker's voice. Manufacturers have also announced their intention to market voice stress analyzers that are pocket sized. One manufacturer even has a wristwatch model.[29]

An even newer truth detection technology is called "micromentaries." This technology involves the examination of videotapes of a subject's non-verbal behavior. The tape records split-second facial expressions, such as eyes that are blinking faster than normal. These facial clues purportedly enable a trained examiner to evaluate the subject's veracity.[30]

COMPUTER TECHNOLOGY

The third and most important technological development of the sixties and seventies that has contributed to the perceived information privacy problem is the dramatic advance in computer technology. The Sentry Insurance report, in which the Harris survey was presented, identifies computerization as a principal cause of the public's increased concern about privacy.

> This trend [public concern about privacy] has stemmed largely from the increasingly technologi-

> cal, computer-oriented nature of our society in which countless determinations, ranging from credit, insurance and job promotions to social security disbursements to census-based allocations of public funds, are now based on the collection of so-called 'personal data.'[31]

A full discussion of recent advances in computer technology is beyond the scope of this paper. However, it is possible to summarize the extent of the increase in computing and information handling power and to identify the effect that these advances are likely to have upon practices and policies for handling personal information.

According to a technology report published in 1977 by the Privacy Protection Study Commission, a two-year independent Federal study group, virtually unprecedented advances over the last 25 years have vastly increased computer speed and storage capabilities for handling information. Specifically, the speed with which computers can handle instructions and thus manipulate information has increased 50 thousand fold in the last 25 years. Today, computers can accept well in excess of 100 million instructions per second. During the same period the information storage capacity of computers increased 10 thousand fold. Today, computers can store and retrieve millions of characters of information (each character represents one unit of information) in a fraction of a microsecond (a microsecond is a millionth of a second). Also during this same 25-year period, the Privacy Commission calculates, computer size has decreased 100 thousand fold and computing cost has been similarly reduced 100 thousand fold.[32]

The implications for information handling and information policy of these advances in automated information technology can hardly be overstated. The Privacy Commission concluded that computers are "dramatically and rapidly altering the way records about individuals are created, maintained and used."[33]

At the risk of oversimplifying the effects of computer advances upon information practices, at least five distinct consequences should be identified.

First, thanks to modern computer technology, the sheer amount of personal information that can be handled (collected, stored and retrieved) is greatly increased. Every study that has looked at the computerization of information has documented this effect. The policy implications are dramatized, for instance, in a CIA report which states that computer technology made it possible for the CIA to collect and store domestic surveillance information about approximately 300,000 individuals. The study concludes, "without such computer support, it would have been impossible to run this program."[34]

Second, the number and types of individuals or organizations that can now use mini- and microcomputer systems to maintain personal data are greatly increased. It is estimated that by the 1980s, computers the size of a large notebook, with enough power to perform every conceivable personal computing need, will be widely and popularly available. In cataloging the potential uses of these microcomputers, one analyst has suggested that small businessmen and professionals will be able to keep better track of their customers and clients. He says, for example, that:

> physicians should be able to store and organize large quantities of information about their patients, enabling them to perceive significant relations that would otherwise be imperceptible.[35]

A third impact of advanced computer technology is the potential of automated transaction records for multiple uses. Discrete elements of personal information maintained in an automated system can be separated, arranged and correlated in numerous configurations suitable for varied purposes. One writer has said that the real impact of computers has little to do with its prodigious archival and broadcasting capabilities, but, rather, lies in its ability to transform a mass of raw data into meaningful and useful information.[36] R. L. Polk and Company, the country's largest producer and seller of mailing lists, reportedly uses automated technology to "massage" bits of personal information from many sources to compile sophisticated profiles of individuals for inclusion in selected mailing lists.[37]

A fourth characteristic of automated information systems is the ease and flexibility with which information can be retrieved from such systems. Automated technology, for instance, can permit users to obtain personal information from a system by submitting not only personal identifiers but various non-personal identifiers. The Privacy Commission's report expresses the fear that this capability may mean that the automation of real estate records, driver registration records and other types of records normally filed by geographic location or other non-personal identifiers will become resources for the construction of personal profiles once this data can be cross-referenced by personal identifiers.

In addition, the ease of retrieval means that another traditional barrier to third-party disclosure—cost of retrieval—will be minimized. As the Privacy Commission puts it, "the cost of retrieving information made it worthwhile for the record keeper to refuse some requests and to limit its responses to others, and thus, in many cases, there was a satisfying coincidence of high-minded principle and economic interest."[38]

The Commission's report notes that because of automation it may soon be cheaper to say yes to an access request than to say no. (Of course, this technological advance has at least one consequence that contributes to information privacy. Individual subjects are also more likely to be able to have access to their records.)

A fifth impact of automation that has implications for information policy is the ability of computers to link records. Modern systems do not need to have huge centralized data bases. Instead, it is relatively easy and inexpensive to link many different and physically dispersed record systems. The net effect is to encourage dossier building and record surveillance on a scale not previously practicable.

TELECOMMUNICATIONS TECHNOLOGY

Advances in computer technology have almost been matched by related advances in telecommunications technology. Our ability to transmit information has increased both quantitatively and qualitatively. The state of a society's technological capability to transmit information has an inevitable effect on information privacy by virtue of the fact that the principal (although by no means the only) information privacy issue is "who gets to see what personal data under what circumstances."

The sheer power of our telecommunications technology means that more information can be exchanged more quickly and at less cost. The Privacy Commission report concludes that in the last 50 years data transmission capacity has increased three orders of magnitude from 3,000 characters per second over 12 voice channels in 1920 to eight million characters per second over 32 voice channels via coaxial cable in the 1960s to many hundreds of millions of characters per second over 100,000 voice channels via helical waveguide and optical fiber systems in the late 1970s and 1980s.[40]

In addition, the nature and quality of transmission capabilities are changing. Today, digital communications make computer-to-computer transmissions commonplace. Indeed, it has created what is in effect a merger of data communications and data processing. At the same time, data transmission satellites and even newer forms of telecommunications technology, such as neutrino beams (elementary particles that have little interaction with matter and thus can actually travel through the earth to transmit data from one part of the world to another), are making long distance, instant transmission of data practicable and inexpensive.[41]

In short, advances in telecommunications, like advances in computers, are making it easier to say yes—to exchange data—and harder to say no.

LEGAL STANDARDS AFFECTING INFORMATION PRIVACY

As noted briefly in the introduction, the term privacy is used by the courts to refer to two relatively distinct legal doctrines: (1) the doctrine that pertains to the individual's freedom to engage in personal activities without intrusion or regulation; and (2) the doctrine labeled here as information privacy that pertains to the control over and use of personal information. This latter information privacy interest has received relatively little constitutional recognition but considerable statutory recognition. This section of the paper summarizes the constitutional, common law and statutory standards for information privacy.

Constitutional Standards

To date, the Supreme Court has handed down three opinions that deal squarely with the constitutional issues raised by information privacy claims. In all three decisions the Court has rejected the claims of the privacy proponents.

In *Paul* v. *Davis*,[42] a Louisville, Kentucky, resident sued two local police chiefs because they circulated a flyer entitled "Active Shoplifters" to merchants in the area. The flyer contained the resident's name and photograph, even though he had been arrested (but never convicted) of shoplifting. In his complaint, the resident alleged that the police chiefs had violated his constitutional right to privacy. The Supreme Court rejected that claim and said that the right of privacy has been applied only to government intrusions into traditional marital activities. The opinion concluded:

> [The resident's] claim is far afield from this line of decisions. He claims constitutional protection against the disclosure of the fact of his arrest on a shoplifting charge. His claim is based, not upon any challenge to the State's ability to restrict his freedom of action in a sphere contended to be 'private,' but instead on a claim that the State may not publicize a record of an official act such as an arrest. None of our substantive privacy decisions hold this or anything like this, and we decline to enlarge them in this manner.[43]

In *United States* v. *Miller*,[44] the Supreme Court ruled that the Fourth Amendment to the Constitution, which protects individuals from unreasonable government searches of their "persons, houses, papers and effects" does not protect the confidentiality of personal information held by institutional recordkeepers. In this case, Miller was convicted of possessing an unregistered still and carrying on the business of distilling whiskey without paying taxes. Prior to the trial, Miller attempted to use his Fourth Amendment rights to prevent his bank records from being introduced at the trial. In finding that the Fourth Amendment had not been violated, the Supreme Court stated that the government had not intruded into Mr Miller's "zone of privacy," i.e., "the security a man relies upon when he places himself or his property within a Constitutionally protected area." The Court went on to find that the bank records were not equivalent to a person's private papers and that, therefore, individuals have no legitimate expectation of privacy in their contents.

> . . . we perceive no legitimate 'expectation of privacy' in their contents. The checks are not confidential communications but negotiable instruments to be used in commercial transactions. All of the documents obtained, including financial statements and deposit slips, *contain only information voluntarily conveyed to the banks and* exposed to their employees in the ordinary course of business.[45]

In *Whalen* v. *Roe*,[46] physicians and patients brought suit against New York State challenging the constitutionality of a New York statute which required that physicians send to Albany a copy of every prescription that they write for certain types of narcotic drugs. The State placed its copy in a centralized computer bank. The statute also included security measures designed to protect the information once it came into the State's possession. In holding that the statute did not violate patients' or physicians' constitutionally protected right of privacy, the Supreme Court found that the enactment of the statute was a reasonable exercise of New York's broad police powers. The Court said that the State's interest in controlling the dissemination of narcotic drugs out weighed the privacy interest of the patients involved.

The *Whalen* opinion is at least somewhat encouraging for privacy advocates because the Court did recognize that the "accumulation of vast amounts of personal information in computerized data banks or other massive government files"[47] could constitute an unacceptable invasion of constitutional privacy rights, depending upon the government's purpose and its controls on redisclosure of the data.

In the period since these three Supreme Court decisions several Federal district and appeals courts have wrestled with the constitutional questions raised by the government's collection and use of personal information. The decisions recognize a limited constitutional right of information but in most instances rule against the privacy claimants. Thus the courts have upheld the government's collection of employee health data for research purposes;[48] the obtaining of "mail cover" information for law enforcement purposes;[49] the collection of psychological information for government employment purposes;[50] and the collection of personal financial information regarding public officials for regulatory purposes.[51]

It appears then that the courts recognize a vague constitutionally based right of information privacy that may put as yet undetermined restraints upon the government's handling of personal data.[52]

Common Law Standards

In the United States law comprises not only codified standards such as those found in the Constitution and in statutes and regulations but also a body of judicially based law inherited from the British that is called the "common law." The courts have used two common law doctrines to place protections upon the use of personal information. The principal focus of these common law doctrines is upon the information practices of private parties. By contrast, constitutional safeguards—such as they are—apply only to governmental behavior.

The first of the common law's information protection doctrines holds that some relationships, by the very nature of their purpose and participants, involve an implicit promise by the recordkeeper that the record subject's information will be kept confidential. Violations of this promise of confidentiality are viewed by the courts as a breach of the recordkeeper's contract. For example, both physicans and bankers have been held liable for unauthorized disclosures of personal information about their patients and customers.[53] Of course, this contract theory only provides protection for information maintained by recordkeepers that has a direct relationship with the record subject. When the information is maintained by third parties who have no direct contact with the subject, the contract theory does not apply.

The second common law theory used by the courts to set information privacy standards is tort law. A tort is a violation by one party of a duty owed to another party that can be remedied by the courts (usually in the form of injunctive relief and money damages). In common law several tort theories are available to penalize recordkeepers who improperly collect, maintain, use or disclose personal information. For example, a recordkeeper may be liable for defamation if the information disclosed is inaccurate and derogatory. Or a recordkeeper may be liable for intrusion or misrepresentation if the method of collection of the data is offensive, overbearing or deceptive.

In addition, in about three-quarters of the states, individuals can rely upon a common law right of privacy to seek damages (and injunctive relief) from parties who: (1) intrude upon an individual's freedom; (2) appropriate the individual's name or likeness; (3) publish personal information that places the individual in a false light; or (4) publish private facts about the individual.[54]

Statutory Standards

INITIAL LEGISLATION

The amorphous nature of the information privacy protections found in the Constitution and the fact that remedies found in the common law vary from state to state and are judicially dependent have helped to make the Congress and the state legislatures receptive to the enactment of statutory information privacy protections.

In the late sixties, the Congress enacted the first of several pieces of information privacy protection legislation. The Omnibus Crime Control and Safe Streets Act of 1968[55] prohibits individuals from using devices that are "primarily useful" to surreptitiously intercept conversations. The act was intended to effectively ban the use of wiretapping and eavesdropping devices in the private sector unless one of the parties to the conversation consented to the intercept. The development of electronic technology to intercept telephone conversations transmitted over microwave carrier channels may have outflanked the law. It is not clear whether this intercept technology comes within the act's "primarily useful" standard.

In 1970, the Congress passed the first of what might be called fair information practices legislation. The Fair Credit Reporting Act (FCRA)[56] sets standards for the collection, maintenance, use and disclosure of consumer credit and other financial and personal information by consumer reporting agencies (firms that perform personal investigations to provide background information about individuals relevant to consumer transactions). The FCRA requires consumer reporting agencies to apprise subjects of the information in their files and permits subjects to correct or rebut information with which they disagree. It also limits the use of certain obsolete and derogatory public record information, and it prohibits the disclosure of information in consumer reports except to customers who will use the data for purposes authorized in the act.

STUDIES IN THE EARLY 1970s

In the early seventies, two landmark studies laid the basis for the next stage in the enactment of information privacy legislation. In 1972, the National Academy of Sciences completed its study of the social and political effects of the computerization of personal information.[57] This empirically based study found that automation of personal data had not, despite popular fears to the contrary, had a negative effect on the development of privacy protection practices and policies. On the other hand, the study indicated that the computer's ability to store, collate and retrieve huge assemblages of personal data created new and greater threats to the preservation of personal privacy.

In 1973, the Department of Health, Education, and Welfare published its report articulating the five fair information practice principles quoted in the introduction to this paper.[58] Those five principles, as they applied to federally held personal information, were made law by the Privacy Act of 1974.[59]

PRIVACY ACT OF 1974

The Privacy Act is landmark legislation. Its provisions require Federal agencies to: (1) explain their record-keeping practices to record subjects and publish descriptions of the contents of and rules for their record systems; (2) provide record subjects with rights of access, correction and rebuttal; (3) comply with certain restrictions on collection of personal information; (4) meet certain data quality and record management standards in the maintenance of personal data; and (5) refrain from disclosing personal information without subject consent unless the disclosure comes within the exceptions for non-consensual disclosure.[60]

To date twelve states have enacted statutes similar to the Privacy Act, governing the handling of personal data in state and local file systems containing personal information.

Upon enactment of the Privacy Act, there was considerable fear that the statute would conflict with the Federal Freedom of Information Act (FOIA).[61] As a technical legal matter, the two statutes are not in conflict. The FOIA makes all publicly held written information available upon request unless the data qualify for withholding under one of the FOIA's nine disclosure exemptions. The Privacy Act prohibits agencies from disclosing personal data without subject consent unless the disclosure comes within one of the Privacy Act's eleven exemptions. One Privacy Act exemption permits agencies to disclose personal data without subject consent if the data *must* be disclosed under the FOIA.

Thus, the FOIA is the controlling statute in the Federal disclosure scheme. If the personal information does not qualify for withholding under an FOIA exemption, the data must be released, not withstanding the Privacy Act. On the other hand, if the personal information qualifies for withholding under the FOIA (and one FOIA exemption covers information "the disclosure of which would constitute a clearly unwarranted invasion of privacy"), the Privacy Act makes the data confidential and the agency must withhold them (even though the FOIA urges agencies to use their discretion in releasing data that qualify for withholding).

This relationship, though technically functional, has two problems. First, the Privacy Act extinguishes discretionary disclosures of personal data. The FOIA favors such disclosures, and thus something of an anomaly is created. A second, and more important, problem lies in the fact that it is the FOIA and not the Privacy Act which has the principal effect upon Federal disclosure policy for personal information. With a few exceptions, the FOIA requires agencies to release personal data unless the FOIA's vague and biased "clearly unwarranted invasion of privacy" standard is met.

The courts have said that the "clearly unwarranted" standard means that agencies and the courts should be biased in favor of disclosure and should weigh the public's interest in access against the subject's interest in privacy.[62] The effect appears to be to permit Federal agencies to disclose a substantial amount of personal data—particularly biographic data.

As for the Privacy Act, it has been criticized for failing to substantially reform Federal agency personal information practices. The Privacy Protection Study Commission's report calls for substantial Privacy Act amendments. In particular, the Commission recommended that the act be revised to: (1) clarify ambiguous language; (2) incorporate reasonableness tests to allow flexibility in implementation; and (3) apply the act to all personal records maintained by Federal agencies—as opposed to the current approach, which only applies the act to "systems of records" from which information can be retrieved by personal identifiers.[63]

Even where personal data need not be released under the FOIA, the Privacy Act has been criticized because its non-consensual disclosure provisions permit agencies to make numerous disclosures without obtaining subject consent. For example, the Privacy Act permits the sharing of personal information within Federal agencies on a "need-to-know" basis. Federal experience in implementing the Privacy Act and state experience with similar legislation suggest that the need-to-know principle does not create significant limitations. Most agencies appear to have interpreted the need-to-know standard to permit significant intra-agency exchange of sensitive personal information that one might reasonably have expected would receive very limited exposure.[64]

The problem is compounded in large, heterogeneous agencies that conduct many different programs. Agencies such as the Department of Health, Education, and Welfare (DHEW) are able to trade sensitive personal information on a "need-to-know" basis among their own units as diverse as the Social Security Administration, the Student Loan Office, the Indian Health Service, the National Institute of Mental Health and the Parent Locator Service. The Office of Management and Budget, which has nominal oversight of Privacy Act compliance, has urged HEW to avoid intra-agency transfer among its sub-agencies unless the data will be used for a purpose compatible with that for which they were collected.[65]

The Privacy Act disclosure provision that has received perhaps the most criticism is its "routine use" standard. This standard permits an agency to disclose personal data without subject consent if the purpose of the disclosure is "compatible" with the purpose for which the data were first collected (and if certain procedural requirements are complied with). Some observers charge that agencies have, on occasion, broadly interpreted the routine use provision to permit a wide variety of nonconsensual disclosures. For example, most agencies use the routine use exception to permit the disclosure of personal information that indicates possible law violations to the Department of Justice, notwithstanding the fact that in most instances the data were not collected for a law enforcement purpose.[66]

The Privacy Act has also received criticism on two other counts. First, the Congress did not provide for the effective oversight or regulation of agency compliance. In consequence, agencies seem to demonstrate differing degress of enthusiasm for the act and have at times adopted apparently inconsistent and self-serving statutory interpretations.

Finally, the Privacy Act's emphasis on subject awareness and vigilance in policing agency recordkeeping practices has been criticized. The effectiveness of the Privacy Act largely depends on record subjects' understanding of their rights and agency recordkeeping practices; their vigorous exercise of their rights to see and copy their records; and their willingness to administratively or judiciously challenge agency information practices. In actual practice, after an initial flurry of interest in records maintained by criminal justice, intelligence and defense agencies, record subjects now appear to be disinterested.

In the period since passage of the Privacy Act, two additional events have occurred that promise to have a significant impact upon the continued formulation of information privacy law.

PRIVACY PROTECTION STUDY COMMISSION REPORT

The first event was the publication, in July of 1977, of the Report of the Privacy Protection Study Commission. The Commission produced a 654-page report containing over 150 legislation recommendations. In large part the report focused on two areas of information policy left unaddressed by the Privacy Act. First, the Commission dealt with the information privacy standards that ought to apply to private sector handling of personal records. Second, the Commission report focused on the creation of privacy safeguards for government access to sensitive personal information maintained by nongovernmental, institutional recordkeepers.

The Commission's recommendations for private sector recordkeepers, such as credit grantors, insurers and employers, are likely to shape the nation's approach to the handling of personal information by those industries. Indeed, the Commission's recommendations have been studied by a Carter Administration task force and, with relatively little change, incorporated into the Administration's privacy program. Legislation to implement that program is now being sent to the Congress.

The Commission's approach to the reform of private sector personal recordkeeping is based upon five policy decisions or strategies: (1) recordkeeping standards should be industry specific—not omnibus; (2) the protection of information privacy must depend primarily upon subject participation rights (such as the subject's right to be given an explanation of recordkeeping practices, the subject's right of access and correction, the subject's right to be able to consent specifically to collection or disclosure of his personal data and the subject's right to bring a civil action for recordkeeping violations); (3) recordkeepers should retain wide discretion to set standards for the type and amount of personal data collected; (4) recordkeepers should retain wide discretion to set standards for data quality and the management and use of personal information within their organization; and (5) record subjects should have an expectation that their personal data will be kept confidential—however, information privacy legislation should not include detailed confidentiality standards that permit or prohibit specific disclosures.

FINANCIAL PRIVACY ACT OF 1978

To date, one group of Privacy Commission recommendations has become law. The Commission's recommendations for the imposition of controls on government access to personal data served as a basis for passage of the Right to Financial Privacy Act of 1978.[c7] The statute also reflects the Congress' displeasure with the Supreme Court's holding in *United States* v. *Miller*, discussed earlier, which strips bank customers of privacy rights in cases where the government seeks access to customer bank records.

The Financial Privacy Act contains four key elements. First, the act requires Federal Government agencies (but not state and local agencies or the private sector) to make all requests for customer bank records in writing. For most agencies the written request must be in the form of a subpoena. Thus the act puts an end to the days when Federal agents could flash a badge and expect bank officers to open their records.

Second, with certain exceptions, Federal agencies must send a copy of the subpoena or other written request to the customer at the same time the subpoena is served upon the financial institution.

Third, the Financial Privacy Act gives the customer an opportunity to contest the disclosure in court. Once in court, the customer must be able to show that his records are not "relevant to a legitimate law enforcement inquiry."

Fourth, the act places limits upon interagency transfer of financial records obtained pursuant to the act.

POLICY OBSERVATIONS

The Information Privacy Framework

At the close of the seventies, a framework for protecting information privacy has been fully articulated. What's more, the framework has been largely implemented (through the Privacy Act) at the Federal Government level and partially implemented (through parallel statutes in twelve states) at the state and local government level. Implementation of information privacy safeguards for the private sector is just beginning.

The framework, as previously discussed, includes in its most comprehensive form five types of provisions. First, information privacy schemes include open record provisions that require the recordkeeper to identify and describe his record systems (to subjects and/or the public) and to explain to record subjects his information practices. This explanation usually includes an identification of: the type of personal information to be collected; its uses; anticipated third party disclosures; the legal recordkeeping authority; and a review of the subject's access, and other participation rights.

Second, the information privacy framework includes subject participation rights. The record subject usually has the right to see and copy his file, to challenge and rebut its accuracy, to consent to or reject proposed collection and (more often) disclosure practices, and to administratively and/or judicially challenge recordkeeping practices that the subject believes violate his information privacy rights.

Third, information privacy schemes often include modest restrictions on the recordkeeper's collection of personal data. For example, with certain exceptions, the Privacy Act prohibits agency collection of data about individuals' exercise of their First Amendment rights. Restraints on collection are perhaps the most effective privacy protection mechanism because information that is never obtained in the first place, quite obviously, cannot be used to harm the individual. On the other hand, legislators have appeared to be wary of imposing substantive collection restrictions. This reluctance is probably attributable to the fact that regulation of the type of information that a recordkeeper can obtain usually has a significant effect upon the recordkeeper's ability to set standards and criteria for operating his enterprise. Instead, access-type restrictions that affect the method of collection, such as those found in the Financial Privacy Act, are far more common.[68]

Fourth, the information privacy framework sometimes includes detailed data quality and record management standards. The Privacy Act, for example, sets standards for the accuracy, relevance, timeliness and completeness of personal data maintained by Federal agencies as well as requiring compliance with employee training, security, audit and other types of record management standards. As another example, the Bank Secrecy Act imposes extensive data maintenance requirements on most financial institutions. It mandates that those institutions keep microfilm records of checks and keep records of most international and large domestic transactions.[69]

The fifth type of provision ordinarily found in information privacy frameworks is confidentiality standards. Typically these standards prohibit disclosure of personal data to third parties unless the subject consents to the disclosure or unless the disclosure comes within an exception doctrine. To date the exception doctrines in most privacy codes have been broadly framed. The generous disclosure exceptions in the Privacy Act have already been discussed. The Privacy Commission's recommendations for private sector information privacy include three broad categories of disclosures that can be made without obtaining subject consent: disclosures made to service the recordkeeping relationship; disclosures made to serve the interests of the record subject; and disclosures made to serve the interests of the recordkeeper or society.

INFORMATION PRIVACY CHALLENGE LIKELY TO GROW

In looking ahead to the 1980s, there is every reason to believe that the challenge to information privacy will continue to grow. Earlier in the paper, five factors were cited as contributing to the emergence of information privacy as an issue of public and political importance: (1) complexity of society; (2) the growth of government; (3) notions of the meritocracy, the "credential" society; (4) diversity and distrust in society; and (5) advances in surveillance and information and telecommunications technology. Not one of these phenomena is likely to disappear and, indeed, some of these developments can be expected to intensify.

For example, the Privacy Commission's report, despite the acknowledged difficulties of technological forecasting, concludes that rapid progress in the development and implementation of information technology can be expected through the 1980s.[70] Developments in telecommunications technology are likely to keep pace. It also seems reasonable to expect continued advances in surveillance technology.

Indeed, experts predict that two areas of surveillance technology in particular are likely to see substantial technical improvements. First, personal surveillance devices that track the whereabouts of the target, permit one- or two-way communication and perhaps even monitor the target's physical or emotional condition are just at the state of the art. The devices could be worn voluntarily by those with a desire for such monitoring or, according to some informal proposals, worn involuntarily by probationers and parolees, aliens or other people of law enforcement or government interest.[71]

Second, many experts believe that more effective devices for measuring veracity, emotional and psychological condition, and perhaps even thought patterns, are likely to be developed in the near future.[72]

Interestingly, the Harris survey found that in just the last year the extent of the public concern about privacy had jumped substantially.

> The American people are greatly concerned about threats to their personal privacy. This concern is pervasive throughout society. Public concern about threats to personal privacy has jumped 17 percentage points in the past year—from 47% in January of 1976 to 64% in December 1978.[73]

The Privacy Commission's report also warns that information privacy problems are likely to intensify. In particular, it concludes that if individuals are to enjoy information privacy rights, policymakers in the 1980s will have to focus upon four problems: greater amounts of personal data and more detailed personal data continue to be collected and retained by organizational recordkeepers, especially in the credit and financial areas; increasingly, recordkeepers maintain such data for purposes unrelated to servicing their relationship with the data subject; increasing personal data are maintained by organizations who in fact do not have a direct relationship with the data subject and who have, therefore, obtained the data from secondary parties; and finally, neither existing law nor technology give the data subject the tools that he needs to protect his information privacy.[74]

All of these developments presage a continued loss of control or power by the record subject over the handling of his personal information. The information privacy framework represents an attempt to protect and restore some of the record subject's power.

POLICY CONSIDERATIONS IN THE 1980s

In framing information privacy safeguards for the private sector as well as refining safeguards that already exist in the public sector, policymakers should take into account a number of considerations.

First, it must be recognized that there is probably little prospect that a consensus will soon emerge regarding the degree to which information privacy rights should be protected or extended. Instead, there are likely to continue to be sharp disagreements about where the balance should be set between a subject's control of his personal information and organizational discretion to collect, retain, use and disclose personal information. Although there are many ways to look at this conflict, ultimately it is a matter of weighing values that invigorate the democracy and nourish personal freedom against values that minimize societal risks, reduce costs and improve organizational effectiveness and efficiency. It is probably unrealistic to imagine that our society will ever decide this conflict in a way that is wholly satisfactory to all groups or that is permanent. Perhaps the most that can be said at this stage is that due to rapid changes in society's need for and use of personal information and changes in information technology, there is today a perceived need to carefully analyze the effectiveness as well as the effects of doctrines that safeguard information privacy.

There is also some consensus today, reflected in the Privacy Commission's report and the President's privacy legislation, that comprehensive information privacy safeguards, somewhat modeled after the Federal Privacy Act, should be extended to the private sector. Although there is considerable empirical and policy support for such a position, policymakers should also recognize that private sector information regulation represents a major step. Inevitably it raises questions about the imposition of excessive government regulation and unnecessary costs.

In regard to specific components of an information privacy framework, practical as well as conceptual problems appear to plague at least a few of these components. For example, logic and experience now suggest that policymakers may not be able to rely upon subject participation rights as a key mechanism for protecting information privacy. Even armed with an arsenal of participation rights such as entitlement to a record practices explanation, and access, challenge and consent rights, many record subjects are not in a position nor, apparently, do they have the inclination to challenge institutional recordkeeping practices.

Several factors appear to support this hypothesis. For one, if it is true, as the Privacy Commission states, that "more and more records about individuals are collected, maintained and disclosed by organizations with which the individual has no direct relationship," such individuals are likely to be in a poor position to police the recordkeeping activities of those organizations. Furthermore, if it is also true that organizations increasingly pay more attention to what other organizations say about the record subject rather than paying attention to what the record subject says, then that individual's chances of participating effectively in the process may be further reduced.

What's more, as society becomes more complex, the nature of its information practices also becomes more complex. This often makes institutional information practices less amenable to personal monitoring and intervention. Finally, many of an individual's most important recordkeeping relationships are with large or intimidating entities such as government agencies, hospitals, employers and various financial organizations. Typically, individuals have little leverage or bargaining power in their dealings with such institutions and may not, therefore, be able to aggressively guard their information rights.

As noted earlier, experience to date with the Privacy Act and similar state legislation mostly bears out these arguments. For the most part, individuals have not used their participation rights aggressively to monitor and police the handling of their personal data. Partly in recognition of this but primarily out of concern for cost, the Congress just amended the recently enacted Financial Privacy Act to delete a requirement that financial institutions distribute to all of their customers an explanation of their notice and challenge rights.[75]

This is not to say that subject participation rights should be eliminated or even curtailed. Rather, the point to be made here is that the emphasis on subject participation in current versions of the information privacy framework, particularly as articulated by the Privacy Commission, may be ill advised. If information privacy is to be expanded and strengthened in the 1980s, it may be that strategies other than subject participation will have to be the cornerstone.

Another agenda item for information policymakers in the 1980s is the need to consider third party disclosure safeguards. To date, confidentiality provisions included in most information privacy schemes have been modest and limited, serving more to endorse existing disclosure practices than to safeguard privacy interests. Confidentiality standards such as those found in the Privacy Act include general exceptions for many noncriminal disclosures.

This hypothesis, if correct, is alarming in two respects. First, information privacy protection schemes, including those found in the Privacy Act, the Privacy Commission proposals, and the Carter Administration's proposed legislation, are usually touted as instruments that accomplish or would cause a significant reform of institutional disclosure practices. Although the proponents of these schemes are by no means attempting to be deceptive, the net effect may be to give the public the false security that information flows that are thought to be unsatisfactory have been or will be significantly reformed.

The second problem caused by ineffectual confidentiality standards is more substantive. Assuming that there is a legitimate need in this society to create mechanisms to safeguard information privacy, disclosure standards are critical. This is especially so if, as suggested above, subject participation rights may be somewhat ineffectual.

Thus it may be that the effective protection of information privacy will require the development and adoption of rules that specifically and definitively prohibit certain kinds of third party disclosures. The Family Educational Rights and Privacy Act of 1974, popularly known as the Buckley Amendment, essentially takes this approach. It prohibits all non-consensual disclosures of sensitive student information except for nine specifically defined types of disclosures. The statute lacks the general, open-ended exceptions for non-consensual disclosures characteristic of many other schemes. After four years, it does not appear that the Buckley Amendment's approach is unduly costly or burdensome.[76]

Policymakers probably will also have to take a long look at the wisdom of imposing restrictions on both the method and the content of data-collection standards.*

Already many standards exist that police the method of obtaining personal information. The Constitution, for example, restricts searches and seizures, compelled self-incrimination and certain kinds of interrogation practices. Federal and state statutes prohibit wiretapping and eavesdropping, and statutes in over a dozen states limit the use of the polygraph. Federal legislation has been introduced that would effectively ban the private use of the polygraph.[77]

A review of information privacy law as well as related proposals and studies suggests three guiding principles for framing information collection standards. First, policymakers should consider the principle that individuals should always know that information about themselves and identified with them has been obtained. If the data are obtained from another recordkeeper, the subject can have such notice if he has the opportunity to consent to the disclosure or if special notice and challenge rights are given to the subject—as in the Financial Privacy Act of 1978—or if he has general knowledge that disclosures of this kind can be made. If the data are obtained directly from the individual, there is probably a consensus that the subject should have specific knowledge. Pretext interviews, wiretapping and covert truth detection devices are offensive precisely because the data target does not know about, or at least fully appreciate, the nature of the information collection process.

Second, policymakers should consider the principle that individuals should almost *always* have the *right* to consent to the collection of data when they are the source of the data. This principle shares the rationale underlying the Fifth Amendment's protection against compelled self-incrimination.

Third, policymakers should consider the need for a principle that states that individuals, even when they consent, should not, when their interests are threatened, be the target of devices that take information from them in a manner that they cannot control. Truth detection devices, brain-wave machines and devices that monitor emotional status, even when the subject knows about and consents to the test, all have the capability of taking information from the subject with relatively little reference to the subject's volition or control. In this sense these machines are innately intrusive and arguably incompatible with our society's view of the individual, who finds himself in an adversarial position.

Another group of information privacy mechanisms that deserves attention from policymakers in the 1980s are various types of data management standards. For example, policymakers should consider standards that permit the disclosure and collection of personal data but restrict its use. Examples of such standards are not uncommon. For example, the Civil Rights Act of 1964 does not prohibit the collection, retention or disclosure of information about an individual's race, color, religion, sex or national origin. However, it does prohibit the *use* of such personal information to make job-related decisions.[78]

Another example of a type of limitation imposed upon the use of data is the broad distinction that is drawn between research and statistical records and administrative records. Research and statistical records are usually, but not always, in a personally non-identifiable form and, by definition, are compiled for a nonpersonal research or statistical purpose. Administrative records are recorded in a personally identifiable form and are used to make direct decisions about the individual.

The Privacy Commission has recommended that research and statistical records be kept functionally separate from administrative records and that data from this source never be permitted to be used to make decisions directly affecting the subject.[79] The Commission's recommendations have been generally reflected in legislation introduced in the 96th Congress by the Carter Administration.[80] Thus, congressional policymakers will have an opportunity to consider the wisdom of imposing this principle of personal information use and data management on the private sector.

Various other data management principles for protecting information privacy have also received some at-

* In one sense data collection is simply another way of looking at data disclosure because disclosure practices of one entity affect the collection practices of other entities. However, it is useful to consider collection issues separately insofar as they: (1) focus on the conduct of the party obtaining data as opposed to transmitting data; and (2) pertain to the obtaining of data directly from the data subject.

tention. Policymakers may want to consider, for example, standards that call for the organization of personal record systems in such a way that medical information, disciplinary or law enforcement information, financial information and other sorts of very sensitive personal data are maintained alone and separately from other data elements. The use of automated information technology probably makes this type of strategy more feasible.

Various organizations and jurisdictions have also experimented with record management strategies that segregate public and non-public information[81] or that actually prohibit the maintaining of certain types of information in an automated form.[82]

Considerable thought should also be given to whether standards should be set to control the data elements that can be used to retrieve personal information from record systems. It can be argued that the use of non-identifiers to obtain *cumulative* personal data from record systems—a process that is eminently feasible in advanced automated systems—should be discouraged, at least in some environments. Automated police blotters and booking systems, as well as numerous public record systems, come to mind.

Another consideration for information policymaking is the pressing need to monitor developments in information technology. The Privacy Commission cites this as one of the requirements that must be met in order to construct an effective information privacy framework.

The Commission's report also stresses the difficulty of such monitoring. It assumes that more or less constant monitoring is not possible and thus recommends that "trigger mechanisms" be identified that will focus public attention upon particular technologies or processes that have significant information handling implications.[83]

The need for monitoring also raises the question of who would perform this activity. Many of the European privacy schemes require the licensing of data banks on the theory that the licensing requirement permits some overview and control of developments in information technology.

In this country the courts have often been left the task of monitoring, evaluating and responding to advances in information and surveillance technology. For example, the courts spent nearly four decades considering advances in eavesdropping technology before abandoning the legal requirement that there be a physical trespass or intrusion before an eavesdrop is illegal.[84]

Most observers believe that the adversarial process is ill-suited to the process of technological assessment. The speed of technological change,[85] its complexity[86] and its frequent partnership with government combine to frustrate effective courtroom evaluation.

The need to monitor developments in information technology and the executive branch's role in the development of technology may mean that increasingly the responsibility for framing information privacy strategies and remedies may be left to the executive branch.[87] This development places a special responsibility on the legislative branch to use its research and public hearing capabilities to participate in the technology assessment and policymaking process.[88]

A final and related policy consideration is the need to consider the development of a regulatory strategy for implementation and enforcement of information privacy standards. The Privacy Commission cites the anti-government, "deregulation" mood of the country as an obstacle to the creation of an agency to enforce compliance with privacy requirements. The recent Harris poll reaffirms that a majority of Americans do not favor the creation of a new governmental privacy regulatory agency.

The public policy debate should focus on the issue of whether it is possible to effectively reform information practices or protect information privacy without a regulatory presence or several regulatory presences. Because information privacy safeguards work to transfer some degree of power from recordkeepers to record subjects, it is probably not realistic to imagine the accomplishment of such shifts in power without regulatory impetus and oversight. The Privacy Act's implementation experience largely supports this supposition.

Interestingly, Western European comprehensive information privacy statutes and proposed legislation invariably include the establishment of a government data protection board or commission with regulatory and enforcement powers.[89] It may well be that in the 1980s the American public will have to choose between its desire to implement effective information privacy safeguards and its desire to avoid the creation of new governmental regulatory entities.

CONCLUSION

In the last 15 years major strides have been made in recognizing and safeguarding information privacy. However, more work needs to be done in developing an ongoing information technology assessment capability, refining the information privacy framework, developing new and alternate approaches to privacy protection and developing strategies for implementing information privacy safeguards in a comprehensive and effective manner.

NOTES

1. The HEW study has been published in many formats. One of the most widely circulated is *Records, Computers and the Rights of Citizens*, MIT Press (1973). See pp. xxiii–xxxv of the summary and recommendations.

2. The United States Supreme Court has recognized these two distinct types of privacy interests. In *Whalen* v. *Roe*, 429 U.S. 589

(1977). pp. 599-600. the Court rejected patient information privacy claims aimed at overturning a New York State drug reporting law. The Court distinguished information privacy claims from behavioral privacy claims.

> Appellees contend that the statute invades a constitutionally protected 'zone of privacy.' The cases sometimes characterized as protecting 'privacy' have in fact involved at least two different kinds of interests. One is the individual interest in avoiding disclosure of personal matters. and another is the interest in independence in making certain kinds of important decisions.

3. For a discussion of information privacy and democratic values. see. Wheelor. S., *On Record: Files and Dossiers in American Life* (1969); Westin. A., *Privacy and Freedom* (1968).

4. For a discussion of information privacy and libertarian values. see. Bloustein. "Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser." *N.Y.U.L. Rev.* 39:962 (1964); Goffman. E., *Behavior in Public Places*. Collier. MacMillan (1963); Henkin. L., "Privacy and Autonomy." *Colum. L. Rev.* 74:1410 (Dec. 1974); and Beaney. "The Right to Privacy and American Law." *Law and Contemporary Problems* 3:253 (1966).

5. See. for example. Askin. Frank, "Surveillance: The Social Science Perspective." *Surveillance. Dataveillance and Personal Freedoms: Use and Abuse of Information Technology*. Burdick Press (1973). p. 73.

> Democratic politics is predicated on the belief that people can reason. grow and change. Anxious men are rarely free men. Those tormented by the thought that it may have been an exercise of freedom that cost them their jobs. deferments. civil rights or liberties—but who never know for sure—are not likely to develop the strengths political democracy needs. Surveillance serves to chill thought and to discourage the risks of freedom.

6. Bouvard. Marguerite Guzman. and Bouvard. Jacques. "Computerized Information and the Effective Protection of Individual Rights." *Society* 12:62. Sept.-Oct. 1975.

7. See. Westin. A., *Privacy and Freedom*. Atheneum Press (1968). pp. 23-65.

8. See. for example. Packard. V., *Naked Society* (1964); Brenton. M., *Privacy Invaders* (1964); Westin. A., *Privacy and Freedom* (1968); Rossenberg. J., *The Death of Privacy* (1969); and Miller. A., *Assault on Privacy* (1971).

9. Harris. Louis. and Associates. and Westin. Dr. Alan. *The Dimensions of Privacy*, Sentry Insurance (May 1979). p. 3.

10. Ibid., p. 5.

11. See. for example. *Personal Privacy in an Information Society*. The Report of the Privacy Protection Study Commission (1977); *Records. Computers and the Rights of Citizens*. MIT Press (1973); Westin. A., and Baker. M., *Databanks in a Free Society*. Quadrangle (1972); *Federal Personal Data Systems Subject to the Privacy Act of 1974*. First Annual Report to the President. Office of Management and Budget (1975).

12. See. *Personal Privacy in an Information Society*. The Report of the Privacy Protection Study Commission (1977). p. 41.

13. Porat. Marc U., "Communication Policy in an Information Society." *Communications for Tomorrow*. Praeger (1978). pp. 34-35.

14. Ibid.

15. See. Reich. "The New Property." *Yale L.J.* 73:733 (April 1967). for one of the most influential discussions of the government's emergent role as a dispenser of "largess." See also, "Government Information and the Rights of Privacy." *Mich. L. Rev.* 73:My-Je 1975 (entire issue). One leading scholar has pronounced privacy to be terminally ill in the modern state. He cites the role of the government as an active force as a death knell for privacy. See Miller. A. S., "Privacy in the Modern Corporate State: A Speculative Essay." *Admin. L. Rev.* 25:231 (Summer 1973).

16. See. Westin. A., and Baker. M., *Databanks in a Free Society*. Quadrangle (1972). pp. 342-343.

17. Ibid., pp. 343-345.

18. See. *Surveillance Technology. Policy and Implications: An Analysis and Compendium of Materials*, Staff Report of the Subcommittee on Constitutional Rights of the Senate Committee on the Judiciary. 94th Cong., 2nd Sess. (1976).

Note too. recent developments in pen register technology. The pen register is a tape recorder/computer which can monitor all numbers called from a phone. the time the calls were made. how long they lasted and when incoming calls were received. An internal relay for switching on an external recorder can be attached to the pen register which will allow the person to listen in on the telephone conversation. According to press reports. AT&T is incorporating into its new electronic switching system an automatic pen register capability which will allow an operator to press a switch and receive over a computer screen the numbers of calls both local and long distance which have been made by a person. AT&T wishes to incorporate the pen register capability in order to eventually be able to have "usage sensitive pricing" which will allow this system to automatically record dates. times and duration of local as well as long distance calls from one phone.

19. See. for example. "No Place to Hide; NSA Eavesdropping Technology." *Newsweek* 86:19-21. Sept. 8. 1975.

20. Ibid., as quoted from a "Meet the Press" television program.

21. See. Belair. R. and Bock. C., "Remote Camera Surveillance of Public Streets." *Colum. Human Rights L. Rev.* 4:144 (1973); "Police Helicopter Surveillance and Other Aided Observations—The Shrinking Reasonable Expectation of Privacy." *Calif. West L. Rev.* 11:05. Sp. 1975.

22. Rowan. *Technospies*. G. P. Putnam's Sons (1978). pp. 149-164.

23. Borchgrave. "Space Age Spies." *Newsweek* (Mar. 6. 1975). p. 37.

24. See, for example. Klass. "Infrared Sensor Sized for Helicopters." *Aviation Week* 108:71-73. Jan. 9. 1978.

25. "Tracking Dangerous People: the Electronic Watchdog We Shouldn't Use." *Psychology Today* 8:84. Jan. 1975.

26. Rowan. *Technospies*. G. P. Putnam's Sons (1978). pp. 157-158.

27. Ibid., p. 220.

And. see. "Privacy: the Polygraph in Employment." *Ark. L. Rev.* 30:35-48. Sp. 1976. For a criticism of the use of the polygraph. see. Shattuck. Brown and Carlson. "The Lie Detector as a Surveillance Device." *ACLU Reports*, Feb. 1973.

28. See. Shattuck. Brown and Carlson. "The Lie Detector as a Surveillance Device." *ACLU Report* (Feb. 1973). pp. 2-6.

29. "End of Trust. Voice Analyzer Watches." *Esquire* 91:8+ Apr. 24, 1979.

30. "Don't Look Now But Your Eyes Reveal Those Little White Lies." *Science Digest* 83:29-33. May 1978.

31. Harris. Louis and Associates and Westin. Dr. Alan F., *The Dimensions of Privacy*. Sentry Insurance (May 1979). p. 3.

32. *Technology and Privacy*. Appendix Five to the Report of the Privacy Protection Study Commission (1977). pp. 10-15.

33. Ibid., p. 1.

The best and most comprehensive discussion to date of the computerization of personal data can be found in the report of the National Academy of Sciences Project on Computer Data Banks. published as Westin. A. and Baker. M., *Databanks in a Free Society*. Quadrangle (1972). See also. "Computerization of Government Files: What Impact Upon the Individual?" *UCLA L. Rev.* 15:137. Sept. 1968; Pemberton. J. D., Jr., "Symposium: Computers. Data Banks and Individual Privacy . . . on the Dangers. Legal Aspects and Remedies." *Mich. L. Rev.* 53:211. Dec. 1968; and Karst. K. L., "The Files: Legal

Control Over the Accuracy and Accessibility of Stored Personal Data," *Law and Contemp. Problems* 31:251, Sp. 1966.

34. Rowan, *Technospies*, G. P. Putnam's Sons (1978), pp. 66-67.

35. Kay, Alan C., "Microelectronics and the Personal Computer," *Scientific American* 237:230-232, September 1977.

36. Bouvard, Marguerite Guzman, and Bouvard, Jacques, "Computerized Information and Effective Protection of Individual Rights," *Society* 12:62, September-October 1975.

37. Flannery, John P., "Commercial Information Brokers," *Surveillance, Dataveillance and Personal Freedoms*, Burdick Press (1973), pp. 224-225.

38. *Technology and Privacy*, Appendix Five to the Report of the Privacy Protection Study Commission (1977), p. 27.

39. Ibid., p. 29.

40. Ibid., p. 35.

41. See, for example, Laskin, Paul L., and Chayes, Abram, "International Satellite Conflict," *Society* 12:30, Sept.-Oct. 1975; Saenz, A. V., et al., "Telecommunications with Nutrino Beams," *Science* 198:295-297, Oct. 21, 1977.

42. 426 U.S. 693 (1976).

43. Ibid., p. 712.

44. 425 U.S. 435 (1976).

45. Ibid., p. 442.

46. 429 U.S. 589 (1977).

47. Ibid., p. 605.

48. *DuPont* v. *Finklea*, 442 F. Supp. 821 (D. Del. 1977); *C. F. General Motors* v. *Finklea* (S. D. Ohio, Oct. 10, 1978).

49. *United States* v. *Choate*, 576 F.2d 165 (9th Cir. 1978).

50. *McKenna* v. *Fargo*, 451 F.Supp. 1355 (D. NJ. 1978).

51. *Plante* v. *Gonzalez*, 575 F.2d 1119 (5th Cir. 1978).

Note that the Supreme Court has also considered whether government officials have privacy rights that limit public access to their personal data. In *Nixon* v. *Administrator of General Services*, 433 U.S. 425 (1977), the Court rejected former President Nixon's claim that his constitutional privacy rights would be violated by permitting the Archivist to screen his papers and tape recordings in order to preserve those of historical value.

52. In *Plante* v. *Gonzalez*, 575 F.2d at 1134, the Federal Fifth Circuit Court of Appeals complained that the Supreme Court's information privacy standards are vague.

> The Supreme Court has provided little specific guidance. Although in the autonomy strand of the right to privacy [impact on behavior] something approaching equal protection 'strict scrutiny' analysis has appeared, we believe that the balancing test, more common to due process claims, is appropriate here. The constitutionality of the [reporting statute] will be determined by comparing the interests it serves with those it hinders.

53. See, for example, *Hammonds* v. *Aetna Casualty and Surety Co.*, 243 F.Supp. 793 (N.D. Ohio 1965); *Milohnick* v. *First National Bank of Miami Springs*, 224 So.2d 759 (Ct. of Apps. Fla. 1969); and see, "Doctor-Patient Relationship: Physicians Breach of Confidentiality Gives Rise to Cause of Action," *Cumber-Sam L. Rev.* 5:169, Spring 1974.

54. Note that in most of these states privacy tort law has statutory recognition. The existence of a privacy tort at common law is disputed by many courts.

The move to enact right to privacy tort legislation is generally credited to a now famous law review article written by Samuel Warren and Louis Brandeis, "The Right to Privacy," *Harv. L. Rev.* 4:193 (1890).

Note that common law and statutory restrictions on the dissemination of personal information may run afoul of individuals' First Amendment rights of free speech. In a series of decisions the Supreme Court has substantially narrowed the extent to which privacy standards can be used to restrict speech about individuals involved in public events, and particularly individuals involved in events of political importance. See, *New York Times* v. *Sullivan*, 376 U.S. 254 (1964); *Time, Inc.* v. *Hill*, 385 U.S. 374 (1967); and, more recently, *Gertz* v. *Robert Welch, Inc.*, 418 U.S. 323 (1974); and *Time, Inc.* v. *Firestone*, 424 U.S. 448 (1976).

55. 18 U.S.C. 2510 et seq.

56. 15 U.S.C. 1681 et seq.

57. Published by Westin, A., and Baker, M., as *Databanks in a Free Society*, Quadrangle (1972).

58. Published as *Records, Computers and the Rights of Citizens*, MIT Press (1973).

59. 5 U.S.C. 552a.

60. Note that during the same period the Privacy Act was passed, the Congress also passed the Family Educational Rights and Privacy Act of 1974, 20 U.S.C. 1232g (popularly known as the Buckley Amendment). The Buckley Amendment provides parents and students with rights of access and correction for most student records held by most educational institutions and limits the non-consensual disclosure of such data.

61. 5 U.S.C. 552.

62. See, *Getman* v. *N.L.R.B.*, 450 F.2d 670 (D.C. Cir. 1971); and *Robbles* v. *E.P.A.*, 484 F.2d 843 (4th Cir. 1973).

63. See, "The Privacy Act of 1974: An Assessment," Appendix Four to The Report of the Privacy Protection Study Commission (1977), p. 77.

64. The description of the Privacy Act implementation experience is based upon the PPSC report cited at footnote 63 and the following published analyses: Belair, R., "Agency Implementation of the Privacy Act and the Freedom of Information Act: Impact on the Government's Collection, Maintenance and Dissemination of Personally Identifiable Information," *J. Marshall J.* 10:465 (Sp. 1977); "Symposium—Openness in Government—A New Era," *Fed. B.J.* 34-279 (Fall 1975); "Access to Information; Exemptions from Disclosure Under the FOIA and the Privacy Act of 1974," *Willamette L.J.* 13:135 (Winter 1976); "Information Privacy and Public Records," *Pacific L. J.* 8:25 (Jan. 1977); *Annual Report to the President*, Regarding Agency Compliance with the Privacy Act. Office of Management and Budget, 1976-1978.

65. Letter from Office of Management and Budget to Mr. John Ottina, HEW Assistant Secretary for Administration and Management, dated Dec. 15, 1975. See, Belair, "Agency Implementation of the Privacy Act and Freedom of Information Act: Impact on the Government's Collection, Maintenance and Dissemination of Personal Information," *J. Marshall J.* 10:465, 503 (Sp. 1977).

66. See, Memorandum from the Attorney General to the Heads of all Executive Departments and Agencies, Office of the Attorney General, Washington, D.C., on the "Implementation of the Privacy Act of 1974 Routine Uses of Information," June 5, 1975.

67. Title XI of H.R. 14279, P.L. 95-630, signed into law on November 10, 1978; effective date March 10, 1979.

68. Note, that the Tax Reform Act of 1926 applies notice and challenge procedures to IRS access to personal records held by third parties (7609 of the *Internal Revenue Code*). In addition, comprehensive medical record legislation now before the Congress contains government access controls of the type found in the Financial Privacy Act. See, for example, S.503. Privacy Act Amendments of 1979; Confidentiality of Medical Records, introduced by Senators Javits, Ribicoff and Muskie.

69. P.L. 91-508; 12 U.S.C. 1730d, 1829b, 1951-1959; 31 U.S.C. 1051-1062, 1081-1083, 1101-1105, 1121 and 1122.

70. See, "Technology and Privacy," Appendix Five to the Report of Privacy Protection Study Commission (1977), pp. 61-86.

71. See, Rowan, *Technospies*. G. P. Putnam's Sons, pp. 157 and 231. Rowan quotes Joseph Meyer, a computer expert for the National Security Administration, as suggesting that those on bail, probationers and parolees be forced to wear miniature tracking devices. Apparently

Meyer envisions that the transponders could also be used for "monitoring aliens and political subgroups" and acting as an "externalized conscience" for the targets. The Center for the Prevention of Violence at UCLA has also done work on transponder technology.

72. "Aura Phenomenon Puzzles Experts; Kirlian Photography," *Smithsonian* 8:109-114, April 1977; and see "No Place to Hide; NSA Eavesdropping Technology," *Newsweek* 86:19-21, Sept. 8, 1975.

73. Harris, L. and Associates and Westin, Alan F., "A National Opinion Research Survey of Attitudes Toward Privacy," Sentry Insurance (May 1979), p. 5.

74. See, "Technology and Privacy," Appendix Five to the Report of the Privacy Protection Study Commission (1977), pp. 17-18.

75. S. 37, 96th Cong., 1st Sess., amending P.L. 95-630.

76. 20 U.S.C. 1232g(b)(1); and see "The Buckley Amendment: Opening School Files for Student and Parental Review," *Cath. U.L. Rev.* 24:588, Sp. 1975.

77. S. 854, the "Polygraph Control and Civil Liberties Act of 1979," introduced by Senator Birch Bayh.

78. Title VII of the Civil Rights Act of 1964, 42 U.S.C. 200e et seq.

79. See, *Personal Privacy in an Information Society*, the Report of the Privacy Protection Study Commission (1977), ch. 15, p. 567.

80. S. 867; H.R. 3409.

81. The Federal Trade Commission and the Food and Drug Administration, among others, organize some of their files according to exempt (non-public) and non-exempt (public) documents. Belair, "Agency Implementation of the Privacy Act and the Freedom of Information Act: Impact on the Government's Collection, Maintenance and Dissemination of Personally Identifiable Information," *J. Marshall J.* 10:466, 485 (Sp. 1977).

82. Iowa's Criminal History and Intelligence Data Statute, for example, prohibits criminal intelligence data from being "placed within a [illegible] data storage system" (749 B.8. of the *Iowa Code Ann.*). See [illegible] v. *Miller*, 383 NYS 2d 95 (1976), which held that the computerization of personal medical records in a state's information system did not violate the subject's constitutional privacy rights.

83. "Technology and Privacy," Appendix Five to the Report of the Privacy Protection Study Commission (1977), p. 35.

84. The breakwater case of *Katz* v. *United States*, 389 U.S. 347, 353 (1967), swept away the requirement that the government engage in a trespass or physical intrusion before the protections of the Fourth Amendment become applicable. In now famous words, the Court declared that "the Fourth Amendment protects people and not simply areas," and, therefore, where a person "justifiably relies" on privacy, it is an unreasonable search for the government to intrude either physically or in any other manner without a warrant. (In *Katz*, the government intercepted the defendant's conversations by placing a wall mike on the ceiling of a telephone booth he frequented.) The *Katz* decision thus overturned the Fourth Amendment's trespass requirement enunciated in *Olmstead* v. *United States*, 277 U.S. 438 (1928), and *Goldman* v. *United States*, 316 U.S. 129 (1942).

85. The degree and speed of our present technological change is unprecedented. For example, in 1940 the Science section of the National Resources Committee could state, "From the early origins of an invention until its social effects the time interval averages about 30 years." (Green, "The New Technological Era: A View from the Law," *Bulletin of Atomic Scientists* 11, Nov. 1967). In the same article Green concludes, "But we know today that technological advance is so rapid that we do not have this 30 year interval. The shorter interval between the origins of a technology and the time its social effects are felt means that the practice of technology with intrinsic hazards will result in more injury to more people at an earlier time. Lawyers have recognized that during this time lag adequate justice may not be done on behalf of persons injured by the new technology. . . ." Ibid. at p. 14.

86. Green, "Technology Assessment and the Law: Introduction and Perspective," *G. Wash. L. Rev.* 36:1033 (1968), p. 1037.

> . . . modern technology produces insidious forms of hazards which do not produce instant dramatic injury but which operate slowly and cumulatively, so that the existence of hazards and concomitant injuries may not become manifest for many years or perhaps a generation.

87. Ibid. Green posits that products of government, industrial partnership are reviewed by the courts within the framework of government criteria.

> When the government through the legislative and the executive controls a technology (both by developing and policing it) the courts are inclined to constrain liability within the range specified by the government . . . .

What this means basically, is that in these technological areas effective protection of the public against the burdens and risks of the technology passes from the courts to the executive and perhaps the legislative branches. If the public interest is to be protected at all, it must be in the formulation of administrative standards or of legislation. See, *Casner* v. *Hecla Mining Co.*, 431 p.2d 792 (Utah, 1967). Despite uncontroverted evidence that any exposure to radiation may cause harm, the court refused to find a causal connection between the radiation exposure and alleged radiation injury where the exposure was within limits established by government authority.

88. See, "Technology and Privacy," Appendix Five to the Report of Privacy Protection Study Commission (1977), pp. 28-30.

In this regard it is interesting to note that the Congress' research arm, the Office of Technology Assessment, currently has underway a major study of three information technologies: electronic funds transfer systems; the electronic mail; and the computerized criminal history exchange program.

89. Thus far, Canada, France, West Germany, Sweden and Norway have enacted legislation. Austria, Belgium, Denmark, the Netherlands and Spain have legislation actively under consideration. Only Spain would not create a privacy regulatory body. See, *Transborder Data Report: The International Report on Information Politics and Regulation* (published monthly) and, in particular, the May 1978 issue.

II Science, Technology, and Socioeconomic Goals

1 The Economics of Productivity: Some Options for Improvement

*by Solomon Fabricant**

SUMMARY

Productivity growth is critical to the standard of living and to the international position of the U.S. Because of the recent decline in the productivity growth rate, our productivity performance is a national concern. Science and technology are important factors in this performance, but there are other factors, also important, that operate in our market economy.

The primary role in raising productivity belongs to private enterprise, which has the specific knowledge of the endless variety of incessantly changing conditions required to deal with the factors relating to productivity in a large number of variegated industries. This fact suggests that selection of governmental options for the improvement of productivity should concentrate on *general* policies that can serve to stimulate productivity in *all* industries.

These are the policies that help: expand the capital equipment of the country, by encouraging saving and investment; improve the quality of labor, by encouraging education and training; and raise the efficiency with which labor and capital are used in production, by encouraging scientific research and the development of new and better technology, materials, business organization, and labor-management relations, among other things. Science and technology are important in this list, but not all-important.

These policies can also be put in terms of removing or reducing obstacles to productivity growth—the taxes, expenditures and regulations of government, as well as the private monopolistic practices, that curb economic freedom or dull incentives. Barriers to price competition, restrictions on entry, and so on, may be intended to serve desirable national objectives other than productivity. But carried too far, their cost in terms of productivity may be higher than is socially desirable.

While the major emphasis of governmental policy should be on general factors, detailed government attention may be required, not only in its own operations, but also in industries—both service and goods producing—in which government is deeply involved as a major purchaser or supplier of funds. This is where some control must be exerted over production cost as well as quantity and quality of output. The problem of cost is usually less the result of backward technology than of inefficient organization, inadequate cost (and productivity) accounting, and weak incentives to economize.

Because too little is yet known about the economic, social and statistical facts and about the relations among the facts, more research on the questions involved should be stimulated and supported. And new policies should be initiated only on a modest experimental scale.

Finally, because productivity is so important to the economic and social welfare of our people, closer and steadier attention should be paid to it. It would be helpful to have an agency charged with the responsibility to keep productivity clearly within the public view, to explain

* Professor Emeritus of Economics, New York University; and Director-at-Large, National Bureau of Economic Research, New York

its meaning and determinants, and to act as a clearinghouse in the continuing and wide-ranging task of improving productivity.

INTRODUCTION

The importance of our subject, productivity, rests on three facts. First, productivity growth is the primary source of a higher standard of living for our people and a vital element in maintaining the economic and political position of the United States among nations. Second, the rate of productivity growth has, unfortunately, fallen sharply in recent years. Third, recent productivity growth in the U.S. has been slow compared to corresponding rates of other industrialized nations; thus there may be implications for our future international economic competitiveness in world markets. No one concerned with the future well-being of our country, therefore, can afford to ignore the possibilities for improving our productivity performance.

Productivity is also a complex subject. The process by which productivity grows is intricate, and the obstacles that stand in the way of faster productivity growth are not always evident even to the trained eye. Further, the information required to improve productivity suffers from gaps and ambiguities. This is especially true in the service sector, which according to available statistics, has both a lower level and rate of growth than does the non-service sector. These gaps and ambiguities may not be ignored when deciding what needs to be done.

THE MEANING OF PRODUCTIVITY

One source of popular confusion concerns the meaning of productivity, of which there are various definitions.

What is generally reported in the official productivity statistics is the change in the ratio of the aggregate number of units of physical output to the aggregate number of worker-hours employed in producing the output. The ratio, "output per worker-hour," is often referred to as "labor productivity."

To measure what many economists prefer to regard as productivity, account is taken not only of hours of work, but also of the services of the capital—tangible and intangible—used in production. Productivity is then defined as the ratio of output to the combination of worker-hours and capital input. To distinguish it from labor productivity, use is made of the term "total (or multi-) factor productivity." Obviously, these two productivity concepts do not mean the same thing. Nor do their measures move parallel to one another, because total input has generally grown more rapidly than the input of worker-hours alone; consequently, total factor productivity has grown less rapidly than labor productivity. Obviously, in addition, the two productivity concepts and measures are related. In fact, as we shall see later, growth of labor productivity may be viewed as determined by growth of total factor productivity together with growth of tangible and intangible capital per worker-hour.

Throughout our discussion, by productivity we shall mean, as does the Bureau of Labor Statistics (BLS), output per worker-hour. When reference is made to total factor productivity, that term will be used.

THE IMPORTANCE OF COSTS

Productivity in any one industry or group of industries, in the sense of output per worker, can always be improved. More or better tools and equipment may be used; or more and better materials and supplies may be obtained from other industries; or investment in new technology or in learning to use old technology may be made; or the industry may be reorganized; or relations within firms may be altered in various ways; or governmental regulations—aimed at reducing environmental pollution, for example—that restrict operations may be modified. Productivity improvement, then, will necessarily entail economic and social costs of one sort or another. Inherent in the question of options to improve productivity, therefore, is the question of how the benefit to be derived from an increase in labor productivity compares with the costs of attaining the increase in productivity. Beyond a certain point, the rate of return so calculated may prove to be unprofitable or even negative from the viewpoint of the firm or industry or of society as a whole. We shall have to keep costs in mind as we proceed in our discussion, recognizing that the information on costs—which may be indirect as well as direct, and long-term as well as short-term—may be scarce, especially when policy requires movement along untrodden paths, as it often does.

INTERDEPENDENCE AMONG INDUSTRIES

What happens or could happen to productivity in one sector depends on what is happening or could happen in the rest of the economy. Thus, to give an obvious example, the level and character of the demand for tourist services—and, therefore, the scale of operations that affects the level of productivity in the hotels, eating places, and amusement facilities that tourists use—depend on the quantity, quality, and costs of transportation goods and services. A proposal to improve productivity in a portion of the economy, then, may not always be evaluated, or even defined adequately, without some reference to policies regarding complementary or substitutional industries.

THE ESTIMATION OF PRODUCTIVITY

The measurement of output per worker-hour is not a simple matter. Because we must depend on available estimates, it is worthwhile to indicate some of the issues involved in making them.

CLASSIFICATION OF OUTPUT BY INDUSTRY

The nation's aggregate product can be classified by the "end" or "final" product consumed or invested in tangible capital, distinguishing particular services (those of a physician, for example) from commodities (such as pharmaceuticals). Or, GNP can be classified by the kind of industry engaged in producing GNP, distinguishing the service industries, such as retail and wholesale trade, from the commodity-producing industries, of which manufacturing is the standard example. Each classification has its uses and advantages.

When productivity is the main concern and a choice must be made, the industry classification is preferred. Employment and the capital and other resources used in production are classified in the basic statistics by industry, not by end product. It is possible to estimate the labor and other resources that, at each stage of production, add value to the materials that finally appear in the form of an end product. For some purposes it is necessary to do so, as in estimating the fraction of the food dollar going to the farmer. But to get at the reasons why this fraction is large or small, it is still necessary to study productivity, the conditions of production, and the possibilities of improving those conditions. This is best done on an industry basis. However, in view of the interrelations among industries that were noted above, and also between consumers and producers, the end product viewpoint provides a helpful supplement to the industry viewpoint, and it should be kept in mind.

THE DISTINCTION BETWEEN GOODS AND SERVICES

In discussions of problems of productivity growth, "the service sector" is often thought to be of special concern. Not only do available statistics show that both the level and the rate of growth of productivity in the service sector are substantially below the corresponding level and rate in the non-service sector, but the service sector has been growing faster than the rest of the economy and now employs two-thirds of all our workers.

According to the conventional classification, the service sector includes wholesale and retail trade; finance, insurance and real estate; a miscellaneous group of "other" service industries including "private households" (mainly hired domestic service); and government. Goods producing industries include agriculture, mining, construction, manufacturing, transportation, communication, and utilities.

However, the distinction between the goods and services sectors of the economy is not as sharp as might be presumed. "It is not evident," as one analyst has pointed out, "that a firm assembling purchased parts creates material commodities in a manner different from a restaurant preparing and serving food, although the Census calls the former establishment manufacturing and the latter trade." And one can sympathize with the dilemmas confronting statisticians in the Bureau of the Census; they classified the Hollywood motion picture industry as manufacturing up through 1929, but in later years placed it in the service sector.

Further, many goods-producing firms provide some of the distributive, repair, and other services needed in their own operations and needed also by their customers—sometimes even their workers and their suppliers. The possibility of change in the distribution of functions among industries must be kept in mind when considering options for improving productivity. Involved also is the changing relationship between the family economy and the market economy. The shift from the housewife's washtub to the commercial laundry and then to the automatic washer-dryer at home is only one example.

DIVERSITY OF INDUSTRIES

Industries vary substantially; the goods-producing industries, agriculture, construction, and manufacturing industries, for instance, obviously differ considerably in the character of the business they do. The service industries are perhaps even more mixed a bag. They vary widely in size, technology, rate of growth, and other characteristics. Some service industries have experienced absolute declines in employment (e.g., personal and repair services); others have had high rates of increased employment (e.g., doctors' and dentists' services).

It follows from the fact that the economy comprises a very large number of different industries that we cannot consider policies tailored to fit closely the specific characteristics of individual industries. Our discussion must be in broader terms.

SECTORAL DIFFERENCES IN MEASURED PRODUCTIVITY LEVELS

The contribution to the GNP from an industry, in relation to the number of employees or the number of hours worked by them, measures the industry's level of productivity at any given time; the major exception is general government, the GNP estimates of which are seriously defective for the present purpose. Excluding general government, then, GNP per worker in the service sector has been about $13 thousand per worker in recent years; in goods production it has been about $17 thousand. So measured, the level of productivity is lower by some 25 percent in the service sector than in the non-service sector. Given the more rapid rise in service sector employment, we may expect to find a "shift effect": a disproportionately high increase in employment in a low productivity sector would cause the rate of growth of national productivity to rise less rapidly than the average of the separate rates in services and non-service, even if these two rates were identical.

Such an effect can be detected in estimates made by the BLS. (These do not quite conform to our definition of the service sector, but they must do.) However, the shift effect has been modest—in the past decade, of the order of 0.1 percentage points; and probably somewhat more, but not over 0.2 percentage points during the preceding decade or two. These effects are to be judged by comparison with annual rates of change in national productivity of the order of 1.0 to 3.0 percent per annum. That is, a rate of growth of national productivity reported to be 2.0 percent a year, for example, would have been 2.1 or at most 2.2, had it not been for the shift effect.

INDUSTRIAL DIFFERENCES IN PRODUCTIVITY TRENDS

Table 1 shows measured productivity growth rates for the major industrial groups and for three different time periods.

The variation in growth rates, among industries, even at the high level of aggregation shown, is considerable. So is the variation among the periods distinguished. The productivity growth rate for four industries increased significantly from the first period to the second, and for two industries, from the second to the third period shown. For *most* industries in both the total service sector and the goods-producing sector, and thus for the economy as a whole, measured productivity growth declined from the first to the second period, and then again from the second to the third period.

Table 1. Productivity growth in the service and non-service sectors. 1950–1977.

	Output share 1977 (percent)	Average 1950 to 1965	Percentage 1965 to 1973	Change per year 1973 to 1977
Service sector				
Trade				
Wholesale	7.3	2.6	3.4	−.8
Retail	10.0	2.3	2.1	.8
Finance, insurance, and real estate	15.4	1.6	.2	2.3
Services, other	12.0	1.2	1.7	−.3
Government	12.5	.4	.5	.1
Total service sector	57.2	1.5	1.3	.6
Non-service sector				
Agriculture	2.9	4.9	3.6	3.0
Mining	1.5	4.3	1.9	−6.1
Construction	4.3	3.4	−2.1	.3
Manufacturing				
Nondurable	9.9	3.2	3.3	2.2
Durable	14.4	2.5	2.2	1.2
Transportation	3.9	3.0	2.9	1.0
Communication	3.2	5.3	4.6	6.7
Utilities	2.3	6.1	3.5	.2
Total non-service sector	42.4	3.4	2.4	1.5
All industries				
Fixed weights (1977 output)	100.0*	2.6	1.9	1.1
Current weights		2.7	2.0	1.1

Note: The productivity growth data relate to output per hour worked by all persons.

* Detail does not add to 100 percent because of rounding.

Source: Department of Commerce (Bureau of Economic Analysis) and Council of Economic Advisers, as published in the *Economic Report of the President*, January 1979, Table 16. The rates for the service and non-service totals are by our calculations, using 1977 output weights.

PROBLEMS OF OUTPUT MEASUREMENT

Serious questions, which we must not ignore, arise about the quality of the reported statistics. For a variety of reasons, output measurement in many industries is not straightforward; their productivity statistics may therefore be flawed. In agriculture a bushel of wheat is a bushel of wheat. In construction, however, the measurement of output in standard units is not nearly as simple.

A major problem in output measurement is the treatment of quality change. If improved products are associated with higher costs to the producer, current productivity measurements capture the resultant increase in output to the extent they are reflected in cost. But many product improvements resulting from science and technology are not associated with higher costs to the producer, and consequently are not currently reflected in output measurements.

Further, output measures do not presently capture a number of dimensions of social value, such as cleaner air and water and safer workplaces, although firms use resources to obtain them.

In the service sector especially, but in other industries as well, the very definition of a unit of output is a major stumbling block that has severe consequences for productivity measurement. For example, the definition, measurement quite aside, of the output of banks is debatable. Banks serve as depositories, make loans and investments, gather and dispense information, clear checks, do some bookkeeping for their depositors, provide financial advice and service in their capacity of trustee, and so on. To further complicate the problem, a good deal of this varied service is—or was (the arrangements are fluid)—paid for not at a stated price, but indirectly, as in the form of compensating balances. It makes a difference which of these services or combination of them is taken as the product, how changes in the volume of each are measured, and with what weights these volumes are combined. How these decisions are made, and therefore how these outputs appear in the national accounts, determines the measure of productivity.

Even more difficult problems arise in the field of education. No one has yet come up with a generally acceptable definition of the output of schools—or of the entire educational process, including the input of the family and the contribution of its environment as well as of the school. In discussions of the question, "What is the output of the school," some have seized on a "practical" definition and measurement of a school's output such as the performance of its pupils on a standardized test. Such a decision is liable to affect the behavior of teachers, who may be led to "teach to the test." The tendency to "teach to the test" is apparent in any industry, goods as well as service producing, whenever consumers judge the quality of the output by some fashionable criterion. But it is especially dangerous in the service industries, where objective criteria of quality are relatively scarce and the pressure to "show progress" is especially strong.

FORCES MAKING FOR PRODUCTIVITY GROWTH

Before considering options for improving productivity, it is essential to be as clear as we can be about the forces that encourage or inhibit it. Quantitative analyses of the sources of productivity growth have commonly focused on the factors that directly or immediately impinge on productivity.

FACTORS AFFECTING PRODUCTIVITY GROWTH

The sources of upward trends in output per worker-hour may be classified in various ways. Here we consider three main groups. One is increase in tangible capital per worker-hour. Another is improvement in the quality of labor. A third is increase in the efficiency with which capital and labor are put to use in production.

For those concerned with productivity, tangible capital is the first factor to come to mind. Plant and equipment have even been called "the" means of production. Whatever the merit of this designation, the more of these needs—and of other forms of tangible capital and the higher the quality of these means—supplied each worker, the higher tends to be the value of the product obtained from an hour of work.

But tangible capital is not the only means of assisting labor in production. Improvements in the quality of labor, the second group of productivity growth factors, also serve to increase output per worker-hour and also, in large part, represent a form of investment, but in human beings—made through education and training. This type of investment is not usually recognized as capital formation, and the families and governments that make expenditures for these purposes have objectives other than enlarging the productive and earning capacity of labor. But they have that purpose also, and whatever the purpose, the expenditures have that effect, on the whole. In this way, "human capital," as economists call it, is built up.

A third group of factors, those which contribute to the more efficient use of labor and capital, is measured by the "total (or multi-) factor productivity" to which reference was made earlier. Most prominent among these factors are advances in science and technology and the speedier diffusion of such advances through improved transportation, communication, and education.

In addition to science and technology there are additional factors contributing to "total factor productivity": the finer division of labor and greater degree of specialization of machines and processes made possible by the general increase in the size of business establishments, industries, and countries, as well as the improvements in transportation and communication and reductions in the obstacles to trade; the improved efficiency associated with reduction in the length of the work day and week; and a better allocation of resources.

It should be understood that the various sources of increase in labor productivity are interrelated; their separate contributions are, therefore, not easy to distinguish. Thus, doubling the rate of increase in tangible capital, for example, where the rates of increase in other inputs remain constant, would not double tangible capital's contribution to productivity, because of diminishing returns as the proportion of capital to the other inputs increased. To some extent, also, inputs complement one another. Technological change, to a significant degree, must be embodied in tangible capital if it is to be put to use in production. The contribution of technology, then, depends in part on the rate of increase in tangible capital as well as on its own rate of increase.

Because so much attention is often paid to technology as a primary source of productivity growth, three important facts brought out in studies of productivity should be underscored. The first lesson is that technology is only part of the larger stock of knowledge to which organizational, managerial, and social innovations contribute. Second, while advances in knowledge, broadly defined, constitute perhaps the most important source of growth in the efficiency with which labor and capital are used in production, there are other factors, not of negligible importance, that also contribute. And third, growth in efficiency is—and has been—joined with growth of tangible capital per worker and improved labor quality to increase output per hour of work.

THE RATE OF R&D AND TECHNICAL CHANGE IN PRODUCTIVITY GROWTH

Information on private sector R&D expenditures, classified by industry and sector, indicates that over 90 percent of research and development is performed in the manufacturing industries. Non-manufacturing industries account for only a minuscule share. However, much R&D reaches these industries embodied in the equipment, materials, and supplies they purchase from manufacturers and utilize in their operations.

In recent years, economists have attempted to assess in a number of different ways the contribution of research and development to output and productivity growth. Results, though gross approximations, provide general support for a strong positive contribution.

One approach attributes a large fraction of unexplained increases in output or output per worker-hour for the economy as a whole to advances in knowledge, which include factors other than R&D. A notable investigation of long-term U.S. economic growth has suggested that something close to half the growth in output per worker-hour in the U.S. between 1948 and 1969 may have been derived from advances in knowledge.

Another approach focuses on the association of productivity and output growth with research and development spending among individual industries. In a recent study (of select U.S. industries) the association turns out to be strong. Industries that have high ratios of research and development to sales (e.g., chemical, electrical, machinery) have experienced substantially higher growth of productivity and output than have industries (e.g., textiles) with low ratios. Other empirical studies (of chemicals, petroleum, and agriculture) in which the effects of increased input of labor and capital were *netted* out have found substantial *increases* in output associated with expenditures for research and development. Specifically increases of some $30 to $50 in annual output were associated with a $100 investment in research and development.

A third type of study has attempted to assess the economic returns for a sample of modest technological innovations. These suggest that for manufacturing during the 1960s, the average productivity returns to research and development conducted by firms in the producing industry amounted to about 30 percent per year; the benefits accruing to firms purchasing from the producers of technological innovations were about 50 percent.

These measurements of the contribution of research and development and technological innovation to economic progress are admittedly rough. One difficulty in obtaining more precise estimates is the inability to measure accurately the influence of new or improved technology on the quality of the goods and services produced. However, there is good reason to believe that if qualitative changes were covered, the measured returns would be higher since a large proportion of R&D activity goes toward improving the quality of goods and services. Second is the difficulty of tracing the contribution of research to improved technology and then tracing technology's contribution to economic progress. Bringing out new or improved products and production processes involves a number of complex interrelated tasks. Third, we lack the information and techniques required to measure, assess, and properly adjust productivity figures for changes in the toxic level and volume of harmful by-products of technological change.

OBSTACLES TO PRODUCTIVITY GROWTH

How strongly the forces of growth operate depends on the obstacles that impede these forces. We want to pay particular attention to those obstacles that are amenable to public policy. For our present purpose, it is convenient to group them under three heads: the "taste" for objectives other than productivity growth, restrictions on the freedom to pursue the objective of productivity growth, and limitations that dull the inducement to pursue this objective.

COMPETING OBJECTIVES

The nation has, we all know, other goals besides that of expanding output per worker-hour in order to increase the real income or leisure time of our citizens. A more even distribution of income and employment opportunities and a more stable, as well as higher, level of income and employment are also economic goals. And there are the so-called non-economic goals, not the least of which are a satisfactory physical and social environment. national defense, and political freedom. Productivity growth, then, is not the only or the overriding national objective.

Apart from national goals, there are the goals of individual workers that tend to compete with productivity advance. The opposition to automation and other technological changes that disturb and threaten to undermine the security of workers in their jobs is an ancient story. More recently there are the so-called deterioration of workers' attitudes toward their jobs and the emphasis placed on the quality of working life, to which reference has often been made in explaining the national productivity slowdown of the past decade. These developments may also be viewed as a shift toward objectives that compete with the drive for real income that rising productivity can bring.

In considering the issues, it is important to understand that investment in competing goals can have a favorable feedback on labor quality and thus serve to offset some of the retarding effects on productivity caused by a slower rate of increase in what is narrowly viewed as "productive" plant and equipment. Further, solution of pollution problems can and often does mean more effective use of the raw materials and waste end-products that now are the pollutants. Investment in the reduction of pollution can be profitable—or less costly than commonly supposed—to business as well as society at large: productivity need not necessarily be reduced or reduced significantly. Improvement in the quality of the environment and reduction in illness and accidents yield benefits that are not now included in our conventional measures of national output. If these benefits were covered, output per worker-hour—not now, but later, after a lag—might tell a different story. Finally, economic progress as measured by national output or productivity is not synonymous with improvement in the quality of life; the value of tradeoffs among goals to improve the quality of life is a social judgment.

CURBS ON ECONOMIC FREEDOM

Monopolistic practices in private industry restrict freedom of entry, hamper price competition, and impede adaptation of industry to technological and other change. The rather tortuous moves underway to revise segments of the financial system and specify their relationships to one another and to the markets they serve provide an example of these limitations and of the difficulties of revising them.

The governmental regulations involved always have, or have been said to have, virtuous objectives. The licensing requirement to which most professionals and many occupations are subject, for example, are supposed to assist in maintaining or raising the level of competence in these fields. But another result, even if unintended, is also to restrict numbers. Another example is provided by building codes designed to protect the building occupants and the community. These codes tend to become obsolete and are not always consistent with one another, thereby raising the costs of home construction.

Within government, various impediments to flexibility, adaptation, and economic efficiency are built into the system by statutory, administrative and even constitutional restrictions. The obstacles are not insuperable, but the task of overcoming the impediments to change is long and arduous.

DULLED INCENTIVES

The incentives to work, to save and invest in tangible and human capital, and to seek higher efficiency in the use of labor and capital may be blunted by taxes and by the provisions written into the tax codes. High and progressive income taxes play a part, especially in the choice between time spent in earning taxable income and time spent at home in nontaxable labor or in leisure. Even the provisions of the unemployment compensation system enter into the calculations of workers and also of employers confronted with the choice between laying off workers or providing part-time work when adjustment to falling demand is necessary. Further, much attention has been paid to the "double taxation" of income from corporate capital and, more recently, to the "hidden" tax imposed by inflation on income from both equity and debt capital under present income-tax accounting rules. These taxes tend to depress the expected rate of return on capital and thus the rate of savings and investment necessary to build up capital per worker.

Another set of factors makes for increased uncertainty and risk, thereby tending to slow up the process of investment and the development of new and improved processes, products and services. One is the cartel-induced rise in the price of petroleum in 1973. Investment of funds and of time has had, and will have, to be devoted to adjusting to radically new relative costs. Businessmen must also cope with complicated energy policies. Another factor is the high and unsteady rate of inflation together with uncertain government response to it. In many industries inflation has complicated labor-mangement relations and aggravated the uncertainties that confront businessmen, besides raising the effective income tax rate and confusing people trying to make sense of conventional accounting statements when planning their investments. Also of concern are the fluctuations in the Federal Government's support of R&D, which could hardly have failed to lower the efficiency with which research and development is planned and carried out. How important these factors, and others that could be added, really are in dampening incentives to raise productivity is a matter of controversy. But it hardly seems likely that the effect has been negligible.

A related factor tending to dull incentives, especially to invest in R&D, is the fact that not all the benefits derived from such investments are captured by those making the investment. Sooner or later, even when kept under cover or protected by patents, knowledge and then use of the new methods, materials, and products can spread to competitors inside and outside the industry, with little or no compensation being paid to the initiating enterprise.

Finally, mention must be made of the institutional arrangements tending to dampen incentives to economize and adjust to changing conditions, that characterize industries such as hospitals, schools, and other nonprofit institutions supported in whole or in part by governmental funds—as well as governments themselves. It is a gross exaggeration to argue, as some seem to, that incentives to improve productivity are altogether lacking in these institutions. There is plenty of empirical evidence and good logic for believing that incentives operate and productivity has grown to varying degrees in the regulated and nonprofit institutions. However, it seems clear that incentives to improve productivity and cut costs can be sharpened further in these organizations.

LIMITATIONS ON PUBLIC POLICY

The options for improving productivity are already implied in the preceding discussion. Before we go further it would be helpful to spell out the reasons why the choice of public policy is—or should be—limited.

THE PRIMARY ROLE OF PRIVATE ENTERPRISE

Before we ask what government can do to improve productivity in any sector, we must consider the primary role of private enterprise.

By far, the greater part of productivity increase has been achieved through private initiatives taken for private benefit. In the future, also, we can expect productivity gains to depend primarily on the efforts of individuals and businesses. They must be given the incentives to increase their investments in training and education and in capital equipment. They must be encouraged to do the formal and informal research needed to develop technological, organizational, and managerial innovations and other means of raising efficiency.

The major contribution of government to productivity increase in the private sector has been in assisting and supplementing these efforts of individuals and enterprise. Government must continue to provide an environment in which our people can look to the future with some confidence and freely seek to attain their own ends. Government must provide or support the facilities, knowledge and information that are of general benefit to the public and that yield returns that cannot be fully captured by investors. Government must continue to make heavy investments in health and education, primarily to reach other national goals but also to raise national productivity.

LIMITED KNOWLEDGE

In pursuing the objective of improved productivity, it is essential to recognize and admit the unpleasant fact that our understanding of productivity is severely limited. In many service industries, even the simple facts about the level and rate of change of productivity—in government, in particular, but also in other service industries—are anything but plain. It is not too much to say that the Department of Commerce only pretends to be measuring the output of general government. Nor do we have a firm grasp of productivity levels and change in a considerable fraction of the non-service or commodity-producing industries. The experts know less than the public believes they know about productivity.

In part, this is due to the difficulties of defining and measuring output. What we mean by more and better output—or a unit of output—is clouded by ambiguity. Every industry's output—indeed every individual product—is a multidimensional bundle of attributes, not all of which are easily, if at all, measurable. And the concept and measure are clouded further by general lack of agreement on the relative values or weights to be assigned to movement along each dimension.

But even if we accept conventional definitions and measurements of output and input, we know little that is specific and quantitative about the factors that determine the relation between them. It should be sufficient here to recall the efforts by economists and public officials to explain the slowdown in national productivity growth recorded in the statistics since 1973, or even earlier, and to recall how different those explanations were. The description in the 1979 *Economic Report of the President* by the Council of Economic Advisers of the Council's attempt to determine and then explain the trend of national productivity in recent years includes a frank confession of ignorance. The suggestions made above about the factors involved in the slowdown would not find universal agreement. The factors are subjects for research.

THE DIRECTION OF PUBLIC POLICY

GENERAL PRINCIPLES

1. Sound governmental policy for improving productivity should, as a rule, be in terms of enhancing the general environment within which *all* industries operate. To improve productivity in industry at large is the best way to improve productivity in specific industries, because it takes account, as it should, of the number, size, and variety of industries, and of the limitations on the scale and flexibility of governmental bureaus and knowledge of government officials.

2. To enhance the *general* environment within which all industries operate means to foster an environment within which the private decisions of savers and investors, producers and consumers, workers and employers, wherever they are, can be made more freely, more confidently, and more rationally than they can now be made.

It follows, also, that pursuit of a program to raise productivity must proceed on a wide front: there is no dominant source of productivity growth to be singled out.

3. After what has been argued here, the reader will not expect a list of specific options; nor would it be possible to give one.

Consider, for example, the fact that a major source of productivity growth is increased capital investment. Any policy to reduce the risk and uncertainty that confronts investors, and to increase the rate of return they may expect, is, then, an option for improving productivity. But obviously it is not so much *an* option as a large family of options. There are many ways to stimulate capital formation. Each way would surely entail direct costs, and have unintended effects, which would need to be weighed. What, to be more specific, can be done to deal with the depressant effect of the Social Security system, as it is now organized and operated, on the rate of national saving without destroying the ability of the Social Security system to meet the objective for which it was established?

It would be foolish, given the limitations of the present document, to try to go into any detail on this sort of question. Besides policies to encourage an increase in the volume of capital formation, consideration would have to be given to policies to strengthen the process of training and education, to support and encourage the development of new basic knowledge and technologies and the diffusion and adaptation of the old and the new, to assist in improving business (and nonprofit) organization and management, to remove obstacles to competition, and to lessen discrimination against minorities. The list of possible policies to improve productivity is long. And many also are the forms that implementation could take in the tax codes, regulations, and government support of research, statistics and education.

4. Another implication concerns the choice among the implements of policy. The implements should be in the form of general rules and regulations, rather than tailored to the details of narrow situations.

In the case of regulation, for example, where highly specific "do's and don'ts" now fill the *Federal Register*, a policy shift toward more general rules, and penalties or rewards—for example, in the case of automobile fuel efficiency—would reduce the governmental burden. Producers and consumers would be encouraged to seek out the most economical means appropriate to their particular technical, economic, and environmental circumstances—circumstances which they are in the best position to know.

5. Because too little is yet known about the connection between any given policy and its results, it is desirable that policy be kept flexible. At the same time, however, shifts in policy should not be so erratic as to create uncertainty and thereby discourage private plans and actions for raising productivity.

Limitations on our knowledge also suggest the importance of research—not only on technology as it is usually defined, but also on organizational, managerial, and motivational questions and on other economic, social, and political problems. Suggested, as well, is the importance of experimenting with new ideas on a modest scale. This requires holding open the possibility of retreat, which, in turn, requires that commitments to keep unsuccessful experiments alive be avoided. It follows that every experiment should provide for prompt and reliable feedback of information on the costs and benefits of the experimental policy—information to be weighed not only by those who may have developed a vested interest in the program, but also by persons who can be more objective.

6. Finally, because productivity is so important to the well-being of our people, steadier and closer attention should be paid to it—not only by the Federal Government, but also by the state and local governments, and not only by government, but also by business, labor, and the general public. There is need for a permanent Federal agency charged with the responsibility and armed with the resources needed to keep the importance of productivity prominently in the public view and to act as a clearinghouse in improving it.

CONCENTRATION ON PARTICULAR INDUSTRIES

Whether there is any need for, or advantage in, concentrating attention on particular industries also requires some discussion.

1. If it is believed that policy should concentrate on those situations in which productivity is low or rising slowly, it should be recognized that such situations can be found in many industries, both service and goods producing.

However, it should also be understood that a low productivity level or rate of growth does not necessarily identify a situation in which the *potential* for productivity improvement is greater than elsewhere. The relatively low level or rate may simply reflect, particularly in service industries, a downward bias in measurement, a possibility that must always be considered. But even if the productivity measure turns out to be sound, it can only hint at the desirability of improvement. The cost of improving productivity in the industry may be higher than it is worth: further study of the reasons for the low productivity and of the cost of dealing with them is indicated.

2. Some have questioned whether activities that have experienced relatively fast productivity growth would not be expected to yield a better return in response to governmental assistance than activities that have expe-

rienced relatively slow productivity growth. Industries with relatively high average rates of productivity growth in one period do tend to have relatively high rates in the next period. Whether they would be more likely than other industries to profit from the assistance of government simply because of their superior historical record is a question, however—perhaps one on which Japanese experience in relation to export manufactures may provide some evidence.

3. Still another criterion of selection of industries for detailed governmental attention should be mentioned. It is, in fact, the one criterion already in use.

Government is deeply involved, not always of necessity, in the details of production in a number of industries. Besides general government and government enterprises, these are the private industries in which government is a major purchaser of goods made to its specifications or to which government funds provide the major source of finance. Some control is therefore exerted over quantity, quality, and cost of production.

That these government controls leave something to be desired is well known. What the reader should be reminded of, in a discussion of options to raise productivity, is this question: Is sufficient attention being given to the possibility that, in some cases of government operation or close involvement, control might be better exerted and productivity raised more effectively by subjecting the activities in question to the forces of competition in a free market?

The reader should be reminded also that when competition is not the solution, as is probably the case in most situations of close governmental involvement, preliminary work still needs to be done to improve matters. There is no ready stock of well-formulated, tested, and otherwise acceptable ideas on which to draw. This is illustrated by the ongoing discussions of the problem of defining and measuring cost of production—the solution of which is a necessary condition for effective cost control and productivity improvement.

The problem of measuring cost is different in government because here even the rudiments of a modern accounting system are lacking. This is one reason why the productivity, output and input data for government are so inadequate. More important, it is one reason why governments pay less attention than they should to making efficient use of the great stock of tangible capital at their disposal. To improve the situation would require a radical reformation of the bookkeeping and budgetary practices of government (some of which are set by legal requirements) along lines on which agreement among accountants has yet to be reached.

2 Technology and the Improvement of Agricultural Productivity

*by Harold D. Guither**

SUMMARY

Agriculture is the nation's largest industry, employing from 17 to 20 million persons. Together with direct support industries, it accounts for about 25 percent of the gross national product.

Agricultural production represents about 2.5 percent of the total national energy requirement. Approximately 17 percent of all U.S. energy requirement is related to the food system. About half of the fuel consumed is petroleum, and another 30 percent is natural gas.

The U.S. farmer is a middleman processing purchased inputs, such as machines, minerals, fertilizers, and fossil fuels, into food and fiber. If the supply of purchased inputs were interrupted, the industry would reach a standstill virtually overnight. Prices of industrial inputs are expected to increase relative to farm product prices and are one of the factors that affect productivity.

The public investment in human capital—developing the minds and capability of the farm and rural population and sharpening management skills—has helped enable the nation to increase its agricultural productivity.

The growth in agricultural productivity is essentially a process of adaptation by the agricultural sector to new opportunities created by the advances in knowledge and by the interindustry division of labor that has accompanied industrialization.

Options for improving productivity of capital are closely allied with energy use. These include (1) making more efficient use of present technologies, (2) conserving fossil fuels by shifting to more solar energy use, (3) increasing alcohol production from farm-produced materials, and (4) shifting to less capital-intensive methods.

The use of insecticides, herbicides, and other chemicals can often be reduced without reducing their effectiveness. Integrated pest management offers real potential for saving energy and improving productivity. Machinery in farming operations could be used more efficiently.

Solar energy is a renewable resource that can be expected to provide a greater portion of the future energy requirements of agriculture. Production of alcohol from agricultural sources is technically feasible, but research and evaluation is needed before definitive conclusions or recommendations can be made.

Organic farming is a popular idea, but research into its energy-saving potential is inconclusive. Returning to a farming and living style like that of the Amish was studied, but energy conservation achievements were found to be greater in Amish homes than on Amish farms.

Large substitutions of resources like labor for mechanized energy are not likely, but increased management ability and time may be substituted for purchased energy. Areas that require additional management input include integrated pest management, fuel-saving meth-

* Professor of Agricultural Economics, University of Illinois at Urbana-Champaign.

ods, and application of fertilizer and pesticides and other cultural practices that reduce machinery use.

Restrictions on farm size to permit more people to live on farms could lead to greater numbers of small farms and farmers surrounded by poverty conditions.

Tax policies have an effect on investment decisions and can influence the trend to larger farms. Regulations dealing with environmental controls add to production costs and reduce yields.

No major scientific breakthroughs in agriculture can be predicted for the next one or two decades, but promising potentials remain for improving productivity by application of known technology.

Public sector research and development remains the primary source of agricultural technology, particularly the biological technology that has expanded the capacity of crops and animals to respond to higher levels of industrial inputs.

Areas of research that might be productively pursued include: climatic processes, integrated pest management systems, basic plant sciences and efficient methods of crop fertilization, control of livestock diseases, improvement of genetically superior animals and reproductive performance, development of alternative energy sources, and improvement of land use.

To facilitate a shift away from fossil fuels, a gradual rise in price can be an effective stimulus. Shielding agriculture from increases in energy prices may have unfavorable long-run consequences.

AGRICULTURE IN THE NATION'S ECONOMY

U.S. agriculture in 1979 will produce a gross output valued between $118 and $128 billion. To achieve this production, farm owners and operators will incur production expenses of $104 to $109 billion, much of which will represent goods and services purchased from U.S. business and industry.[1]

INDUSTRY SIGNIFICANCE

Agriculture is the largest industry in this country, employing some 17 to 20 million American workers. About 4.3 million persons work directly in farm production. About 8 to 10 million are involved in storing, transporting, processing, and merchandising farm commodities. That accounts for one out of every five jobs in private industry. In 1975 nearly 3.3 million persons were employed by eating and drinking places. The agricultural food system thus employs more workers than any other industry, either directly or indirectly. These figures do not include the related jobs in manufacturing agricultural equipment and equipment components.

Agricultural production itself accounts for only 3.5 percent of the total U.S. gross national product. When industries producing inputs or processing and distributing agricultural products are included, however, this rises to approximately 25 percent of the total gross national product. Agricultural assets totaled $671 billion in 1977—equal to 75 percent of the capital assets of all manufacturing corporations in the United States.

In recent years international trade in agriculture products has made an increasing contribution to the economic well-being of the U.S. agricultural economy. In the calendar year 1978 agricultural exports totaled $29.4 billion; agricultural imports totaled $14.6 billion, in contrast to a nonagricultural trade deficit of $46.8 billion. Because of a positive agricultural trade balance, the net deficit in the U.S. trade balance was reduced to $32.2 billion. The increased role of agriculture in foreign trade adds to the uncertainties with which farmers must contend and which the nation must recognize.[2]

ENERGY INPUTS USED IN AGRICULTURE

Total energy use for agricultural production, approximately 2.5 percent of the national total, includes energy used in direct farm production, for indirect inputs, and for capital inputs. It excludes transportation of both production inputs to the farm and produce from the farm to market.[3]

Indirect energy use occurs in the manufacture of fertilizer, pesticides, feed, and feed ingredients. Capital inputs also include the manufacture of farm machinery and equipment, agricultural steel, and farm trucks and trailers and the construction of farm buildings.

Approximately 17 percent of all U.S. energy requirements are related to the food system.[4] Approximately half of the fuel consumed in the food system is petroleum, and another 30 percent is natural gas. For many purposes natural gas is the most desirable and the least expensive alternative. Yet its users are faced with an uncertain supply-price outlook.

Gasoline and diesel fuels for tractors, combines, and other farm machinery constitute four-fifths of the direct energy requirements for crop and livestock production.[5] In 1972 electricity accounted for about 11 percent and fossil fuels about 89 percent of farm energy expenditures.[6]

Energy has been utilized in order to facilitate four major objectives: (1) reducing labor requirements—for example, mechanization on the farm and convenience foods in the home; (2) increasing the quantity and variety of food produced while using less land; (3) lowering the risk of crop failure or food spoilage—for example, through irrigation, crop drying, and refrigeration; and (4) reducing waste while maintaining or creating a quality of food preferred by consumers, as evidenced by their willingness to pay the necessary higher costs.

American farmers spend 4 to 8 percent of their operating budgets on energy. Fertilizer production consumes large quantities of natural gas, and adequate amounts of chemical fertilizer are vital to the success of farmers' operations. One estimate is that one-third to one-half of agricultural yield is due to the use of chemical fertilizer.[7]

U.S. farmers have become dependent upon nonrenewable energy because it has paid to do so in the past.[8] However, the growing scarcity and rising cost of those materials point to a changing resource environment for agriculture.[9]

FACTORS IN AGRICULTURAL PRODUCTIVITY

Four factors enter into the agricultural production process: land, labor, capital, and management. Loomis emphasizes that one of the most dynamic characteristics of U.S. agriculture is the constantly changing input mix.[10]

In reality, the present day U.S. farmer is a middleman processing purchased inputs, such as machines, minerals, fertilizers, and fossil fuels, into food and fiber. If the supply of purchased inputs to U.S. farmers is interrupted, the industry will grind to a virtual standstill overnight.[11]

LAND

Land is the base on which all agricultural output begins. The most recent estimates of the nation's total land area, published in 1974, reported 1,897 million acres, of which 333 million were classified as cropland, 603 million as forest and woodland, 767 million as pasture and range, and 194 million as urban land and land for special uses. The amount of land needed for a productive, efficient farming operation will depend upon the available labor, capital, and management resources, as well as the types of commodities being produced.

LABOR

Labor demonstrates the outstanding change in the input mix. The movement of labor from, and the movement of labor-saving inputs and techniques into, agriculture characterize this shift. Affecting the supply of farm labor are the returns from employment in agriculture relative to those from employment in the nonagricultural sector. Available statistical evidence suggests that the difference between the returns from nonfarm and farm employment causes a pull on the labor force from agriculture to industry. However, the mere presence of an economic pull is only a necessary, not sufficient, condition for mobility. There must also be real opportunities for labor to transfer from agriculture into industry.

The second facet of the labor cost-input picture is that of the relative costs of labor and substitute inputs. The movement of labor out of agriculture and the movement into agriculture of such labor-saving inputs as machinery, motor vehicles, fertilizer, and improved feeds and seeds have been essentially simultaneous.[12]

Agricultural labor in the United States is performed by three distinct groups. Farm operators, numbering about 2.5 million in 1974, do most of the agricultural work. Farm operators' labor is supplemented by nearly one-half million unpaid family workers and an estimated 2.7 million hired farmworkers, some working only part of the year.

The hired farm work force, about 3 percent of the civilian labor force, is distributed unevenly across different types of crop farms and geographic areas.[13] The demands for labor may be intense during harvest periods and relatively light during parts of the growing season. Farm workers fully employed in the growing season may be underemployed or unemployed during other parts of the year. In livestock and dairy areas the needs for labor are distributed more evenly throughout the year than in crop-growing areas.

CAPITAL

Capital in agricultural production encompasses a wide range of goods and technologies that are combined with land and labor. There is a near consensus that industrial inputs, though not critically scarce, will not be as cheaply available in the future as in the past. The prices of these inputs may continue to increase relative to farm product prices, instead of lagging behind product prices, as in earlier decades.

The role of changing relative prices as one of the factors affecting productivity is illustrated by machinery and fertilizer. This changing price structure cannot be disassociated from technology, as it is through technology that new and improved inputs become available to agriculture at favorable prices.

Binswanger and Ruttan point out that a decline in the price of fertilizer relative to the price of land can be expected to induce a rise in fertilizer used per hectare and also induce advances in crop technology such as the development and introduction of more fertilizer-responsive crop varieties. As the price of labor rises in relation to the price of land, farmers can be expected to attempt to reduce labor input per unit of land by substituting fertilizer and other chemical inputs, such as herbicides and insecticides, for more labor-intensive practices.[14]

Because of the physical response of plant growth to fertilizer, the increased output per unit of fertilizer input is economically favorable on many farms. Also, the use of fertilizer results in a reduction in amount of labor and other inputs per unit of output through allocation of relatively fixed inputs to an increased volume of output.[15]

MANAGEMENT

Management constitutes a fourth production factor. Management refers to that often intangible talent or ability of the farm operator or decisionmaker to combine land, labor, and capital into the lowest-cost and most efficient mix of inputs to produce a given farm product.

Management is a major factor affecting productivity. Technology may be available and economic conditions may justify changes in inputs, but these forces become meaningful in a productivity sense only when the farm manager acts in response to them. Here, we are in the areas of education and decisionmaking as they influence changing productivity.[16]

The public investment in human capital—developing the minds and capabilities of the farm and rural population and thereby sharpening management skills—has helped enable the nation to increase its agricultural productivity. As Hayami and Ruttan point out, "Public support for education and research as an instrument of economic progress represents a major institutional innovation in modern society."[17]

Improvements in human capital in the form of educated, innovative farmers, competent scientists and technicians, and perceptive public administrators and business entrepreneurs are essential to generating improvements in agricultural productivity.[18]

TOTAL PRODUCTIVITY AND FACTOR USE

It is difficult to generalize about the productivity of American farms. Soil fertility, climate, weather conditions, and technology all contribute to the final output per unit of input.

The conventional approach to the measurement of total factor productivity involves the computation of an index of total output and an index of all factor inputs. Total factor productivity is then simply computed as the ratio of the output index to the input index.[19]

Increases in total factor productivity have been important sources of output expansion in the past. From 1940 to 1970 total measured physical inputs in agriculture increased only 4 percent while output increased 58 percent.[20]

There are many forces behind the increase in productivity of our total farm resources. Most important is technology—the ever-changing body of knowledge, techniques, and instruments used in producing economic goods.

The growth in agricultural productivity has been closely related to the progress of the entire economy. Throughout the economic development of the United States, increased productivity in the agricultural sector has released resources and facilitated industrialization. Industrialization, in turn, has fostered and facilitated technological developments adaptable to agriculture and thereby has stimulated increased productivity in agriculture. The two forces have complemented each other.[21]

Actually, the growth in agricultural productivity is essentially a process of adaptation by the agricultural sector to new opportunities created by the advances in knowledge and by the interindustry division of labor that has accompanied industrialization.[22]

Hayami and Ruttan also point out that a progressive industrial economy also contributes to growth of agricultural productivity through a *greater* capacity to support agricultural research; through its capacity to support both general education and production education in rural areas; through its capacity to support the development of more effective transportation and communications systems; and through the pervasive strengthening of other elements in the physical and institutional infrastructure serving rural areas.[23]

Technological changes affect the productivity of capital and labor. The observed differences in land and labor productivity are associated with differences in the level of biological and mechanical technology of various countries. Hayami and Ruttan believe that industrial inputs are responsible for advances in both biological and mechanical technology. Nations that have achieved relatively high levels of either land or labor productivity have been relatively successful at substituting industrial inputs for the relatively scarce resources—either land or labor.[24]

The percentage changes in productivity generally follow the percentage changes in farm size. The presumption is that technical change is the mechanism that induced both productivity increases and enlargement of farm size.[25]

Productive assets per farm worker and per farm have risen from $3,300 and $6,200 in 1940 to $97,601 and $164,885 in 1975. (Productive assets are defined as real estate, plus machinery, inventories, and cash on hand.) Even in real terms, these are sixfold increases. Further, man-hours in agricultural production have fallen with a concomitant rise in output per man-hour during the period. At the same time, units of machinery are decreasing, but their values and sizes are increasing. The trend toward increased capital-labor ratios and unit sizes of machinery is apparently linked to increased farm size through declining average production costs.[26]

A strong trend toward specialization has characterized the "modernization" of every industry, including agriculture. Among the consequences in agriculture are specialized and animal monocultures. Monocultures facilitate mechanization and automation; but they also require using a vast array of chemical and mechanical measures to cope with diseases and insect problems.[27]

The consequences of this massive infusion of technology have been widespread. Agricultural output has increased, but it has become concentrated in areas away from markets. The concentration of production in remote areas has required the picking of crops that have not yet ripened, their storage and increased shipping costs. This also sometimes has lowered the quality of final products and has aroused the ire of consumers. The extent of consumer dissatisfaction has never been measured, but it is certain that the technologies designed to capture economies of scale have taken their toll.[28]

However, agricultural productivity has benefited the American consumer in two ways. Abundant supplies of a wide variety of foods have resulted in lower prices compared to those in other developed countries.[29] The ability of a small proportion of the population to produce the total national food supply enables the remainder of the population to engage in production and distribution of other goods and services, contributing to a higher standard of living for the entire population.

ENERGY USE IN PERSPECTIVE

Until the mid-1950s, total agricultural yields increased more rapidly than energy use, but since then, the rate of increase in energy has been faster than the rate of increase in yields. Not only does a farm use energy directly to power machinery and in the form of fertilizers, for example, but energy has indirectly led to greater agricultural production by eliminating the necessity for woods formerly used to supply household fuel and for pastures required for draft animals.[30]

Energy is a serious problem that goes far beyond any one sector of the economy. The effects on agriculture will be determined in large part by the national policies that are adopted. Standby allocation schemes are in place for vital functions such as food production, but the high degree of economic interdependence diminishes their value, except in the short run. Having fuel available is of little use if energy for manufacturing and distributing tires, spare parts, and other inputs is not.[31]

PRESENT FACTOR USE AND PRODUCTIVITY: ASSESSMENT, ISSUES, AND EXPECTED CHANGES

LAND USE

Land use adjustments that affect ultimate output can be categorized into three types: crop specialization, shifts within the cropland base, and land use conversions. Between 1969 and 1975 the acreage planted to the seven major crops (corn, oats, barley, sorghum, soybeans, cotton, and wheat) accounted for nearly all of the increase in acreage of the 20 principal crops.

The acreage planted to all feed grains increased only modestly, but corn acreage increased at the expense of oats and barley, while sorghum remained a constant proportion of the total feed grain acreage. Corn is higher yielding than the crops it replaced and has greater feed-equivalent value. Consequently, the 9 percent increase in composite yields between 1969 and 1975 can be attributed to crop specialization. Because this shift in cropland intensity has, for the most part, concluded, further contributions to output from this source cannot be expected.[32]

A new appreciation of the crucial role of farmland has developed during the 1970s. The perceived scarcity and higher price of inputs such as fertilizer, pesticides, and herbicides contributed to skyrocketing land prices.[33]

Differences of opinion exist on the extent of the land base or potential land base, partly because measurement is imprecise, data are unavailable for some regions, and experts don't agree on definition of terms.[34] To some, "potential arable land" refers to productive capacity or ultimate limits of land development, leaving costs and benefits aside. To others, it is some unspecified point on the land development schedule, with given levels of technology and prices.

The United States has relied mostly on increased technology to meet expanding production needs. While U.S. crop output from 1940 to 1972 increased nearly 70 percent on essentially the same land area, farm labor inputs fell by two-thirds, fertilizer use increased nearly ninefold, and mechanical power and machinery inputs grew by 237 percent.[35]

Competition for land from such forces as urbanization and surface mining become policy issues because of the dramatic, and often irreversible changes they make. The amounts of land threatened each year may be small, but the issue is continually of policy interest.[36]

LABOR USE

The distinguishing feature of U.S. agricultural development over the period 1880-1970 is the primary reliance on growth in land area per worker as a source of growth in output per worker.[37]

In the developed countries, where demand for labor in the nonagricultural sector pulled a significant amount of labor out of agriculture, the rates of growth in labor productivity have been increased by improvements in the land-labor ratio.[38] The progress of mechanization has been a primary source of growth in output per worker.

Given the present costs of labor in the United States, any effort to substitute human or animal labor extensively for mechanical power would result in sharply higher food prices.[39]

If the 160-acre limitation on farms receiving Federal water were enforced, it could encourage a shift toward labor-intensive fruits and vegetables because large acreage capable of using capital-intensive equipment would no longer be profitable.[40]

Barkley cautions about shifting to less capital-intensive farming without considering the consequences. Although one of the reasons for choosing a different method is to relieve agriculture's dependence on capital by substituting labor for nonlabor farm inputs, this switch in factor proportions leads to slower-paced agriculture, increased opportunity to farm, and less nonfarm capital.

The result of the switch would be lowered land-labor and capital-labor ratios. This, in turn, would lead to more farms—and perhaps an increased market demand for land to be used for the traditional labor-intensive, single-family farm. This kind of technological transference is apparently the kind of change that many advocates of "more appropriate" technology have in mind. It has a romantic appeal that fits with the American myth of agrarian fundamentalism; it has contemporary appeal because it offers the opportunity to substitute labor for capital.

But Barkley is apprehensive about a shift to intermediate technology. It could solve rural problems or it could make them worse, depending upon whether the changes in agriculture production are labor-intensive and increase the rural community population or are labor-extensive and lead to rural unemployment and migration from the rural areas. Utilizing more human labor and animal power is not a viable alternative because there is insufficient cropland to produce feed for millions of draft animals that would be required to partly or completely replace tractors.[41]

CAPITAL AND ENERGY USE

The dependable production of food during the transition from fossil fuel sources to renewable energy sources poses a major challenge to U.S. agriculture.

For the present, efficient use of present technology has shown that productivity of capital can be increased.

The diesel engine has significantly higher fuel efficiency than the gasoline engine does, both on an energy basis and on a gallon basis. In one recent test diesel tractors averaged about 38 percent higher horsepower hours per gallon than comparable gasoline models. This fuel efficiency is also reflected in self-propelled combines.

The implications of these developments would encourage greater use of diesel engines in major farming operations. Such economic incentives, however, must be matched by energy policies that allocate increasing quantities of diesel fuels to farming communities as greater shifts to diesel fuel burning engines occur. Rigid allocation policies based on past usage will discourage a shift to diesel equipment or disrupt agricultural production when shortages take place.

Electrical energy applied to stationary farm routines has increased worker productivity and enhanced worker safety.[42]

The use of insecticides, herbicides, and other chemicals can often be reduced, without reducing their effectiveness. Farmers can carefully analyze the need for the fertilizers, especially through soil testing, and use them accordingly.

Insulation of farm buildings requiring climate control can keep out winter cold and summer heat, reducing heating and cooling costs.[43]

Tillage has been reduced over the past 10 or 15 years; however, increased herbicide requirements tend to nullify part of the savings, and minimum tillage can be more readily justified on the basis of soil and water conservation and reduced labor requirements than on the basis of energy savings.[44]

A 1974 estimate indicated minimum tillage practices are in effect on as many as 33 million acres. An assessment of the long-run effects of its continued adoption concluded that by the year 2000 annual harvested acreage could increase by 20 million acres and crop production could increase by 5 percent, mainly through increased multiple cropping made possible by minimum tillage.[45]

A specific savings of irrigation energy is achieved by establishing programs to test pumping plant efficiencies. The University of Nebraska found that 58 percent of the irrigation pumps tested in one program used twice as much fuel as should have been required. Common problems were worn impeller seals and poor impeller adjustment. Proper matching of wells, pumps, and engines can both conserve fuel and cut production costs.[46] Such adjustments offer an opportunity to improve productivity of capital using present technology.

Virginia dairy farms with no-till corn production have higher returns than Virginia dairy farms with conventional corn systems because with the former system higher corn grain and corn silage yields are achieved and—more importantly—more acres of sloping land can be used annually for growing corn. Although the no-till corn system requires less engine fuel per acre for land preparation, the no-till system uses more energy inputs in the form of nitrogen fertilizer and pesticides.

Higher returns with the liquid manure system, compared with the solid manure system, show another opportunity for improving productivity.[47]

The grape pest management program in Erie County, Pennsylvania, has demonstrated the value of regular pest scouting and the application of insecticides and miticides based on needs. Presently an ad hoc committee of agricultural leaders in the county is exploring the concept of establishing a pest management cooperative. Pest management growers usually spot sprayed for pest insects and mites, treating generally less than 20 percent of their vineyard areas; nonparticipants, by comparison, treated 75-100 percent of their areas.[48]

A lack of information about opportunities for improving capital use (rather than lack of motivation) may exist. Some practices widely accepted on most commercial farms were still not being used on many farms in a Kentucky study. These findings suggest opportunities for more in-depth extension efforts with small farmers.

Technology has been developed to process and convert animal wastes into sanitary, palatable, and nutritious cattle feed. The Ceres Ecology Corporation of Sterling, Colorado, working with scientists at Colorado State University, has put a full-scale plant into operation that processes the waste from 8,000 beef cattle per day.

The new nitrogen inhibitors appear to have good effectiveness at reducing the energy inputs to corn production; they inhibit the release of the nitrogen into a form that will leach away before the corn roots have a chance to utilize it. These inhibitors are already being recommended and used in the Midwest. Their use can be expected to move to other corn producing areas.

A more "old-fashioned" practice is to include more legumes such as alfalfa into a rotation. Increased alfalfa production could result in major energy improvements if losses during harvest are minimized and if marketing is profitable through livestock or as dry hay.

Corn drying is an economical practice at current prices because it saves the significant field losses of corn that can occur if the crop is allowed to stand in the field late in the fall before harvest.

It is not profitable at current prices to delay harvest to reduce energy drying of corn, but developments in crop residue gasification, heat pumps, and solar drying could greatly reduce the major fuel requirements of drying.

Production of a double crop (e.g., wheat followed by soybeans) is 25 percent more energy efficient than wheat alone but less energy efficient than soybeans alone and less profitable than corn or corn-soybean rotations at current prices.

A possible method of drying corn with lower energy input is the use of part of the corn residue for drying fuel. Either incineration or gasification of the crop residue could provide enough energy for current drying methods. Less than 50 percent of the cobs could be used in a gasifier to dry the corn down from 28 percent to safe storage moisture content.

Efforts now underway to lower fossil fuel requirements for crop drying include (1) the increased feeding of high-moisture corn to animals on or near the farm producing the corn and (2) the combining of high-speed high temperature driers with storage systems utilizing low-temperature or natural air drying.

Timely operation and timely application of energy-based inputs can be one of the most important factors in increasing the overall energy efficiency of a cropping system. The use of livestock for utilization of the hay and the subsequent use of manure to further replace fertilizers add to the reasons for looking to legumes for improving energy efficiency. A winter legume cover crop between soybeans and corn in the rotation can improve energy efficiency by 20 percent and can aid in soil and water conservation, but it is not economically feasible at current nitrogen fertilizer prices.[49]

Imposition of energy restraints would appear to sharply reduce field crop acreage and to increase vegetable acreage.

FUTURE APPROACHES TO SAVING ENERGY

APPLICATION OF SOLAR ENERGY

Alternative energy sources based on something besides fossil fuels show considerable potential for future adoption. For example, solar heating systems, such as grain dryers and water heaters, and windmills (or wind turbines) for electrical power generation may prove economically feasible. Solar energy appears to be especially well suited to low-temperature drying, which utilizes small increases in drying air temperature over extended periods. Although a constant source of energy for drying is preferable, intermittent application of heat is acceptable in low-temperature drying. The long-time drying can continue even when rainy weather or heavily overcast days keep the sun's energy from penetrating the earth's surface.

Solar energy can be stored, for example, by heating a tank of water for use at night and on cloudy days. However, storage systems are bulky, require additional equipment for reclaiming the stored energy, and, consequently, are costly to install and maintain.

Economic studies of solar grain drying have lagged behind the feasibility studies. Solar energy is a renewable resource that can be expected to provide a greater and greater portion of our future energy requirements.

The economic comparison between conventional and solar photovoltaic energy systems for irrigation suggests that the latter will become profitable in the middle to late 1980s if the cost of solar modules follows Department of Energy projections. These results are remarkably robust and insensitive to reasonable variations in discount rates, fuel escalation rates, and support system costs. Not surprisingly, solar-powered systems become viable earlier in those areas with larger amounts of year-round insulation or longer irrigation seasons. However, because of the expected continual decline in the costs of solar modules between now and the year 2000, the optimal year of investment is later than the initially viable year.[50]

Widespread implementation of the outputs of solar energy programs can make agriculture more energy self-sufficient, provide jobs in rural areas, and reduce importation of oil. The shift to solar energy production and use and the reduction in fossil fuel energy use will require new capital investments on farms, construction and installation of new equipment, and adaptation of farm production facilities. All these changes will require labor and provide employment.

Faced with finite reserves of fossil fuels, agriculture will be forced to exploit other sources of energy. Therefore, the development of solar energy and hydrogen as future fuels deserves attention.

However, the physical flow of inputs and outputs is only one aspect of the choice of alternative cropping systems and energy-saving practices. Such things as management requirements, risk, resource bottlenecks, and economic feasibility of alternative cropping systems have more to say about their potential adoption than their product energy efficiency does. The analysis of energy flows is specifically linked with analysis of the economic feasibility, potential bottlenecks, and other management factors for each cropping system.

BIOMASS AND GASOHOL

The production and use of biomass as an energy source is of particular interest to agriculture because of the close relationship of biomass production to food, fiber, and forest production. The development of biomass for energy and petrochemical substitutes will provide new markets for agricultural products and will make major contributions to rural energy self-sufficiency.

Generation of methane from manure provides a fuel, permits handling of a pollutant in an environmentally acceptable manner, and provides a protein-rich residue which can be fed to livestock.

Ethanol proponents also say they have developed a new, cheaper process for making ethanol out of plant leaves, stalks, and husks. George Taso at Purdue University has developed this process, which would cut production costs by reducing the time required for fermentation and would insure an inexhaustible supply of raw materials without diminishing world food supplies. This process could produce twice the ethanol using half the energy of a process using whole grain.[51]

Small-scale alcohol plants near the source of biomass fuel are another promising concept, and the Department of Energy is interested in farm-based alcohol plants. The brewer mash insoluble residue can be fed to animals without further processing. However, generally more process energy is required to produce grain alcohol than is contained in the alcohol produced. The big hope here is that some innovative approaches to the fermentation and distillation systems will overcome this limitation.[52]

In Ontario potatoes, sugar beets, and Jerusalem artichokes are looked upon as potential raw materials for alcohol production and as a way to dispose of grain surpluses and low-quality grains infested with insects, mold, or disease. Ontario officials see crop distillation as most practical when done by a custom operator with a portable distillery on a tractor trailer.[53]

In 1975 Brazil established PORALCOOL—the National Alcohol Program. The program calls for blending ethyl alcohol with gasoline to save foreign exchange, increase gross national product, and expand the domestic capital goods industry. A pilot plant uses manioc (the cassava root) as a raw material and is fueled by burning eucalyptus wood, freeing alcohol production from dependence on imported petroleum (except to provide the cultural energy input used to produce the crop).

Fermentation of corn to ethanol for industrial uses is roughly economical with corn prices at $2 per bushel. The potential market would be for about 110 million bushels of corn per year.

Waste materials could become energy sources. Gasohol from this source may be economical. However, the process is still in the laboratory stage of development. The pilot plant stage must be completed before researchers will be confident of the economic and technical performance of the process.

Fermentation of corn to ethanol for blending with gasoline to produce "gasohol" is not economical at current or expected future prices for corn or gasoline. Some form of government subsidy would be necessary to make the gasohol proposition economically viable.

One issue that has created considerable interest is the net energy efficiency of the corn-to-ethanol conversion process. It is unclear whether the process of converting corn to ethanol consumes more energy or less energy than the process of converting petroleum products to ethanol consumes.

Another unanswered question concerns removal of the crop residues from the field, the effect of removal on soil fertility and tilth, and the development of equipment to collect waste material without interfering with the harvesting of the grain crop. Methods of handling the residue, storing it, and supplying it to a processing plant throughout the year will also need to be worked out.

LESS CAPITAL-INTENSIVE FARMING

Another approach to saving energy is less capital-intensive use and some increase in labor use. One study selected 14 matched pairs of cornbelt farms, one set using organic methods without chemical fertilizers or pesticides and the other using conventional methods.

The organic farms produced a somewhat smaller crop output per acre of cropland than their conventional counterparts produced; however, the variable costs of crop production per acre were higher for the conventional group. The organic group used considerably less energy per dollar of crop output and per acre of cropland and had slightly greater labor input per acre.

However, the samples used in this study may not accurately represent the larger populations of farms from which they were drawn, and errors could have arisen from reliance on farmers' reports on yields, field operations, and purchased inputs and on other data used in the calculations. The main conclusion reached was that further research into organic methods of crop production (as well as methods that lie between the conventional and organic alternatives) was warranted.

Another study compared farming methods of the Amish in Pennsylvania, Illinois, Wisconsin, and Nebraska with the conventional farming style of their neighbors.

In central Pennsylvania the yields from the two forms of agriculture were much the same, with the Amish yields 4 percent higher than that of their neighbors. No penalty in the form of reduced production stemmed from the reduced energy use of the Amish, supporting the optimistic hypothesis that high agricultural production can be maintained with reduced use of energy on the farm as well as in the home.

In Nebraska, however, the Amish energy ratio was 49 percent higher than that of the older order Amish of Pennsylvania, and their yields were 47 percent lower. This group apparently failed to take advantage of scientific developments as extensively as other Amish farmers.

In eastern Illinois the energy ratio of the Amish was 25.3 percent higher, but their yields were 38.1 percent lower. The results support the pessimistic hypothesis that a decline in energy available to agriculture would cause a decline in food production.

The conservation achievements of the Amish were greater in their homes than in their farming. Their major contribution to energy conservation is the limited demand they make on available resources to support their way of life.

Even though the results are mixed, the study suggests a potential for reintroducing human labor without major losses in production as long as key supplies such as fertilizers are available.[54] The Amish farmers in Pennsylvania apparently used fertilizer much as other progressive farmers did; the Amish in Nebraska did not.

Breimyer assesses the capital use and energy alternatives and points out that choosing between trying to protect agriculture's access to remaining supplies of nonfarm inputs of extractive origin, such as oil, and the opposite course of developing new, imaginative alternatives from produced raw materials is inevitable. Whenever a farmer plows under legumes (to substitute for fertilizer), or dries corn with sunlight (a substitute for natural gas or electricity), or delivers fresh vegetables to a nearby village market (a substitute for long distance transportation) "he counteracts by agrarian means the consequences of an oil supply that is not only running out but being reserved by its holders."[55]

TAX POLICIES AND THEIR EFFECTS ON FARMING DECISIONS

Tax policies at Federal, state, and local levels of government affect farming decisions. The impact upon efficiency and income distribution is debated by agricultural economists.

Tax policy can be a means of discouraging establishment of very large farms. But how real estate, income, capital gains, and estate and inheritance taxes affect farm operators' decisions needs to be evaluated.[56]

The tax system is often seen, both inside and outside of agriculture, as a way to write off costs against ordinary income, and through capital gains pay a lower rate than on ordinary income.

Progressive income taxes continue to place restraints on farm size but not to the extent that they once did. The tax rate changes of the last ten years favor increases in the scale of farming and expansion through borrowed capital. In 1978 changes decreasing tax rates mean higher after-tax income, but the effect will be to support the movement toward fewer and larger farms.[57]

The ability of the farmer to benefit from special tax provisions is directly related to the farmer's marginal income tax bracket. The higher his tax bracket, the greater the benefits. The high-income farmers who incorporate their farm operations can realize additional tax savings and gain access to subsidized capital through retained earnings to buy more land, resulting in a greater concentration of land ownership among fewer and larger farms.

Investment credit can be an effective reduction in the price farmers pay for machinery and other eligible equipment. Along with accelerated depreciation and additional first-year depreciation, tax savings of up to 50 percent of the purchase price can sometimes be realized in the first year of purchase.[58]

The tax-saving benefits accruing to nonfarm investors can lead to a distorted pattern of income distribution and a stimulus for forming a small number of large units in order to take maximum advantage of credit, tax, and price support policies. The risk-bearing capacity of the family type farm is its greatest comparative advantage; the reduction of risk, therefore, tips the makeup of farm units toward large-scale highly capitalized enterprises.[59]

However, tax shelters for farming operations are not necessarily a waste of capital resources. Williams views the firms most directly involved in creating tax shelters as managed by intelligent, reputable, and responsible people for whom the long-run benefits of access to risk capital are too great to jeopardize through abuses for short-term gains. For example, through funds from private investors the cattle-feeding industry can adopt a new lower-cost technology and management practices and improve efficiency.[60]

MANAGEMENT OPTIONS FOR IMPROVING PRODUCTIVITY

The options for improving productivity through management include upgrading educational achievement of farm operators, determining the most efficient size of operation, use of integrated pest management, double cropping, and new management technology employing computers.

EDUCATION

Peterson and Hayami, in summing up the literature, report that the productive value of education has its roots in (1) the worker effect and (2) the allocative effect. The first increases the marginal product of labor given the level of other inputs. The second enhances the worker's ability to acquire and interpret information about costs and new inputs.[61]

Doering suggests that large substitutions of resources like labor for mechanized energy are not likely but that increased management ability and time may be substituted for purchased energy. Areas that require additional management input include integrated pest management, fuel saving methods, application of fertilizers and pesticides and other cultural practices to reduce machinery use.

FARM SIZE

Much has been written concerning the relative efficiency of different scales of operation. Some range in size is probably necessary to permit the best use of the varying natural capabilities of farm operators. Even if there were no differences in the relative technical efficiency of farms of different sizes, a pattern of differences would make more efficient use of managerial resources.[62]

Associated with, and partly because of, changes in the composition of inputs, the average size of farm business has also changed. Increases in the scale of farm operations, that is, control over resources of production, are associated with adoption of technology. The dynamics of this interrelationship of expansion and technology usually result in lower per unit cost, particularly of fixed inputs. Because of more efficient use of inputs, the increases in sizes of farm businesses that have occurred and are now taking place result in higher productivity for agriculture as a whole.[63]

Technological explosion in farming has not destroyed the family farm and the self-employment associated with farm production. The most logical reason seems to be the biological nature and spatial dispersion of farm production which makes a large concentration of capital, management, and labor more difficult in farming than it is in other industries. So, competitive efficiency or least-cost production is attained at a smaller firm size in farming than in many other kinds of production. These smaller sizes are being adapted to the managerial and working capacities of self-employed family labor.

Although larger farms have an income advantage, costs per acre changed very little from 500 to 1,200 acres on Illinois corn and soybean farms. And although benefits from government farm support programs have been directly proportional to the size of the farm, policymakers may want to recognize in future policy considerations that policies that encourage development of corn and soybean farms above 1,200 acres may actually encourage a degree of inefficiency and loss of productivity.

Significant economies of size were noted in California dairy herds of up to 750 cows, but most cost reductions were achieved by using dairy herds of 375 to 450 cows. But specialized management is essential to achieve the cost savings.

Restricting farm size will reduce overall economic productivity. However, the significant advantages of larger farms seem to result from higher sales per acre and greater access to more irrigated land and from superior management.

Efficient family-size corn farms would achieve most of the possible economies of production. Very large farms may realize additional savings from reduced input costs, marketing advantages, financial leverage, and tax minimizing strategies.[64]

In Yolo County, California, farms of about 600-800 acres appeared to compete on a unit cost and profit basis with much larger farms. At least 200 acres were required to break even. In the Imperial Valley minimum costs per dollar of total revenue began at 2,000 acres, but most cost savings had been reached at about 700 acres. In the Westlands District most economies of scale had been reached at 800 to 1,000 acres.

The U.S. Department of Agriculture, in assessing the effects of enforcing the Reclamation Act of 1902, concluded that the California cotton farm of 640 acres has a 14 to 16 percent cost advantage over the 160-acre farm; the 480-acre farm in the North Platte valley of Nebraska has cost advantages of 15 to 20 percent over a 160-acre farm in the Columbia Basin; and the 320-acre farm has a cost savings of 8 to 12 percent over the 160-acre farm.

The reason for cost savings is the ability to spread certain fixed costs, such as management, supervision, and costly machinery and equipment, over more units of output. Integrated firms can reduce the middleman's margin and handling costs since the input firm provides inputs for its own farming operation.[65]

The more serious threat to the moderate-size farm is the very large-scale family farm which grows even larger. Heady recommends research for this intermediate-size farm, which has few opportunities to grow larger and which contributes so much to the viability of rural communities.[66]

Interest in the size of a farm is more than just an interest in finding the productive and efficient combination of resources. Low incomes of farm families and distribution of agricultural resources among farmers are major concerns. Questions concerning economic efficiency are frustrating to the ardent supporters of smallness who feel that somehow criteria for economies of scale ought to be changed so that the small-scale operation in agriculture becomes economically viable where it would not have been before.[67]

INTEGRATED PEST MANAGEMENT

The interest in more efficient pest control has encouraged development of integrated pest management programs. The falling productivity of insecticides during the later 1960s, when the benefits of insecticides didn't seem sufficient to pay the costs, is a factor in this development.[68]

Integrated pest management is defined as a system of using all suitable techniques and methods in a compatible manner that maintains the pest population at levels below those causing economic injury.

These programs are likely to move toward interdisciplinary and multiple component approaches, such as natural enemies and minimal insecticide use, in which improved pest monitoring and prediction systems are essential elements but the final delivery message to the producer is clear, plain, and unencumbered with excessive details.

DOUBLE CROPPING

Double cropping (following wheat with a soybean crop in the same year) improves the output from a given acre of land and improves the output/input ratio of energy considerably. It is risky from the weather standpoint, but the simulator accounts for this risk. This practice is on the increase, and the resulting energy efficiencies will recommend it further.

The comparison of crop returns in Illinois suggests that double cropping wheat land can compete with the rotation of traditional single crops. Four crops in three years—corn-soybeans-wheat, followed in the same year by either soybeans or grain sorghum—would be slightly more profitable than the usual corn-soybeans or corn-soybean-wheat rotation.

COMPUTER USE

Computerized information systems could conceivably improve management decisions in the future, as computer processing power will become readily available to most small businesses, including farms. However, nothing on the horizon in computer technology can replace the human skill of understanding the problems, choosing appropriate data bases, and applying the required processing needed to transform data into useful information.[69]

THE INFLUENCE OF INTERNATIONAL TRADE

International trade in U.S. agricultural products influences the allocation of resources on U.S. farms. In 1977 31 percent of the harvested crop acreage was exported, most of which was food grains, feed grains, and oil seed crops. In terms of the percent of production exported, the Department of Agriculture estimated that 73 percent of the rice, 69 percent of the sunflower seed, 58 percent of the cattle hides, 56 percent of the wheat, 55 percent of the soybeans, 45 percent of the almonds, 40 percent of the cotton, 36 percent of the tobacco, 35 percent of the dried prunes, and 28 percent of the corn moved into the export market.[70]

Variation in the amounts of individual commodities exported each year is directly related to supplies produced in other parts of the world, either in the importing countries or competing export countries. Consequently, world prices and prices received by U.S. producers are directly affected by the production of the export commodities in other parts of the world. Although exporters must report sales to the USDA, the Federal Government has made no effort to control or limit exports since 1975.

National policy written into the 1977 Agricultural Act and the general opinions of producers oppose government intervention in the export of U.S. farm products. The on-farm commodity reserve created in the 1977 Agricultural Act is a major effort to stabilize prices during periods of reduced supplies or rising demand. But U.S. policy has not faced up to the crucial question of how a major national shortage of production inputs or domestic farm products would be handled in view of our strong export market and the benefits of a positive trade balance in farm commodities.

For example, the full allocation of all petroleum needs for agriculture and the withdrawal of that priority allocation in the spring and summer of 1979 suggest potential problems in the future if further shortages of these products occur.

The rising prices of grain in 1975 and the temporary limitation of exports stimulated by pressure from consumer and labor groups also show what can happen if national policies are not thoroughly planned in advance for such emergency situations. During the marketing year following the reduced 1974 crops, exporters maintained their demand and the domestic users cut back consumption of feed grains, resulting in lower meat supplies in succeeding years.

When domestic supplies of commodities are reduced while large volumes of exports continue, higher domestic prices can be expected. Actually, the buyers abroad and the buyers at home are competing for a shorter supply. Domestic prices will rise to ration the shorter supplies available. But if exports are reduced, causing larger trade deficits, and the foreign exchange value of the dollar declines, then U.S. consumers will have to pay more for imported goods.

Nor has a policy toward the potential competition between U.S. products and imports of certain foods been fully developed. Martin and Rochin point out that trade and immigration policies affect agricultural labor markets by altering the demands for labor. As Mexican irrigation systems expand and as the real cost of transporting agricultural commodities between the United States and Mexico and the Orient declines, the production of labor intensive products may shift to countries with lower labor costs.[71]

LOOKING AHEAD: PRODUCTIVITY ISSUES AND POLICY OPTIONS

LAND USE

The major options for improving productivity of land center around (1) retaining "prime" lands for agricultural production; (2) letting the market allocate land for agricultural use and other purposes; and (3) subdividing large farm units, particularly those receiving water from federally financed projects, into 160-acre operating units as provided by Federal law.

Data from the 1975 Soil Conservation Service Potential Cropland Study identified 78 million acres of noncropland with a high potential and 33 million acres having medium potential for conversion to cropland. However, when all high potential acreage was analyzed, only 15 million acres had no limitations to development. However, 7.4 million acres converted to urban and water uses between 1967 and 1975 were estimated to be from prime farmland. The data suggest that continued loss of prime farmland at the 1967-75 rate could eventually lead to a loss of prime farmland that could not be replaced except with inferior land with lower productivity.

Although the complete preciseness of the data is uncertain, the general picture reenforces the viewpoint that combined Federal, state, and local government efforts to retain prime lands for agricultural production have long-range benefits for overall agricultural productivity.

The alternatives to preserving prime lands are to spend more energy to obtain higher production from the remaining farmlands or to bring marginal lands into agricultural production.

On the other hand, Gardner cites the economic consequences of land preservation legislation:

> If the market is removed from the allocation task and is replaced by essentially nonmarket allocation criteria, the consequences in terms of economic efficiency and equity should be carefully evaluated.
>
> When property rights are altered, such as would be the case if land disposition is restricted, the signals are distorted, and the market does not work efficiently. If society really has a scarcity of agricultural land and food and fiber prices are not artificially controlled at lower than equilibrium levels, the land market serves perfectly well to reveal this scarcity. If the effective market instrument is replaced by the prime land retention schemes where inadequate economic or political criteria dictate resource allocation, there is no reliable way of knowing whether agricultural land is becoming scarcer or more plentiful.[72]

LAND-LABOR RELATIONSHIPS

Subdividing large farm units to 160 acres could lead to more intensive labor use but would also probably reduce the productivity of labor, as more intensive cropping would be encouraged. Observations have shown that in the developed countries, where demand for labor in the nonagricultural sector has pulled a significant amount of labor out of agriculture, the rates of growth in labor productivity have been increased by improvements in the land-labor ratio.

In the less developed countries, where labor productivity growth has been primarily brought about by the increase in land productivity, the input of fertilizer per hectare has increased more rapidly than the input of machinery per worker. In the developed countries, in contrast, the progress of mechanization has been a primary source of growth in output per worker.[73]

Those who advocate more labor and less machinery per unit of land should recognize that this choice will affect productivity and that the environment of the United States is one of a developed country rather than a less developed one.

The options for labor use really focus on the question of whether the combination of factors used in agricultural production will include more labor or less labor.

Cost advantages of labor-saving technology will be lost if policies restricting size of farm are imposed. But such efficiency losses will vary according to the degree of restrictiveness. If size limitations appear desirable on social and other grounds, the sacrifice in efficiency should be clearly recognized.[74]

CAPITAL AND ENERGY USE

The options for improving productivity of capital in agriculture are closely allied with energy use. These include (1) making more efficient use of present technologies, (2) conserving fossil fuels by shifting to more use of solar energy, (3) increasing alcohol production from farm-produced materials to reduce use of petroleum-based fuels, and (4) shifting to less capital-intensive farming methods.

Some samples of more efficient use of present technology can be described. Fuel could be conserved at the present level of mechanization by avoiding the use of tractors delivering more power than is required for a specific cultural operation. Furthermore, tractors could be operated at speeds and in gears that minimize fuel consumption for various tasks. Increased custom hiring of certain cultural operations like combining or harvesting should result in more effective application of costly equipment that might otherwise remain idle for most of the crop year. Aerial instead of tractor application of seed, fertilizer, or pesticides might reduce energy consumption. However, the scientists preparing a National Academy of Sciences report and suggesting these approaches admit that knowledge of the energy savings attributable to these is lacking.[75] More consideration could be given to costs of smaller machinery packages or the opportunities to rent or contract for machinery on small units.

Large energy savings could not be accomplished by adopting less energy-intensive technologies, but individual farmers could be made less vulnerable to fluctuations in energy prices and supply.[76]

The cheap energy era is behind us, the alternatives are all expensive, and it will require long lead times to commercially develop significant quantities of alternate fuels. There are indirect or hidden effects of energy shortage and allocation schemes. Fuel allocation, guaranteeing 100 percent of the fuel needs for agriculture, for example, may not help much if other needed inputs become scarce and other bottlenecks occur.

Energy policy for agriculture will also affect decisions and productivity. Some believe higher energy prices would encourage efficient energy management, increase domestic oil and gas production, and make alternative energy sources economically viable.

Although tax incentives could encourage energy saving, the need for such incentives is questionable if farmers can be shown how good management can provide more profits while helping save energy.

So, it must be recognized that the economics of agriculture will henceforth comprise not only the economics of the land, together with capital from superimposed industrial processes, but also the economics of access to mineral raw materials on which our industrial technology is based.

The practice and politics of agriculture will be divided between: (1) sheltering agriculture from the injurious effects of more costly raw materials and (2) fostering the protection and development of the natural processes of plant and animal production that sustained mankind before chemical fertilizers or petroleum were ever heard of.[77]

Fuel scarcities and high costs are likely to be major motivating forces for moving away from direct use of fossil fuels for motive power. However, Doering suggests that this will not be possible all at once. In the short and intermediate term, a policy of investing in storage facilities rather than substituting human labor or animal power for fossil fuels may be more appropriate.[78]

TAX POLICY

Decisions on tax policy affecting agriculture involve these options: (1) keep present policies; (2) restrict investments by outside investors and business firms; or (3) restrict or eliminate cash accounting, deductible investments, investment credit, and capital gains now open to farm producers. Because the policies that apply to other businesses usually apply to agriculture, any changes in these policies—as well as in policies in effect especially applicable to agricultural production—should be reviewed carefully; otherwise, incentives to expand production and improve productivity could be jeopardized.

RESEARCH AND TECHNOLOGY

Public sector research and development remain the primary source of agricultural technology, particularly the biological technology that expanded the capacity of crops and animals to respond to higher levels of industrial inputs.

The continued high rate of productivity growth in the agricultural sector cannot, however, obscure the fact that the rate of productivity growth in agriculture has slowed down over the last decade. The serious issues related to the future agriculture performance of the U.S. are: (a) whether the U.S. agricultural research system is maintaining its capacity to support productivity growth and (b) whether the institutional constraints which condition the structure of American agriculture are becoming an increasing burden on productivity growth. Ruttan questions whether the United States is continuing to make the investments required to release the technical constraints on agricultural productivity.

Looking ahead, no major scientific breakthrough comparable to hybrid corn or DDT can be reasonably predicted for the next one to two decades, although some believe promising potentials remain for improving productivit" from the application of known technology and from ne .echnology now in the research and development phases.

Although one cannot predict the outcome of investments in research, many seem to strongly feel that reductions in research investments may explain why average agricultural output per hectare of cropland remains low relative to the levels that prevail in many other developed and developing countries.

In developing a long-range policy for maintaining agricultural productivity in the face of energy shortages, a number of options can be identified. Anticipated price increases and continuing supplies appear preferable to the serious problems that a disruption in availability of energy supplies could produce for the U.S. food system.

The recent concern with climate has underscored the fact that our fundamental knowledge of climatic processes and mechanisms is not sufficient to permit forecasting. Reliable forecasts of climate even one growing season ahead would be extremely useful, and predictions of trends one to two decades ahead are sorely needed for national policy planning.[79]

From both an environmental and energy perspective, more intensified research in pest control could include (1) integrated pest management systems based on cultural, genetic, and conventional chemical techniques, (2) selected biologically based interference with the reproductive processes of the pest species, (3) microbiological and natural biological agents for pest control, and (4) breeding for pest resistance.[80]

Some options for plant science research include (1) strengthening research support for the basic plant sciences, (2) intensifying and expanding international efforts to preserve, conserve, and exchange genetic resources, (3) broadening the range of parameters in plant breeding research, (4) developing more efficient methods of crop fertilization, (5) accelerating work on conventional breeding for higher and more dependable yields, (6) developing a variety of intensive cropping systems for the tropics, and (7) increasing the protein value of grain.

Specific livestock research areas identified as having significant potential for increased production both in the U.S. and developing countries include (1) controlling reproductive and respiratory diseases, (2) developing genetically superior animals, (3) improving nutrition efficiency, and (4) increasing the reproduction performance of all farm animal species.

ENERGY POLICY

A realistic national energy policy for our U.S. food system could include (1) adequate funding for basic and applied research relating energy supplies and utilization to an assured food supply; (2) adequate funding for basic and applied research in order to develop alternative energy sources and to effectively demonstrate their uses; (3) guaranteed supply of fossil fuels for production agriculture, including input supplies and transportation services, until alternative energy sources are assured; (4) a national land use policy to protect remaining cropland; and (5) deregulation of petroleum and natural gas prices to encourage an orderly shift to alternate energy sources.[81]

The potential for contradictory and counterproductive activities dealing with agricultural production is tremendous. As an example, some officials have logically urged the agricultural research and extension establishment to devote its efforts toward perfecting and disseminating more energy-efficient agricultural systems.

At the same time, many officials appear to want to exempt agriculture from the energy price increases that could be expected under the President's energy plan in order to hold down food prices.

If this latter policy prevails, there may be no incentive to adopt energy-efficient systems in agriculture. For if agriculture is guaranteed the energy it needs at current prices, there is little incentive for technological development or adoption.[82]

Strangely, the new cadre of energy specialists has tended to ignore the relative price issue, yet it is relative prices that bring about or negate resource substitution.

SOCIAL AND ECONOMIC GOALS

The goals for the future direction of U.S. agriculture must be identified. Some policy decisions may be made on the basis of economic efficiency, others on the basis of social equity. The persistence of small farmers and reversal of the recent rural urban migration indicate that more emphasis on research that would enhance the well-being of small and part-time farmers may be warranted.[83]

Should farm size be limited to allow more farmers to have access to land and farm operating opportunities? Heady and Jensen caution against putting an upper limit on farm size in order to increase farm numbers and people living on farms. They suggest a policy of having farms operate on a scale to insure a favorable living level for a reasonable number of farm families rather than having an excess of small farms surrounded by poverty conditions.[84]

Small farms do provide diversity in the structure of agriculture and an outlet for productive use of family labor. Small farm operators whose farm operation is only a supplement to off-farm income can survive longer than can full-time farmers with inadequate resources.

The smaller farmer may find direct marketing a means of supplementing income. Although direct marketing accounts for only a small percentage of total food marketed (less than 3 percent of fruits and vegetables), it is nevertheless important to a large number of small farmers, particularly those near population centers.

The resurgence of farmers' produce markets, pick your own fruit and vegetable farms, and simple consumer food cooperatives are probably more symbolic of agrarian than of industrial food delivery. And although current quantities in such direct marketing are trivial, the potential for the future could be substantial.[85]

POLICY GOALS IN PERSPECTIVE

Agriculture has special and specific needs for energy. The viability of the U.S. food system, which provides an estimated 90 percent of our domestic food needs plus another $30 billion for export, is essential.

Exports provide food needed by many countries—developed, developing, and centrally planned. Our food imports, equivalent to about half the value of our exports, provide tropical products which we cannot produce plus other foods which supplement our own production. To reduce exports would enlarge our trade deficits, raise the cost of our imports as exchange rates become less favorable, and eventually lead to lower standards of living for the American public.

The forces influencing the direction of future U.S. farm and food policy have divergent economic, political, and social implications for farm operators, farm workers, rural citizens, and consumers. Social objectives can be in direct conflict with policies that could promote improved economic productivity.

NOTES

1. U.S. Department of Agriculture, *Agricultural Outlook*, Economics, Statistics and Cooperatives Service, AO-39 (December 1978), p. 2.

2. United States General Accounting Office, *Changing Character and Structure of American Agriculture: An Overview*, CED-78-178 (September 26, 1978), p. 5.

3. Federal Energy Administration, *Energy Use in the Food System*, Booz, Allen, and Hamilton, Inc., Management Consultants (May 1976), p. iv-6.

4. Ibid.

5. Van Arsdall, R. Thomas, "Energy Requirements in the U.S. Food System," *Agricultural Outlook*, U.S. Dept. of Agriculture (March 1976), p. 18.

6. National Academy of Sciences, *Agricultural Production Efficiency*, Committee on Agricultural Production Efficiency, Washington, D.C. (1975), p. 125.

7. Van Arsdall, R. Thomas, and Rall, Edward, *Fueling U.S. Agriculture*. *The Farm Index*, U.S. Dept. of Agriculture 16:7 (July 1977), p. 4.

8. United States General Accounting Office, op. cit., p. 27.

9. Breimyer, Harold F., "Agriculture's Three Economies in a Changing Resource Environment," *Amer. J. of Agr. Econ.* 60:1 (February 1978), p. 37.

10. Loomis, Ralph A. and Barton, Glen T., *Productivity of Agriculture, United States, 1870-1958*, Technical Bulletin No. 1238, U.S. Dept. of Agriculture (April 1961), p. 11.

11. Fischer, Loyd K., "Environment, Technology and Farm Size," *Can the Family Farm Survive?* Special Report 219, Agricultural Experiment Station, University of Missouri-Columbia (1978), p. 49.

12. Loomis, Ralph A. and Barton, Glen T., op. cit., p. 20.

13. Martin, Philip L. and Rochin, Refugio, "Emerging Issues in Agricultural Labor Relations," *Amer. J. of Agr. Econ.* 59:5 (December 1977), p. 1046.

14. Binswanger, Hans P. and Ruttan, Vernon W., *Induced Innovation, Technology, Institutions and Development*, Baltimore: Johns Hopkins University Press (1978), p. 59.

15. Loomis, Ralph A. and Barton, Glen T., op. cit., p. 23.

16. Ibid., p. 20.

17. Hayami, Yujiro and Ruttan, Vernon W., *Agricultural Development: An International Perspective*, Baltimore: The Johns Hopkins Press (1971), p. 136.

18. Ibid., p. 85.

19. Christensen, Laurits R., "Concepts and Measurement of Agricultural Productivity," *Amer. J. of Agr. Econ.* 57:5 (December 1975), p. 910.

20. Schuh, G. Edward, "The New Macroeconomics of Agriculture," *Amer. J. of Agr. Econ.* 58:5 (December 1976), p. 803.

21. Loomis, Ralph A. and Barton, Glen T., op. cit., p. 18.

22. Hayami, Yujiro and Ruttan, Vernon W., op. cit., p. 74.

23. Ibid., p. 81.

24. Ibid., p. 67, 84.

25. Gardner, B. Delworth and Pope, Rulon D., "How is Scale and Structure Determined in Agriculture?" *Amer. J. of Agr. Econ.* 60:2 (May 1978), p. 297, 298.

26. Ibid., p. 298.

27. Fischer, Loyd K., op. cit., p. 51.

28. Barkley, Paul W., "Some Nonfarm Effects of Changes in Agricultural Technology," *Amer. J. of Agr. Econ.* 60:2 (May 1978), p. 310.

29. U.S. Department of Agriculture, "Food Price Still Climbing in Most Surveyed Countries," *Foreign Agriculture* (August 1979), pp. 18-19.

30. Johnson, Warren A., Stoltzfus, Victor, and Craumer, Peter, "Energy Conservation in Amish Agriculture," *Science* 198 (1977), p. 373.

31. Skold, Melvin D. and Penn, J. B., "Production Potentials in U.S. Agriculture," *Research Issues Facing Agriculture and Rural America, Economic Research Service*, U.S. Dept of Agriculture (September 1977), p. 91.

32. Skold, Melvin D. and Penn, J. B., op. cit., p. 78.

33. Breimyer, Harold F., op. cit., p. 43.

34. U.S. Department of Agriculture, *Perspectives on Prime Lands*, Background papers for seminar on retention of prime lands, (July 16-17, 1975), pp. 37-60.

35. United States General Accounting Office, op. cit., p. 51.

36. Skold, Melvin D. and Penn. J. B.. op. cit.. p. 81.

37. Binswanger, Hans P. and Rutton. Vernon W.. op. cit.. p. 57.

38. Hayami, Yujiro and Ruttan. Vernon W.. op. cit., pp. 71-72.

39. American Society of Agricultural Engineers. Task Group on Energy and Food, "An ASAE Public Policy Issues Report," *Agr. Engineering* 60:1 (January 1979). p. 32.

40. Martin. Philip L. and Rochin, Refugio, op. cit.. p. 1050.

41. National Academy of Sciences. op. cit.. p. 126.

42. American Society of Agricultural Engineers, Task Group on Energy and Food. op. cit.. p. 33.

43. Van Arsdall. R. Thomas and Rall. Edward. op. cit.. p. 5.

44. Doering. Otto C. III, and Peart, Robert M., *Evaluating Alternative Energy Technologies in Agriculture,* NSF/RA-770124. West Lafayette, Indiana: Agricultural Experiment Station. Purdue University (June 1977), p. 14.

45. Skold. Melvin D. and Penn. J. B.. op. cit., p. 89.

46. American Society of Agricultural Engineers. Task Group on Energy and Food. op. cit.. p. 35.

47. Burton. Robert and Kline. R. G.. "Adjustments in a Farm Business to an Energy Crisis." *Amer. J. of Agr. Econ.* 60:2 (May 1978). pp. 256-7.

48. Jubb. G. L., Jr.. Obourn. T. H., and Peterson. D. H.. "Pilot Pest Management Programs for Grapes in Erie County. Pennsylvania." *J. of Econ. Entomology.* 71:6 (December 1978). p. 915.

49. Doering. Otto C. III. and Peart. Robert M.. op. cit.. p. 11. 13.

50. Katzman. Martin and Matlin. Roland W.. "The Economics of Adopting Solar Energy Systems for Crop Irrigation." *Amer. J. of Agr. Econ.* 60:4 (November 1978). p. 652.

51. Teague, Carolyn. "Congress Looks to Gasohol in Search for Fuel Options." *Congressional Quarterly* (April 29. 1978). p. 10?7.

52. Stout. Bill A.. "The Energy Alternatives." *Agr. Engineering.* 60:2 (February 1979). p. 12.

53. Kirk. Mart. "A Canadian Views Alcohol as a Farm Fuel." *Agr. Engineering* 59:5 (May 1978). p. 13.

54. Johnson. Warren A.. Stoltzfus. Victor. and Craumer. Peter. op. cit.. p. 378.

55. Breimyer. Harold F.. op. cit.. p. 46.

56. Stanton. B. F.. "Perspective on Farm Size." *Amer. J. of Agr. Econ.* 60:5 (December 1978). p. 734.

57. Carman. Hoy F.. "Changing Federal Income Tax Rates and Optimum Farm Size." *Amer. J. of Agr. Econ.* 54:3 (August 1972). p. 491.

58. Woods. W. Fred. "Access to Capital in Agriculture: The Federal Tax Issues." *Can the Family Farm Survive*? Special Report 219. Agricultural Experiment Station. University of Missouri-Columbia (1978). p. 45. 46.

59. Raup. Philip M.. "Some Questions of Value and Scale in American Agriculture." *Amer. J. of Agr. Econ.* 60:2 (May 1978). p. 307.

60. Williams. Willard F.. "How Large Farming Operations Use Tax Influenced Investment." *Income Tax Rules and Agriculture*. Special Report 172. Agricultural Experiment Station. University of Missouri-Columbia, 1975, p. 16.

61. Peterson, Willis and Hayami. Yujiro. "Technical Change in Agriculture." *A Survey of Agricultural Economics Literature*. Volume 1. Lee R. Martin. Editor, Minneapolis: University of Minnesota Press (1977). p. 511.

62. Mighell. Ronald L.. *American Agriculture. Its Structure and Place in the Economy.* New York: John Wiley and Sons (1955). p. 70.

63. Loomis. Ralph A. and Barton. Glen T.. op. cit.. p. 24.

64. Krause, Kenneth R. and Kyle. Leonard R., *Midwestern Corn Farms: Economic Status and the Potential for Large and Family Size Units.* Agricultural Economic Report 216. Economic Research Service, U.S. Dept. of Agriculture (November 1971). p. iii.

65. Krause. Kenneth R. and Kyle. Leonard R., "Economic Factors Underlying the Incidence of Large Farming Units: The Current Situation and Probable Trends." *Amer. J. of Agr. Econ.* 52:5 (December 1970). p. 775.

66. Heady. Earl O., "Agricultural Production Economics in the Future." *Lectures in Agricultural Economics.* Bicentennial Year Lectures. U.S. Dept. of Agriculture (December 1. 1976). p. 156.

67. Doering. Otto C. III, "Appropriate Technology for U.S. Agriculture: Are Small Farms the Coming Thing?" *Amer. J. of Agr. Econ.* 60:2 (May 1978). p. 293.

68. Carlson. Gerald A.. "Long Run Productivity of Insecticides." *Amer. J. of Agr. Econ.* 59:3 (August 1977). p. 547.

69. Harsh, Stephen B.. "The Developing Technology of Information Systems." *Amer. J. of Agr. Econ.* 60:5 (December 1978). pp. 910-911.

70. U.S. Department of Agriculture. *1978 Handbook of Agricultural Charts.* Agricultural Handbook No. 551 (November 1978). p. 71.

71. Martin. Philip L. and Rochin. Refugio. op. cit., p. 1050.

72. Gardner. B. Delworth. "The Economics of Agricultural Land Preservation." *Amer. J. of Agr. Econ.* 59:5 (December 1977). p. 1033.

73. Hayami. Yujiro and Rutton, Vernon W.. op. cit.. pp. 71-72.

74. Seckler. David and Young. Robert A.. "Economic and Policy Implications of the 160 Acre Limitation in Federal Reclamation Law." *Amer. J. of Agr. Econ.* 60:4 (November 1978). p. 581.

75. Raup. Philip M.. op. cit.. pp. 126.

76. Klepper. Robert. Lockeretz. William. Commoner. Barry. Gertler. Michael. Fast. Sarah. O'Leary. Danile. and Blobaum. Roger. "Economic Performance and Energy Intensiveness on Organic and Conventional Farms in the Corn Belt: A Preliminary Comparison." *Amer. J. of Agr. Econ.* 59:1 (February 1977). p. 1.

77. Breimyer. Harold F.. op. cit.. p. 46.

78. Doering. Otto C. III. "Agriculture in the Year 2000. An Energy Perspective." *Amer. J. of Agr. Econ.* 59:5 (December 1977). p. 1068.

79. Skold. Melvin D. and Penn. J. B.. op. cit.. p. 95.

80. Corley. E. L. and Turnbull. James. *Agricultural and Food Research Issues and Priorities. A Review and Assessment.* U.S. Dept. of Agriculture. Science and Education Administration (December 1978). p. xii.

81. American Society of Agricultural Engineers. Task Group on Energy and Food. op. cit.. p. 36.

82. Doering. Otto C. III. op. cit.. p. 1069.

83. Corley. E. L. and Turnbull. James. op. cit.. p. 57.

84. Heady. E. O.. and Jensen. H. R., *Farm Management Economics.* Englewood Cliffs. N.J.: Prentice-Hall (1954). p. 449.

85. Breimyer. Harold F.. op. cit.. p. 41.

3 The Role of Science and Technology in the Containment of Health Care Costs

by Kenneth E. Warner*

SUMMARY

The high and rapidly growing costs of American health care constitute today's principal health policy concern. Both existing and proposed regulations are directed at moderating medical care cost increases, with attention focused on hospital costs and, within hospitals, on the acquisition and use of sophisticated technology.

Technological change is a central characteristic of modern medical care. Science continues to make important contributions to the high technology of medicine, that which is preventive or curative, inexpensive, and easy to use. Over the next few years, for example, several new drugs and vaccines should simultaneously improve health and decrease health care costs. By contrast, much new equipment-embodied technology will add to costs while performing "halfway" medical functions such as maintenance rather than cure of seriously ill patients.

Depending on how "technology" is defined, empirical analysis indicates a wide range of "culpability" in technology's contribution to the cost inflation problem. Ironically, much policy emphasizes control of the acquisition of expensive capital equipment (e.g., Certificate of Need and the proposed ceiling on capital expenditures), even though analysis suggests that capital costs per se are not a major factor in inflation.

The tendency to characterize technology as a cause of cost inflation must be balanced by recognition that the "technology problem" is a logical outcome of widespread third-party liability for the costs of medical care. Excessive acquisition and use of technology, and the emphasis on quality-enhancing rather than cost-reducing technology, reflect demand in a reimbursement environment in which resource consumption is divorced from financial responsibility. Piecemeal, bureaucratic regulatory mechanisms hold the potential of moderating inflation but seem unlikely to alter fundamentally either the problem or its result. By contrast, a reimbursement system which truly rewarded economic efficiency would create an immediate demand for cost-saving technologies.

The potential for science and technology to contribute to health cost containment lies as much outside as inside the medical arena. Historically, the greatest health contributions of science and technology have come from the prevention of illness, both in the medical care setting (e.g., immunizations) and through public sector alterations in the environment (e.g., sanitation). More recent prevention efforts related to the environment and to individual health behavior appear to be quite cost effective. Future scientific and technological developments related to genetics, behavior, and the physical and social environment might simultaneously promote health and contain health care costs.

Science and technology can be expected to continue contributing to improvements in health. Many health-promoting developments in the nonmedical areas seem

* Associated Professor of Health Planning and Administration, School of Public Health, University of Michigan.

capable of also contributing to the containment of health care costs, though their total impact will be distinctly limited given the concentration of health care expenditures in the delivery of medical care. Barring significant changes in the medical care reimbursement system, medical technology seems unlikely to contribute greatly to health cost containment. Rather, regulatory policies may diminish its contribution to inflation.

INTRODUCTION

The costs of American health care—their magnitude and rate of increase, sources, and what to do about them—constitute today's single most discussed health care issue. Concern vacillates between the amount of the present burden and the potential of the future burden. The former was over $200 billion in 1979, or more than a month of labor per American worker. The potential future burden is reflected in the rate of growth of health care expenditures which, in recent years, has exceeded the general increase in consumer prices by about 50 percent. As a result, health care expenditures have grown from 4.5 percent of the gross national product (GNP) in 1950 to almost 9 percent today. Perhaps more dramatically, if general inflation were to proceed at a rate of 9 percent per year from now until the end of the century and health care costs were to continue rising at a rate 50 percent greater than general consumer prices, in the year 2000 health care expenditures would exceed 20 percent of the GNP, or one out of every five dollars.

Clearly, health care costs are the focal point of current Federal health policy debates and initiatives. In a political environment dominated by concern about inflation and deficit spending, the pressures of health cost inflation are felt even more acutely by the Federal Government than by the population at large. The Federal share of health expenditures has been growing rapidly for more than a decade, from 21 percent in 1965 to more than 40 percent at present. Federal payments cover a significant proportion of the costs of inpatient care, and hospital costs—representing more than 40 percent of the total cost of care—have been inflating more rapidly than any major component of the health care sector. Thus, the Carter Administration has identified its hospital cost containment legislation as ". . . one of the most essential anti-inflation measures Congress will ever consider." Other components of the governmental effort to control health cost inflation include Certificate of Need (CON), other features of the Federal health planning law, restrictions on Medicare reimbursement, the work of Professional Standards Review Organizations (PSROs), promotion of Health Maintenance Organizations (HMOs), medical device regulation, formation of the National Center for Health Care Technology, and support of research on inflation and cost containment. It is noteworthy that several existing and proposed regulations are aimed at controlling costs through rationalizing—which usually means restricting—the acquisition and use of technology. Two prominent examples are CON and a national ceiling on capital expenditures. The latter was a central feature of the Administration's original hospital cost containment legislation and is now being proposed as an independent legislative effort.

To date, the Federal Government's principal response to the *health* cost problem has been to attack *medical care* expenditures, either piecemeal (e.g., CON) or institutionally (e.g., hospital cost containment). By far the largest source of the problem, the medical care system, is an obvious target for remedial measures. But this response has been symptomatic of a general failure to distinguish "health" from "medical care." Frequently simply a matter of semantics, this distinction serves as both a point of departure and an organizing framework for this paper's discussion of the role of science and technology in the containment of health care costs. Thus, the next section of the paper identifies the determinants of health, an appreciation of which is essential to understand how science and technology might impact health outside the medical care delivery system. This discussion is followed by an examination of the roles of science and technology in the delivery of medical care, in the past, at present, and with a conjectural glance at the next few years. The fourth section considers the past, existing, and potential roles of science and technology in the nonmedical determinants of and influences on health. The concluding section summarizes the main points and offers some derivative policy-relevant thoughts.

INFLUENCES ON HEALTH

The preeminence of medicine in the popular conception of health care dates primarily from the World War II period. According to many observers, it is a function of the technological revolution in medicine, originating with the development of antibiotics and growing with the emergence of increasingly sophisticated and complex medical hardware. The cures attributed to the "wonder drugs" and promised by ensuing technological developments have been credited with producing an overly optimistic reliance on medical care as the means of rectifying malfunctions in the human body. The failure of medicine to achieve the successes unrealistically desired of it and the soaring cost of medical care—itself a function in part of the technological revolution—have contributed to the rediscovery of the multiplicity of nonmedical personal and social influences on health status.

No single classification of the determinants of health is perfectly satisfactory, but the major determinants include:

- genetic factors;
- individual behaviors (diet, exercise, tobacco and alcohol consumption, stress, etc.);
- environmental influences, including:
  the physical environment (air and water pollution, radiation hazards, occupational hazards, etc.),

  the social environment (crowding conditions and their sequelae, etc.), and

  the political and economic environments (distribution of power and its uses, distribution of income, rate and nature of changes in economic conditions, etc.); and
- use of personal health services.

Two features of these classes of health determinants warrant brief attention. First, only the last class—personal health services—exists solely to influence health status. Genetic, behavioral, and environmental factors influence health quite independently of any conscious intent. That is, environmental pollution comes into being for reasons unrelated to health, and, unattended to, adversely influences health status. Similarly, the Industrial Revolution has eased our lives and increased our material well-being, but affluence and automation have contributed to less healthy sedentary lifestyles and patterns of food and drug consumption. One reason for the increased utilization of personal health services is to redress the damage done by these other factors. It is estimated, for example, that nearly 20 percent of the cost of medical care derives from illnesses attributable to cigarette smoking and alcohol abuse.

Modifications of or interventions in these areas can have a positive influence on health. For example, changes in lifestyle may have been contributing to recent increases in life expectancy, specifically linked to decreases in rates of heart disease. The important point is that genetic, behavioral, and environmental factors can influence health both "passively" (i.e., unconscious or unintended) and "actively" (i.e., purposeful modification). By contrast, personal health services invariably represent a conscious intervention to ameliorate, correct, or prevent other negative influences on health.

The second feature of the classes of health determinants is that three of the four classes represent primarily *individual* influences on health, whereas the remaining class contains *collective* influences on health. Genetic and behavioral factors and utilization of personal health services have discrete effects on the health of individuals, while environmental factors tend to influence the health of groups of people collectively. For example, if a city is heavily polluted, many residents will suffer adverse consequences. They cannot avoid the problem individually, short of moving to another community or perpetually wearing gas masks. Similarly, any intervention that reduces this health hazard will have beneficial consequences for all those who suffer from it, and not exclusively for single identifiable individuals. By contrast, genetic defects cause problems for individuals; through amniocentesis and counselling, individuals can reduce their own risks. An individual can convert a sedentary lifestyle into an active one. An individual can seek and receive health-promoting medical care.

These distinctions have implications for strategies to contain health care costs, and for the role of science and technology in achieving this goal. For example, the collective influence of environmental factors allows numerous people to benefit from single programmatic or technological efforts. This generally means that the additional cost of serving an additional individual is negligible, an attractive feature for dealing with common health problems in concentrated populations.

SCIENCE AND TECHNOLOGY IN THE DELIVERY OF MEDICAL CARE

Technologically, the medical care delivery industry is something of an anomaly. In the public mind, medical care suggests a world of pills, solutions, translucent tubes and large gleaming stainless steel devices. In fact, in many respects the field of medicine is technologically backward. The provision of care continues to be a labor-intensive activity and technological advances which have pervaded the mainstream of everyday life are only belatedly finding applications in medical settings (witness the use of microprocessors).

Nevertheless, every month one finds scores of articles in the professional literature and popular press praising or denouncing new medical devices, drugs, and procedures. Despite the discordance over the value of the new technology, the volume and nature of attention reflect the widespread agreement that the delivery of medical care is experiencing a technological revolution of unprecedented magnitude. Both current developments and those just over the horizon strongly suggest a continuation of this trend into the foreseeable future, though important aspects of its character may change.

Medicine's technological revolution dates from the 1940s, when science introduced perhaps the single most significant contribution of modern medicine to human welfare: the antibiotics. In a short period, these drugs relegated common scourges—major sources of suffer-

ing—to the history books. The antibiotics and other chemical products and biologics (e.g., Salk vaccine) are what Lewis Thomas has characterized as the true high technology of medicine: effective, preventive or curative, inexpensive interventions which are easy to use.

Witness to the excitement and optimism that these developments engendered was the rapid growth in the National Institutes of Health (NIH). For roughly two decades following World War II, NIH's budget grew at a rate of 20 percent per year. But the gleaming image of biomedical science tarnished over time as research failed to produce "wonder drugs" comparable in import to the antibiotics. Public interest in biomedical research was sparked momentarily in the early 1970s by initiation of the War on Cancer, but politically motivated expectations for "a cancer cure" were not realized. Coupled with declining public optimism, the concern over health care costs has contributed to more than a decade of budgetary stagnation at NIH.[1] Recent social science research raises the question of whether this budgetary restriction may be penny wise and pound foolish. The tangible fruits of basic biomedical research are unpredictable, invariably deferred, and frequently difficult to attribute to individual research efforts. Nevertheless, longitudinal analysis suggests that biomedical science has returned health dividends and possibly kept health care expenditures below the levels which otherwise would have been expected.[2]

Although the dramatic impact of the early wonder drugs has not been repeated, research continues to make chemical and biological advances which significantly affect the practice of medicine. For example, as a modality, cancer chemotherapy remains experimental, yet many childhood victims of acute leukemia treated with chemotherapy are today healthy adults, and drugs constitute an integral component of successful therapies in Hodgkins disease, breast cancer, and other malignancies.

Several new drugs may enter the class of "high technology" by improving health and lowering health care costs in the next few years. Some examples are:

- The newly licensed *pneumococcal vaccine* holds the potential of containing costs by limiting the severity of pneumonia and thereby reducing the number of patients requiring treatment, particularly among the high-risk elderly population.[3]

- Introduced just two years ago, *cimetidine* appears to have dramatically decreased the incidence of surgery, and hence the length of hospitalization, for duodenal ulcer patients. Associated cost savings should be appreciable and could grow as physicians increasingly treat other ulcer patients with the drug.[4]

- A new anti-convulsive drug, *valproic acid*, may help more than half a million epileptics and thereby avoid hundreds of millions of dollars in medical and welfare costs, as well as significant indirect costs associated with productivity losses.

Several diseases, including multiple sclerosis, rheumatoid arthritis, and diabetes, may be the product of viruses which, although not readily susceptible to treatment, may be preventable. The search for suitable vaccines is a research effort with potentially dramatic returns in reduced expenditures and suffering.

Other drug research areas are more problematic. Once a hotbed of research, the pharmacological control of mental disorders has not realized the potential many scientists anticipated. The costs of mental illness are immense, so progress in the control of such illness could yield economic as well as humanitarian rewards.

Along with its successes, the growing medical pharmacopoeia has produced health and cost problems. Adverse drug reactions, often reflecting drug interactions, account for numerous hospitalizations and increased lengths of stay. Information systems to monitor and avoid adverse reactions are under study in several clinical settings. The cost containment potential of such systems is today a matter of conjecture.

The early "wonder drugs" successfully attacked a group of devastating acute and infectious illnesses. In part due to this success, the problems remaining for science and technology to attack have been the more intractable chronic diseases. The failure of research and engineering to realize victories comparable to those of the early drugs well may reflect the greater complexity and intractability of these remaining diseases.

In relative terms, the character of the bioscience revolution has shifted from a biological and chemical emphasis to a mechanical one. Many of the vaunted developments of the day reflect the application of engineering knowledge to the production of equipment-based technology. The most discussed development of the 1970s, computerized tomography (CT scanner), is a perfect case in point. Unlike their biochemical predecessors, the new devices generally are not intended to serve preventive or curative functions. Rather, they are diagnostic (e.g., automated lab tests; CT) or oriented toward maintenance, i.e., controlling but not eliminating a problem (e.g., renal dialysis). Maintenance technologies, in particular, require repeated or continued application. Often called "halfway technologies," such devices tend to be complex and expensive.

In health policy circles and in the public mind, "medical technology" has come to refer to sophisticated capital equipment generally accompanied by a large acquisition price tag. Unlike their biochemical predecessors, the new equipment-embodied technologies are frequently criticized as contributing significantly more to cost inflation than to the betterment of health care. As is discussed below, the new technology does seem to be responsive to the idiosyncratic economic environment of medical care, but the nature of technology's contribution to inflation is only poorly understood. Much policy is based on the perception that the capital costs of sophisticated, elaborate technologies are responsible for the lion's share of technology-induced inflation (e.g., CON and the proposed national ceiling on capital expenditures), but empirical evidence suggests otherwise. The importance of capital acquisition appears to be growing, but the major costs of equipment-embodied technology are the operating costs, particularly including personnel costs (e.g., specialized technicians).[5] More significantly, considerable evidence identifies mundane technology with a low cost per use as the major villain in the technology-cost drama. The accessibility, ease of application, and third-party financial liability associated with many diagnostic tests have led to their excessive application, translating a small cost per use into large aggregate expenditures. For example, from 1951 to 1971, the number of laboratory tests per diagnosis of perforative appendicitis in the Palo Alto Medical Clinic rose from 5.3 to 31.0, with no significant changes in patient outcome. The average cost of treatment (including all inputs, not just lab tests) rose from $516 to $2,062. Controlling for general inflation, one still finds a 156 percent increase in the cost of treatment.[6]

The phenomenon represented by this example is a major source of difficulty in translating the cost containment potential in many technologies into realized containment of costs. Numerous new technologies are less expensive than their predecessors to acquire (i.e., capital costs) and/or use (i.e., operating costs); were they used simply as substitutes, they would reduce health care costs. Often, however, their convenience, simplicity, and like features combine with a permissive economic environment to produce more frequent use than was the case with their predecessors, and than appears to be necessary. Thus, in the aggregate, cost containment potential is lost. Low-cost automated EKGs are an example in many delivery settings, as is CT scanning. The latter is a demonstrably cost-effective technique for making certain diagnoses, but in numerous institutions its novelty and noninvasiveness have resulted in significant increases in the numbers of diagnostic tests performed relative to presenting signs and symptoms.

It is important to understand the medical economic environment, for economic incentives play a significant role in explaining the nature of medical technology development and patterns of technology acquisition and use. Technology does not simply "force" itself into the system. It is demanded—by hospitals, physicians, and patients—by a system that appreciates its services, and, at minimum, tolerates, or more often ignores, its cost implications. Extensive third-party payment, through both private insurance and public programs (primarily Medicare and Medicaid), removes the economic burden of resource consumption from those who produce it. If a patient's insurance will pay for a diagnostic procedure, the cost of that procedure is of no dir[illegible]ern to either the patient or the physician. Thus i[illegible]e p[illegible]dure offers a chance, however minuscule, of [illegible]vin[illegible] or solidifying a diagnosis, the procedure m[illegible]rdered. The burden is spread over the collectivity of the population paying taxes or insurance premiums. In the aggregate, such individual resource consumption decisionmaking spells high costs.[7]

Escalation in the high costs of care has resulted in part from growth in the extensiveness and depth of insurance coverage. In 1950, private health insurance paid for only 9 percent of personal health care expenditures, with government covering only 20 percent; 68 percent—over two-thirds—represented direct payments by patients. (Philanthropy accounted for the remainder.) By 1976, direct payments covered only one-third of the total, with private insurance having grown to 26 percent and government 40 percent, all of the government growth having occurred since 1965. Currently, over 90 percent of hospital charges are paid by private and public third-party reimbursement. The advantages of this growth in coverage are obvious. The danger is that the more providers and consumers are isolated from direct financial responsibility for use of resources, the greater will be uses of procedures and devices that are only marginally worthwhile. In the limit, instead of having to live within a budget—a feature of virtually all other facets of life—medical practice may be constrained primarily by the state of the art in science and technology.

Vie[illegible]his light, the profusion of new technology is seen [illegible]n outcome of the medical cost environment, [illegible] simply a producer of it. Private industry has perceived the medical sector as one which is receptive to innovations that will do new things, or old things in new and better ways. The medical economic environment has induced private industry to search for new technologies with only minimal concern for their cost implications. This is consistent with the general observation in economics that in industries in which consumer demand is highly sensitive to market price, technical innovations tend to be cost saving and oriented toward performing existing tasks at less expense. Conversely, in industries in which demand is relatively insensitive to price, technical innovations tend to be quality enhancing. By insulating both consumers and providers from the financial implications of their decisions, third-party reimbursement makes the demand for medical goods and services quite insensitive to their cost.

A manifestation of this economic environment is [illegible]e rapidity of technical change in certain areas of medicine and its consequences. Many technical innovations become scientifically obsolete before they are mechanically obsolete. The rate of scientific obsolescence for many medical technologies is five to eight years, meaning that capital costs must be spread over a shorter period than the machine's mechanical life. An interesting case in point is that of CT scanning. Only a few years old, scanning has already run through several technological generations, and industry is currently developing new capital-intensive diagnostic technologies which promise to make CT scanning obsolete within a few years. Thus the capital cost of a scanner—often in the vicinity of three-quarters of a million dollars—is currently amortized over as little as five or six years, and foreseeable new developments may force a shortening of that period. Concern in the medical community is with modernity and quality (an elusive, ill-defined characteristic). Cost is, at best, a secondary consideration.

The corollary to the economic inducement to orient medical technical change toward the enhancement of quality is that there may be a large reservoir of significant cost-saving medical technical innovations. The lack of demand for such innovations means that relatively little effort has been devoted to searching for them. This theoretical conclusion is supported by industrial observations of a significant untapped market for cost-saving innovations.[8]

The future role of science and technology in containing medical care costs seems to rest fundamentally on the reimbursement and regulatory environment. Assuming that the cost of care remains a national concern, we can anticipate more restrictive reimbursement practices and regulation. Restrictiveness per se, however, may not prove to be as important as the specific nature of reimbursement and regulation. For example, hospital cost containment proposals currently under consideration have features with contradictory effects with regard to the incentive to develop cost-saving capital-intensive equipment: the overall limit on hospital expenditures in the principal hospital cost containment bill would encourage the search for cost-saving technology, while the separately proposed ceiling on capital expenditures would discourage the search for capital-intensive forms of cost-saving technology.

Numerous "perverse" effects of medical technology regulation have been documented; future ones seem likely. In the former category, a prominent study concluded that early CON regulation of hospital bed growth effectively constrained the number of beds but did not contain costs, since resources were poured into increasing the intensity of technology per bed.[9] For the future, the medical device amendments of 1976, coupled with CON, have raised the spectre of a regulatory hassle which will pose a disincentive for all medical technology research and development, and create a "technology lag" like the purported anti-innovation effects of drug regulation. Similarly, decisions by public and private sector insurors not to reimburse for certain technology-based procedures may discourage technology R&D in general, as well as decrease the utilization of specific equipment.

A new feature of the health care scene, the National Center for Health Care Technology (NCHCT), could influence technology development and use within the next few years. The mission of NCHCT is to assess the safety, efficacy, and cost effectiveness of existing, new, and emerging technologies. Although the Center is not itself a direct policymaking organ of government, it will perform or support technology assessments which may serve as the basis for reimbursement—or nonreimbursement—decisions by both public and private third-party payers. Clearly, the activities of NCHCT hold the potential of affecting the amount and nature of health care technology R&D, as well as influencing the diffusion and use of individual technologies.

Independent of the reimbursement and regulatory environments, general scientific and technological trends will influence the shape of the medical technology of the future. The cost implications of new technology will reflect a mixing of scientific and economic factors. An example will serve to clarify this point. The role of microprocessors in medical care is currently very limited, but the potential suggests multiple, important applications in the future. Through miniaturization and the associated portability, and given the low cost of the computing technology, the microprocessor promises to "personalize" medical care. Many patients will be able to get their own microprocessor-based devices which they can wear or carry with them. Already in operation are prototypes of prostheses driven by the wearer's brain, acting in concert with a microcomputer. A simple microprocessor-based machine can continuously monitor a hypertensive person's blood pressure readings, infusing medication during periods when the pressure exceeds predetermined levels. Some years into the future, diabetic individuals may be able to wear (or have implanted) small devices which will continuously monitor glucose levels, infusing insulin as needed. Given the plummeting costs of the basic computer technology, the cost containment potential of microprocessor-based devices appears great. However, whether microprocessor-based medical technology will emphasize quality enhancement or cost reduction, or conceivably a combination of the two, will depend in part on the nature of reimbursement. If extensive third-party coverage continues, with cost- or charge-based reimbursement, device developers will have a strong incentive to harness the power and convenience of the basic technology to perform new medical tasks, with relatively little attention being paid to cost.[10] However, the portability permitted by the microprocessor could be exploited to develop miniaturized versions of existing large equipment which currently necessitates hospitalization, thus decreasing the need for lengthy expensive hospital stays. The technology developers seem likely to respond to market demand.

Certain applications of the new miniaturized computer technology will almost definitely contribute to cost containment. Indeed, the most likely significant near-term use of microcomputers is in management functions in small medical care delivery organizations (solo or small group practices, clinics, etc.). Due to their expense and size, the computers of the 1960s and 1970s have served primarily the larger delivery units (e.g., hospitals). In many of these larger units, patient information systems have been increasing the quality of recordkeeping and usage of data. Over the next several years, refinements in this technology, as well as increasing familiarity and experience with its use, should continue to assist larger delivery units to improve the efficiency of their operations, increasing quality while simultaneously lowering costs. As a result of the new microtechnology, however, the more dramatic development in the 1980s should be the spreading of computerized accounting and patient record management to the smallest delivery unit. This should contribute to greater efficiency, though who will reap the economic rewards (i.e., physicians or consumers) remains to be seen.

Other computer-based technology developments are both "sexier" and more controversial. Robots are appearing in some new hospitals in the form of automated programmed carts which move from one patient room to the next, dispensing prescribed drugs. The cost containment potential of automated materials management delivery systems is currently problematic. As developments such as these ultimately should survive or fail on grounds of efficiency, any adverse effect on costs in the next few years should not persist, whereas any realized cost savings would be expected to grow with technical and operational refinements.

The regulatory and legal environments of health care may introduce some new twists into the development, use, and cost implications of the computer-based medical technology of the future. The computerization of functions currently performed, and judgments made, by medical personnel could have some subtle yet far-reaching implications. For example, if a device were to make a diagnosis or "order" administration of a drug and the patient experience a negative reaction, who would be liable? The patient's physician? The device manufacturer? The programmer? The engineer? Beyond the question of who would be liable is the issue of the nature of the liability. Product liability is much stricter than service liability. Thus, an error associated with a "thinking" device might be subject to greater liability than an error produced (directly) by human judgment. Malpractice implications are obvious, raising the concern that device developers might choose to refrain from working on socially desirable innovations for fear of malpractice liability.

Developments in medical and surgical procedures will undoubtedly continue to have significant influences on the costs of health care. As noted earlier, new drugs may permit the shifting of significant numbers of patients from surgical to medical care, with attendant decreases in hospitalization and other costs. Surgical developments, both mechanical and procedural, will advance the frontiers of medical science and may or may not advance the frontiers of health care expenditures. Will the 1980s witness the emergence of a surgery comparable to coronary artery bypass? This controversial surgery costs hundreds of millions of dollars yearly. It appears to reduce the suffering of angina patients, but not (or at least not clearly) to increase longevity. To many members of the academic and health policy communities, the issues of cost and quality of care raised by this and other procedures recommend increased evaluation, and conceivably regulation, of medical and surgical procedures. Like the debate of a decade ago on medical device regulation—the machinery of which is being put into place—the early 1980s should witness a heated debate on the merits of extending regulation to procedures, the last bastion of medical autonomy.

In contrast with the engineering achievements that have dominated the recent era of medicine's technological revolution, the character and products of basic biomedical research are less clearly a function of reimbursement and regulatory structures. The general perception is that basic research is less subject to the vagaries of the ultimate patient care market than is applied research and development; and the products of basic research are less predictable. New scientific initiatives might produce means of preventing or curing illness, which, like the antibiotics, could contribute significantly to cost containment. An obvious example of a research area with potential is recombinant DNA research which could yield plentiful sources of insulin and other biologically necessary substances, allow correction of genetic defects, and so on. Even though DNA research seems unlikely to significantly affect either health care or health care costs over the next five years, impacts may be felt soon thereafter. Over the long run, the output of basic bioscience research depends on the level of funding and serendipity. If real research resources devoted to basic biomedical science continue to decline as they have for more than a decade, we will witness a further narrowing of one of the nation's promising avenues for health promotion and cost containment.

A second type of research is particularly pertinent to health cost containment and its relationship to science and technology: health services research, the application to substantive health system problems of the skills and perspectives of the social and behavioral sciences and the techniques of statistics and epidemiology. Health services research can elucidate the relationships between technical health care phenomena and their social ramifications. In recent years, researchers have analyzed the determinants of health cost inflation and helped to identify the specific roles of technology in that inflation. The work of this cadre of analysts has served to inform health policymakers. Its importance grows as the health care marketplace deviates further and further from the economic norm and as government's roles in health care financing and regulation expand. The potential usefulness of health services research is also growing as this relatively new field matures. Given all of the health policy initiatives recently begun or under consideration, several of which require the regular input of health services researchers (e.g., the work of the NCHCT), this research could assist in determining cost-effective deployment and utilization of health care technology.

A prognosis for the near-term future is that the cost of medical care will continue to rise at a rate in excess of general prices; overall, technology will continue to contribute to that rate of increase. However, the prognosis is less negative than it was a few years ago. Beefed-up regulatory mechanisms are now beginning to function and major third-party reimbursers appear more willing to refuse to reimburse for certain procedures of undemonstrated efficacy. Perhaps more important, the government's concern with cost containment has gradually diffused into the public consciousness, from big business management to labor unions to county medical societies. This seems likely to diminish the excessive use of technology, but a new national mood does not appear sufficient to redress the scope and magnitude of the problems existing under the current rules of the game.

Under those rules, the best that can be hoped for is that regulatory efforts, research, and heightened cost consciousness will moderate technology's contribution to inflation. Alternative reimbursement or regulatory mechanisms could change this situation quickly, however. For example, a truly effective cap on total hospital expenditures would create immediately a significant demand for cost-saving techniques and technologies. Both theory and the observations of industry suggest that this demand could be satisfied effectively.

SCIENCE AND TECHNOLOGY AND THE NONMEDICAL DETERMINANTS OF HEALTH

"An ounce of prevention is worth a pound of cure." Today's cost consciousness has turned this public health maxim into a question and altered the unit of measurement. Government officials and public health analysts are asking, "Is a dollar of prevention worth sixteen dollars of cure?"

Clearly there is no single answer to this question. Some prevention activities are demonstrably cost effective, while the value of others is suspect. Most prevention activities have not been scrutinized analytically to determine whether or not they are cost effective. Nevertheless, the idea of using prevention as a means of containing health care costs is being firmly implanted in the public's mind due to the efforts of both government and the private sector. A prominent example of the latter is the series of Blue Cross and Blue Shield magazine ads which encourage regular exercise as a means of keeping insurance costs down.

While prevention has long been a popular public health theme, historically both public and private sector resources devoted to prevention have been minimal. In recent years, only 10 percent of Federal health expenditures has been allocated directly to disease prevention and control (outside of medical care organizations), environmental control, consumer safety, and health research. However, the current Federal budget includes significant increases in resources targeted toward such efforts as control of smoking and alcohol abuse, health hazard appraisal, occupational health programs, maternal and child health, and health promotion/disease prevention research. Although the total resources remain small, this budgetary growth is particularly impressive in light of the Administration's generally overriding concern for balancing the Federal budget, with the associated implications for spending on discretionary programs. Publicity about the government's new prevention initiative has accompanied the budgetary commitments.[11]

Low-cost prevention of illness holds considerable appeal as an approach to containing the costs of health care. While the inclination is to assume that the payoff to prevention activities generally lags the activities by many years, there are obvious cases where this is not true—for example, immunizations—and evidence of rapid returns even where the assumption seems most reasonable, as with efforts to change behavior. For example, following decades of steady increases, heart disease mortality has been falling over the past decade. The same period has witnessed increases in exercising and decreases in ingestion of cigarette tar and nicotine and dietary cholesterol. Causation has not been established, but the relationship is certainly suggestive and encouraging.

In this section of the paper I will note examples of both past and potential cost-containing applications of science and technology to each of the three nonmedical classes of health determinants identified at the beginning of the paper.

GENETIC FACTORS

Over the millenia, genetic factors have served as a natural means for strengthening the species. However, the growth of technical knowledge and material wealth has made us more able and willing to combat the individual ravages of genetics. Medical science can correct many genetic defects and material and spiritual affluence have been directed toward building social support systems to address the needs of those for whom defects cannot be corrected or compensated.

Genetic defects impose at least two major costs on society: the economic burden of maintaining and supporting those in need of help and the emotional costs borne by the families of individuals with incapacitating or terminal defects. Other than through medical correction, the only effective means of decreasing such costs is to prevent the genetic defects from occurring. To date, this has resulted primarily through the counselling of potential parents. The knowledge base for such counselling has come from research.

Recently, a technology has emerged to increase prospective parents' information on the "genetic health" of the fetus. Amniocentesis permits the early identification of Down's syndrome and other genetic problems. For parents who would consider abortion, the technique provides very useful information which can contribute to the containment of long-run social and emotional costs.

The future role of genetic science and technology most likely will differ from the present one. To date, science and technology have contributed to increased knowledge and improved prediction and diagnosis. At the frontier of biological science is investigation of genetic manipulation: means of repairing genetic defects in the fetus and altering genetic composition prior to conception. The prospects are both thrilling and terrifying. Genetic repair could eliminate or significantly reduce the costs of genetic defects without requiring destruction of the fetus, which is itself emotionally costly. In the extreme, genetic manipulation could prove to be the most effective means of containing health care costs: scientists could produce the proverbial super-race, with built-in immunities to the majority of afflictions known to human-kind. Health care costs would melt away as illnesses were engineered out of existence. Needless to say, the social costs of a manufactured super-race might vastly exceed the derivative savings in health care costs. Science and technology in this area will not have a significant impact in the near-term future—the five-year outlook for health care costs will not be affected by genetic engineering—but ensuing five-year outlooks will have to pay increasing attention to this area of science and technology.

INDIVIDUAL BEHAVIORS

The influence of individual behaviors on health has received much attention in recent years, which has resulted in attitudinal and behavioral changes regarding diet, exercise, smoking, and so on. Science has contributed to knowledge of the health implications of behavior and hence presumably to much of the behavioral change. To the extent that the changes are truly significant, and to the extent that such changes contribute to improvements in health, science should be credited with containing health care costs.

Attempts to modify behavior, particularly through publicity efforts (e.g., the anti-smoking campaign), have their detractors, the argument being that simple persuasion rarely produces much behavioral change. However, a lot of behavioral change may not be necessary to justify the investment of effort in persuasion. In terms of resources consumed, that investment can be very small. Consequently, a small result in behavioral change may more than pay back the investment.

The principal past and near-future scientific contributions to health cost containment due to behavioral changes are likely to be concentrated in two areas of research: the relationships between behaviors and health effects and the determination of cost-effective means of disseminating findings in a persuasive manner. The amount of health-promoting behavioral change which can be induced might prove to be small, but, again, the rate of return to the very low research expense could be considerable.

Science and technology have contributed to health behavior change in more mechanistic ways. A large share of smoking research dollars has been devoted to the search for less hazardous cigarettes. Documented decreases in tar and nicotine intake and recent discoveries concerning physiological changes in smokers suggest that this effort has been useful, combined with the publicity which has encouraged the shift to lower tar and nicotine cigarettes. Again, the research cost has been relatively small and the potential cost containment substantial.

Mechanical apparatus has contributed to behavior change. Stationary bicycles and other devices have been designed to increase the convenience and enjoyment of exercise and thereby to capture a portion of the "exercise market." To date, the health promotion and cost containment implications of all such equipment remain unknown.

ENVIRONMENTAL INFLUENCES

Many of the major victories against human illness have come from improvements in the quality of the environment. Early sanitation efforts contributed immensely to the health of people in the developed countries, and similar efforts are now taking place in or being explored for developing countries. In our own nation, recent concern with the health effects of pollution has produced regulation and control of the private use of air and water to dispose of wastes. Regulation has created a demand for means of achieving cost-effective disposal of pollutants. This demand has translated into the growth of a flourishing pollution control industry.

The health cost containment implications of pollution control are a function of (1) the relationship between reductions in pollution and improvements in health, (2) the magnitude and nature of reductions required by law (which affects the preceding point), and (3) the efficiency or cost effectiveness of pollution control technology. In the public policy arena, environmentalists and industry are debating the appropriateness of regulation and its extent, given the cost to industry and hence the consumer. This debate serves to emphasize the need to investigate carefully both the means of achieving pollution abatement and the amount which should be achieved. For example, there is some evidence that strict automobile pollution abatement requirements may be costing society more than they are worth, while controlling effluents from industrial smokestacks has been a cost-effective activity.[12] In both cases, improvements in pollution abatement technology hold the potential of contributing to more efficient abatement and hence health cost containment.

The environment can be a source of transmittal of a health-endangering substance, as in the case of pollution, or it can be a means of introducing a health-promoting substance, as in the case of fluoridation of municipal water supplies. Studies of alternative approaches to improving dental health invariably conclude that fluoridating the water supply is by far the most cost-effective technique. Despite this, many water supplies have remained unfluoridated. Over the next several years, progress in decreasing this number will directly contribute to containing the costs of dental care.

Several issues will be on the frontier of environmental public health concern for the near-term future. Pollution control—how much, what type, and how to achieve it—should remain of interest for a long time to come. This issue is becoming even more challenging in the face of the energy crisis, itself a major environmental issue with potentially profound implications for health and health cost containment. In the case of both coal and nuclear power, we confront a trade-off between the health risks associated with producing energy and the health (and other) costs associated with a signficant decrease in the supply of energy. Scientific and technological efforts to develop new "clean" sources of abundant supplies of energy (e.g., solar power; fusion) are directed toward resolving this dilemma. Unfortunately, economically viable solutions seem a couple of decades in the future.

A similar problem currently at the frontier of public policy is the management of toxic substances, in consumer products (e.g., saccharin) and at the workplace (e.g., asbestos). Policy initiatives in the next few years will affect the shape of the economy and its health consequences. The issue is how far, and at what cost, we will go to reduce a health risk (and, usually implicitly, the health costs associated with the risk). In some instances, some degree of risk reduction seems clearly desirable and consistent with the objective of health cost containment—for example, the risk exposure of miners and asbestos workers—but, again, the questions are which risks to reduce, how much, and how? Scientific and policy studies can help to resolve these questions, and technological developments can improve the efficiency with which we reduce risks.

As a final example of a current and near-future "environmental" problem, consider the toll associated with motor vehicle accidents. Not only do such accidents account for a disproportionate amount of disability and death among the young, they often produce particularly expensive disabilities requiring long and costly hospital stays and frequently involving substantial long-term costs of care after the patient leaves the hospital. The principal conscious approach to this problem has been to impose safety requirements on the manufacture of automobiles. Motorcycle helmet laws are also directed toward providing the driver with physical protection. The nationwide reduction of the maximum highway speed limit to 55 miles per hour, originally intended to conserve energy, has reduced serious accidents and as such has contributed to health cost containment. Other behavioral alternatives include stricter enforcement of speed limits and mandatory use of seatbelts. Technological alternatives include additional car safety features (e.g., the much-debated air bag) and new efforts in building safer highways (e.g., road surfaces with greater traction in ice and rain; breakaway signs; flexible shoulder barriers). Here, as with all the environmental measures, we need vast increases in our knowledge of both the technical features of alternative approaches and the health and other cost implications.

At the outset of this paper, I subdivided the class of environmental influences on health into the physical, social, and political and economic. Discussion in this section has concentrated on the physical environment, reflecting the present policy emphasis and the problems most readily (or clearly) susceptible to scientific and technological solution. In a treatment of environmental influences on health, however, it would be negligent to fail to consider the somewhat less tangible but nevertheless significant social, political, and economic environmental impacts on health. For example, it is argued that deterioration of certain aspects of the social environment has contributed to mental illness: the decrease in the primacy of the family unit, the associated unavailability of personal support systems (relatives, religious leaders), crowding and the associated lack of privacy. Both social and physical science need to be applied to develop a basic understanding of the causes and magnitudes of health problems produced by social factors, but the complexity of these relationships makes it unlikely that near-future "scientific breakthroughs" have the potential to lessen the problems significantly. Nevertheless, the long-term interests of social welfare (and of health cost containment as a component thereof) recommend devoting considerable research attention to this area.[13]

Certain aspects of political and economic influences on health have been much studied. For example, students of economic development have noted frequently that both health and development are associated with a more egalitarian distribution of income and that the development process contributes to improvements in health quite independently of, and apparently more effectively than, direct personal health services. More recently, a few scholars have begun to devote attention to the influences of political and economic factors on health in industrialized countries. The work of Harvey Brenner and others has uncovered adverse health consequences from macroeconomic fluctuations.[14] Some of these are neither new nor surprising—e.g., unemployment has a large health toll—but others are: for example, it appears that *any* macroeconomic change has short-run health costs, including periods of rapid economic growth. Similarly, feelings of political involvement or powerlessness can translate into invigorated or diminished health, respectively.

As in the case of the health influences of social conditions, the health consequences of political and economic systems and system changes are too poorly understood to consider these as areas in which definitive health cost containment can derive from near-term policy changes. Nonetheless, these appear to be areas warranting research attention, with implications possibly becoming well understood and relevant for debates on future policymaking.

CONCLUSION

Rising health care costs are likely to be with us for a long time. To avoid obsession with the negative consequences of this phenomenon, it is imperative to keep in mind that not all cost increases are undesirable. Some reflect genuine improvements in the product produced by the medical care system. Rising health care expenditures also reflect the fact that wealthier societies wish and can afford to spend proportionately more on medical care services. The goal of health cost containment should be to decrease the unnecessary or less effective expenditures—those on the marginal lab test, the unnecessary surgery, the inappropriate physician visit.

The apparent intractability of health cost inflation results in part from the inherent tension between the competing desirable goals of promoting access to medical care, particularly for the poor and elderly, and constraining the rise in costs of care. Attainment of the former—the goal of the past decade—must bear some of the responsibility for making the latter a problem. Thus the challenge is to contain costs without jeopardizing the improvements in access which have contributed to the cost inflation.

A major policy issue is whether Congress will ever be willing to bite the bullet of true medical system cost containment. The measures passed to date are BBs, generally aimed at individual technology or facility acquisition decisions. These policies irritate the medical care system but seem unlikely to affect medical cost inflation significantly. Given the huge sums of money invested in the acquisition and utilization of individual technologies and procedures, a single BB might restrict the expenditure of tens or even hundreds of millions of dollars—definitely a worthwhile result—but the percentage impact on systemwide cost inflation would not be substantial. The aggregate effect might prove to be a discernible dent in the rate of inflation, but equally possible is no effect, or conceivably a negative impact as a result of bureaucratic and allocative inefficiencies induced by piecemeal regulation.

A second issue, closely related to the first, is what type of serious bullet legislators might eventually be willing to bite, should they show a genuine interest in systemwide control. Experience outside of the health arena as well as within it suggests that legislators will not consider seriously major systematic changes devoid of bureaucratic regulatory control. For example, the growth of Health Maintenance Organizations and voucher funding of national health insurance might go a long way toward solving the cost inflation problem while encouraging high-quality care, but the likelihood that Congress would adopt such an approach is small.[15] Even the most systematic cost containment proposals to receive serious legislative consideration must be classified as partial and seem doomed to achieve only a slight moderation of health cost inflation. The Administration's hospital cost containment bill, although appropriate in spirit, has been watered down for political reasons and fails to plug leakages which might translate hospital cost containment into ambulatory cost inflation. In other words, certain costs would "leak" outside of the hospitals and into other, unregulated delivery settings.[16]

In short, barring an unanticipated radical change in the basic reimbursement rules of the game, medical cost inflation seems likely to persist. The piecemeal regulatory approach may moderate certain cost increases, though it is difficult to predict at what price in terms of overall efficiency and quality of care. At present, the general societal atmosphere and the threat of severe regulation are probably the Administration's most effective allies in its campaign to contain costs.

As the second section of this paper emphasized, the role of science and technology in medical care has often been antithetical to the cause of cost containment. But the nature of medical scientific and technological change is simply a response to the unorthodox economic environment of the provision of medical care. Science and technology could be more effectively harnessed to moderate medical costs if the reimbursement system provided incentives to seek cost-saving means of delivering care. Currently it does not and no major reforms seem likely in the next few years, with the possible exception of some form of limited hospital cost containment.

The contribution of science and technology to medical cost inflation may decrease somewhat if piecemeal regulatory mechanisms deter the development and acquisition of technology (e.g., CON; FDA medical device regulation). The danger is that regulation will produce undesirable side effects, for example, a "technology lag" similar to the drug lag attributed to FDA regulation of new drugs.

The problem of health cost inflation, and the possibilities for its containment, take on a new dimension when one distinguishes "health care" from "medical care." Diligent use of nonmedical approaches to health care has the potential of simultaneously serving the goals of health promotion and cost containment. As a monolithic concept, "prevention" can be oversold, but certain prevention activities are almost certainly worthwhile and others appear promising. The need is to identify those which are cost effective.

Science and technology appear to have made their greatest contributions to health by affecting its nonmedical determinants. The provision of clean water supplies, sanitary disposal of human and animal wastes, and public education are all activities which are so basic and pervasive in our society that we take them for granted. Yet they have contributed immensely to the long and healthy lives which citizens of the developed world have come to expect. Of late, the medical care "cost crisis" has begun to focus attention on the cost containment potential of these and other nonmedical influences on health. Surprisingly little is known about the cost containment implications of disease prevention strategies, but certain generalizations can be drawn from the available evidence and simple logic:

- Several inexpensive prevention activities eliminate costs that society would bear in their absence. A prominent example is water fluoridation.

- Certain highly effective prevention activities are responsible for increasing medical care costs, though in a manner virtually everyone would find desirable. Activities which reduce premature adult mortality contribute to a population of the elderly whose medical care needs exceed those of the young. Thus, an aging population almost necessarily increases medical care costs. This is a contributor to the medical care cost inflation problem today, one that will increase as the population ages further. Nevertheless, it is the kind of problem that we should welcome if it reflects an improvement in the quality as well as length of life.

- As many prevention activities are quite inexpensive, their health-enhancing or cost-containing potential need not be great in order to warrant their application. An example is the anti-smoking publicity and the apparent associated decrease in tar and nicotine consumption. Similarly, recent (inexpensive) emphasis on the link between individual behavior and cardiovascular health has changed the behavior of Americans with respect to exercise and diet, changes that may be contributing to the decreases in cardiovascular disease and mortality which we are now experiencing.

- Certain relatively recent public health prevention activities appear to be returning new health dividends, with an implication of health cost containment potential. Industrial pollution control falls into this category.

The evidence suggests that both tried-and-true prevention techniques and novel attempts to alter behavior, the environment, and even genetic makeup, warrant attention as mechanisms to contain health costs, as well as to promote health. Science and technology have played central roles in many prevention efforts and their potential for the future seems considerable. Particular areas of scientific investigation that appear attractive are:

1. Biomedical research (basic research, like that supported by the National Institutes of Health, to develop knowledge of these processes).

2. Epidemiological research (population-based studies to assess factors associated with the in-

cidence and prevalence of important disease patterns).

3. Health services research (economic and other social and behavioral science analysis to evaluate the social implications of disease prevention and treatment efforts).

The first two areas can contribute understanding of basic disease processes and the health effects of various interventions. The third area translates "production relationships" (the technical cause and effect linkage) into social implications and values.

In the technology arena, continued developments in pollution control, reduction of occupational exposure to toxic substances, increased nuclear safety and so on hold health promotion and cost containment potential. Two particularly important developments relate to computer technology. One is exploration of applications of the new microprocessor-based technology to disease prevention. The other is work on software systems to improve health data collection, surveillance, and analysis. An example of how the latter might be combined usefully with scientific investigation is in the study of environmental carcinogenesis. Estimates of environmentally induced cancers range up to 90 percent of all cancers. Knowledge of cancer causation might be converted into cancer prevention. Acquiring the data necessary to develop understanding and analyzing them scientifically represent mammoth tasks, but they could have a tremendous payoff in the battles against cancer and health cost inflation.

The exciting potential of prevention efforts relates to their potential for significant rates of return. That is, investments in many prevention activities can be quite small and returns comparatively quite large. However, the distribution of health care costs among medical care and nonmedical activities is so heavily skewed toward the former that nonmedical prevention interventions seem most unlikely to resolve the cost inflation dilemma. Rather, they represent an important component of a package of prevention and medical system regulation which, if pursued conscientiously and vigorously, can moderate health cost inflation. The overall prognosis is not entirely encouraging, but neither is the patient's condition terminal.

NOTES

1. In real terms—that is, the amount of research which the budget will support—the NIH budget has been declining in recent years. This is particularly true of basic bioscience research.

2. A recent empirical study concluded that, from 1930 to 1975, biomedical advances accounted for a small decrease in medical expenditures (Mushkin, Selma, et al., "Returns to Biomedical Research, 1900-1975. An Initial Assessment of Impacts on Health Expenditures," Report #A1 (Revised), Public Services Laboratory, Georgetown University, December 1977). This finding is not consistent with those of several earlier studies which found positive aggregate effects of technology on expenditures, but given the methodology of most of these studies—in which "technology" or "biomedical advance" is treated as a residual once other influences are taken into account—none of the results should be considered definitive. A more important finding of Mushkin et al. was that biomedical advances were cost-containing for certain diseases (e.g., infective and parasitic diseases) and cost-increasing for others (e.g., cancer and circulatory diseases). As the authors summarize their findings (p. 15), "Those innovations that resulted in a substantial decrease in the incidence of the disease itself appear to be cost-reducing while those that resulted in improved techniques of treatment of the disease seem to be cost-raising."

3. U.S. Congress, Office of Technology Assessment, *Selected Federal Vaccine Immunization Policies: Case Studies of Pneumococcal Vaccine*, 1979.

4. The cimetidine case illustrates the complexity of the mechanism for a seemingly cost-saving innovation to produce significant cost savings. While preliminary analysis indicates a substantial decrease in surgical care of duodenal ulcer patients and in their hospital length of stay, the number of hospital admissions has not decreased, though cimetidine can be administered on an outpatient basis. In addition, there appears to have been a small but perplexing increase in perforated and bleeding duodenal ulcers. The drug has been used for half of all patients with gastric ulcers, despite the Food and Drug Administration's failure to indicate the appropriateness for this use. And the drug has been used for long-term use since its introduction, though the FDA did not approve other than short-term use until the summer of 1979 (Wylie, C. Murray, "Duodenal Ulcer Before and After Cimetidine: Time Trends in Hospital Care," paper submitted for publication; also private communication with Dr. Wylie).

5. In a nonrepresentative sample of Boston area hospitals, Abt Associates found that purchases of major movable equipment accounted for less than 5 percent of hospital costs. However, total equipment expenditures per day rose 23 percent per year from the mid-1960s to the mid-1970s. Thus, the contribution of equipment acquisition to hospital cost inflation was quite small but growing. Abt concluded that "if major movable . . . equipment constitutes an important source of hospital cost inflation, it must be because of the complementary inputs required to operate and service it." (Abt Associates, Inc., *Incentives and Decisions Underlying Hospitals' Adoption and Utilization of Major Capital Equipment*, prepared for the National Center for Health Services Research and Development, Contract No. HSM 110-73-513, September 1975.) Other studies find operating costs running from one to three or more times capital costs.

6. Scitovsky, Anne, and Nelda McCall, "Changes in the Costs of Treatment of Selected Illnesses, 1951-1964-1971," NCHSR Research Digest Series, DHEW Pub. No. (HRA) 77-3161, 1976.

7. A typical failure to fully comprehend the role and implications of third-party reimbursement is represented in the propensity of some critics to blame physicians for overuse of certain procedures. The assumption that physicians should perform the appropriate social benefit-cost calculation puts them in an untenable position: to represent society's interests, they may be asked to deny a patient a procedure which might benefit the patient and in any case would not harm the patient economically. This violates the medical ethic of representing the patient's best interests. Were the patient responsible for paying the cost of the procedure, then the physician's benefit-cost calculation should properly weigh the patient's economic sacrifice against the potential for medical benefit.

8. I discussed this theoretical conclusion in "Effects of Hospital Cost Containment on the Development and Use of Medical Technology," *Milbank Memorial Fund Quarterly/Health and Society 56* (Spring 1978), pp. 187-211. Becton-Dickinson, a major hospital sup-

plier, has expressed a similar belief (as quoted in *Blue Sheet, the Drug Research Reports*, January 11, 1978).

9. Salkever, David, and Thomas Bice, *Hospital Certificate-of-Need Controls: Impact on Investment, Costs, and Use*, Washington, D.C.: American Enterprise Institute for Public Policy Research, 1979.

10. As an NIH engineering scientist responded when asked to estimate the cost of a certain individualized monitoring device, "Cost? I could build you one for $200 or one for $20,000." The microcomputing hardware is cheap. A major cost of microprocessor-based technologies will be the software, and that cost can vary considerably according to the degree of sophistication desired, number of functions, and so on.

11. In the summer of 1979, the Surgeon General released a lengthy report on preventing disease and promoting health. Its title is *Healthy People: The Surgeon General's Report on Health Promotion and Disease Prevention.*

12. *The work of Lester Lave and Eugene Seskin (e.g., Air Pollution and Human Health*, Baltimore: Johns Hopkins University Press, 1977) is the principal source of this assessment. It should be noted that this work has been criticized on technical grounds.

13. A large proportion of visits to physicians appears to be motivated by the need for emotional or psychological support. In years gone by, such support was more readily available within the home (e.g., the supportive live-in grandparent) or other social institutions (e.g., the minister, priest, or rabbi). The deterioration of family units and of religious involvement may have caused emotional problems; in any case, it has decreased the availability of personal solutions to the need for support. Increasing reliance on the medical system accompanies decreasing preparation of physicians to deal with emotional and psychological problems. Due to the rapidly growing body of technical knowledge, emphasis on the science rather than art of medicine, professional prestige and income considerations, physicians have become increasingly specialized, trained to deal with specific organ systems and not with the social and emotional problems of the whole person. Thus, a significant contributor to the high cost of medical care is the often ineffective use of expensive physician time. Development of less expensive and more appropriately trained support mechanisms, either within or outside of the medical care setting, holds the potential of making a significant contribution to health cost containment. Within medicine, generalist physician assistants and substitutes can play this role effectively.

14. See, for example, U.S. Congress, Joint Economic Committee, *Achieving the Goals of the Employment Act—Thirtieth Anniversary Review*, Vol. 1, Paper No. 5 ("Estimating the Social Costs of National Economic Policy: Implications for Mental and Physical Health, and Criminal Aggression"), Washington, D.C.: U.S. GPO, 1976.

15. A specific proposal was prepared for DHEW Secretary Califano in 1977 by Alain Enthoven of Stanford University. It is described in Enthoven, "Consumer-Choice Health Plan," *New England Journal of Medicine* 298 (March 23, 1978), pp. 650-658, and (March 30, 1978) pp. 709-720. HMOs have received some Federal support, but the maze of requirements and regulations in the legislation may have been quite counterproductive. For an interesting discussion of this, see Paul Starr, "The Undelivered Health System," *Public Interest* 42: 66-85, 1976.

16. A striking example of this phenomenon is related to the development of Medical Service Plans (MSPs), private physician group practices which contract with hospitals to staff a service and then charge their own fee for service, independent of the hospital. Physician reimbursement thus leaks out of the hospital. It could be removed from the hospital's expenditure base in order to determine allowable increases, but the physician service component of the hospital stay would not be subject to the ceiling and could inflate at any rate the MSPs chose and the market would bear.

4 Enhancing the Contributions of Science and Technology in Environmental, Health, and Safety Regulations

*by Eugene P. Seskin and Lester B. Lave**

SUMMARY

The difficulties in formulating sound environmental, health, and safety (EHS) regulations flow from the lack of well-defined goals, the absence of agreement on the best methods of achieving objectives, the costliness of proposed solutions, and the underlying scientific, social, and legal complexities. The contributions of science and technology can lead, in part, to the calming of a debate that has become increasingly shrill and, at times, irrational. While ultimately the goals of EHS regulations must emanate from society's value judgments, learning how to achieve chosen objectives in an efficient manner is extremely important.

In this paper we explore the complexities of the EHS regulatory process with special attention focused on how science and technology can foster improvements in it. Dealing with the entire subject is not possible in a relatively short paper such as this one; hence, we have chosen to emphasize aspects concerning health. We begin by describing briefly the current regulatory environment: EHS regulations have large costs and benefits, which are often difficult to evaluate. We then turn to the major goals of regulation and follow this with an examination of the form and timing of regulations.

Next, we shift the emphasis somewhat by discussing informational needs for improving EHS regulations. First, we describe difficulties with estimating the costs of such regulations and ways in which these might be mitigated. We then turn to the more difficult issue of estimating the benefits of EHS regulations. Here, two basic approaches are mentioned: experimental and statistical.

We devote a major portion of the paper to issues concerning the estimation of health benefits from EHS regulations. Specifically, we discuss the strengths and weaknesses of two basic approaches: laboratory experimentation (including animal bioassays and short-term techniques using microorganisms), and epidemiology (broadly defined). We note that contributions are needed in the development of research methods (including appropriate statistical techniques), in determining the value of analyzing additional data, in measurement techniques (such as generating better estimates of exposure and dose), and in information gathering (including augmenting existing data and collecting new data); all can provide a more adequate basis for EHS regulations. Finally, we discuss the issue of monetizing estimated health effects and point to needed research on such factors as the psychology of risk bearing.

* Dr. Seskin was Senior Associate, Research, Resources for the Future, Present Address: Bureau of Economic Analysis, Department of Commerce. Dr. Lave is a Senior Fellow at the Brookings Institution.

Brief evaluation is then given to institutional arrangements required to create an environment suitable for carrying out the needed research in these areas. In particular, we note the importance of promoting interaction between disciplines and of adopting a long-term view of the problems.

INTRODUCTION

In the 1960s, Americans recognized increasingly that the environment was in a deplorable state and getting worse and that new risks to their health and safety were being discovered constantly. The institutional framework for protecting the environment, health, and safety was deemed a failure and a new approach was initiated. This new approach was a radical departure from the ways in which the United States had previously regulated the economic actions of firms.[1] It consisted of broad government regulation whose goal was "simply" to protect the EHS concerns of all Americans with little attention to resource requirements or scientific foundation. While the tools of regulation varied, the regulations themselves were often quite detailed, consisting of design and performance standards for individual machines, products, processes, and services deemed to affect the environment, health, and safety.[2] A host of new agencies were created to administer these regulations,[3] and vast resources were devoted to rule-making, litigation, and compliance. Since health and safety are essential to the enjoyment of life, the legislation creating these agencies enjoined them to protect people and the environment with little regard to the feasibility or cost of doing so.

From the vantage point of mid-1979, an evaluation of EHS legislation that has been passed and the subsequent regulations that have been promulgated results in contradictory conclusions. On the one hand, EHS regulations enunciated noble goals, represented major social experiments, and were successful in lowering some risks and enhancing some aspects of environmental protection. On the other hand, EHS regulations have not lived up to expectations, have had costs running into the billions of dollars annually, and are seen as requiring significant change if their goals are to be accomplished. An examination of the background legislation makes it evident that the mechanism chosen by the Congress for attaining EHS goals was basically judicial in nature, with a preoccupation for legal processes and remedies. At the same time, this legislation evidences naiveté about scientific analysis and technology, the latter often entering the picture only to determine if some action were feasible. Even here, the Congress has grown impatient at times and required that something be accomplished, even if it were not feasible with known technology.[4]

One reason for dissatisfaction with EHS regulations concerns the fact that while we are currently experiencing the costs of such regulations, many of the associated benefits will not be fully realized for some time. Before the 1960s, society bore the untoward consequences (social costs) of an economy relatively free of EHS regulations. By the mid-1980s, we should live in a safer, less polluted environment, although it will be one with more costly goods and services. During the 1970s and early 1980s, EHS regulatory costs have begun and will continue to have measurable cost impacts, yet many of the untoward consequences will persist until the mid-1980s. Thus, we are in the midst of a period when we must bear both types of costs.

The costs and slow progress have led to a polarization.[5] The regulations are defended by consumer advocates and environmentalists who think even more should have been accomplished using more stringent regulations. They attack critics of such regulations, calling them selfish, profit seeking, and contemptuous of the public interest. At the same time, the regulations are under fire by business groups who see them as excessively costly, capricious, violating privacy, inhibiting small business, and inefficient. Rather than working together for better regulations, the two groups are often pitted against one another questioning both the intelligence and motives of their opponents. If allowed to spread, this polarization will breed cynicism and weaken our social fabric.

The contributions of science and technology can change the current situation in which each side tries to shout louder than the other, exaggerating its claims. Although values are fundamentally at issue, science and technology could serve an increasingly important role in defining what EHS effects would be mitigated by proposed actions, what EHS objectives can be attained, and the most efficient means of reaching those objectives. That is, science and technology have the potential for changing the nature of the exchange; science and technology will not tell us how safe or cautious we ought to be, but they will serve to clarify the costs and risks associated with proposed actions. They will show us how to lower existing risks and how we may create other risks if we are not careful.

In regulating the environment, health, and safety, principal uncertainties and areas of ignorance concern the costs of such regulations, the benefits of the regulations, and the best means of accomplishing objectives. Our overall objective in this paper is an assessment of the potential contributions of science and technology in lessening these uncertainties. This topic is too vast to be covered in a relatively short paper; we cannot hope to be definitive. Instead, we have elected to pick and choose among the issues, with a resulting emphasis on aspects concerning health.

Although we stress the problems associated with current EHS regulations, since these are areas in which science and technology can make important contributions, we recognize the critical need for such regulations: while economic development has contributed greatly to our standard of living, it has also created problems for the environment, health, and safety. Most people would agree that it would be a bad bargain to have command over a large amount of goods and services, but have a forbidding environment, be disease-ridden, and face high risks. Along with improvements in our economic well-being, we desire enhancement of our environment, health, and safety. However, it is not obvious how EHS regulations can be formulated to provide this enhancement while at the same time preserving our level of private consumption. This paper is focused, in part, on ways in which this might be accomplished. Our basic answer is that science and technology provide two crucial inputs: they can be used to show how each goal can be accomplished at lowest social cost and they can be used to quantify the tradeoffs among goals. In short, they can provide the information necessary to accomplish society's objectives.

But there is a great intellectual gap to be bridged before arriving at these conclusions. We begin with an examination of the current costs of EHS regulations.

THE COSTS OF EHS REGULATIONS

No one is certain how much the myriad new regulations are costing us or whether the benefits of the regulations exceed their costs. The costs, however, are highly visible and have been the focus of one line of attack on EHS regulations. Using very crude techniques, DeFina estimated that the administrative costs of regulation in 1976 were $3.2 billion and that compliance costs were $62.3 billion for a total of $65.5 billion, about 30 percent of which is due to EHS regulations.[6] A government estimate of the total costs of regulation suggests that they are more on the order of 1 percent of gross national product—about $14.9 billion in 1977.[7] Still other figures relate only to the Federal Government budgetary expenditures on administering regulations. For example, table 1 provides a summary of such costs (by function) for the fiscal years 1974 through 1979. EHS regulations appear to be responsible for most of the total Federal regulatory expenditures. Finally, Denison has estimated that by 1975, the last year for which he provided figures, output per unit of input in the nonresidential business sector of the economy was 1.8 percent smaller than it would have been if business had operated under 1967 conditions—1.0 percent was ascribable to

Table 1. Expenditures on Federal Regulatory Activities, by Fiscal Year, in millions of dollars.

Area of Regulation	1974	1975	1976	Annual Increase	1977	Annual Increase	1978	Annual Increase	1979	Annual Increase	Increase 1974-79
Consumer Safety Health	1,302	1,463	1,613	10%	1,985	23%	2,582	30%	2,671	3%	105%
Job Safety and Other Working Conditions	310	379	446	18	492	10	562	14	626	11	102
Environment and Energy	347	527	682	29	870	28	989	14	1,116	13	222
Financial Reporting and Other Financial	36	45	53	18	58	9	70	21	69	−5	92
Industry-Specific Regulation	245	269	270	—	309	14	340	10	341	—	39
	2,240	2,683	3,064	14%	3,714	21%	4,543	22%	4,823	6%	115%

PERCENT DISTRIBUTION OF FEDERAL REGULATORY EXPENDITURES

Consumer Safety and Health	56%
Job Safety and Other Working Conditions	13
Environment and Energy	23
Financial Reporting and Other Financial	1
Industry-Specific Regulation	7
	100%

Source: R. DeFina and M. L. Weidenbaum, "The Taxpayer and Government Regulations," St. Louis, Center for the Study of American Business, Washington University, March 1978, p. 3.

pollution abatement, 0.4 percent to employee health and safety programs, and 0.4 percent to the increase in dishonesty and crime.[8] Except for the budget data, these estimates are little more than guesses, based on scant evidence. But if they are in the right "ballpark," they imply that a substantial portion of the growth in U.S. economic productivity has gone to "finance" EHS regulations and that measured real income will rise in the future less rapidly than in the past.

THE BENEFITS OF EHS REGULATIONS

It must also be remembered that the above estimates represent gross, rather than net, costs of regulation. That is, they do not include estimates of the offsetting gains, or benefits, attributable to the regulations. Unfortunately, relatively little attention has been paid to this side of the equation. In part, this is because the benefits of such outcomes as cleaner air and safer products are deemed to be self-evident. However, there is also the analytic problem that the benefits are more difficult to measure than the costs of regulation. Costs usually take the form of items such as expenditures on labor and materials; benefits are less tangible since they involve such factors as health and environmental improvement. Furthermore, the benefits are frequently spread among a large population. For example, it is relatively easy to ascertain the costs to a steel plant of installing pollution control equipment, but it is much more difficult to quantify—let alone express in dollars—the decreases in medical care for the surrounding population because the air is cleaner.

Yet, the potential benefits from EHS regulations are large and real. For example, the President's Council on Environmental Quality (CEQ), in releasing its latest annual report, estimated that Federal air pollution standards alone were saving the nation about $22 billion a year in averted damages.[9] Elsewhere, we have made similar estimates of the health benefits that would result from significant reductions in the ambient concentrations of specific air pollutants.[10] Thus, whether estimated by economic methods or inferred from the political process, the benefits associated with EHS improvements are significant and worthy of paying large costs.

BALANCING THE COSTS AND THE BENEFITS

It is inconceivable that EHS regulations of the sort promulgated during the last 15 years could be implemented without inefficiencies, waste, mistakes, and intense controversy. Furthermore, while public opinion polls make it evident that society desires effective EHS regulations, there is also an indication of concern for the huge costs of regulation and the impacts on the economy.[11] Thus, in trying to achieve the goals of regulation and at the same time maintain a consciousness of the associated costs, the regulatory agencies face the extraordinarily difficult task of balancing protection against costs. Here, science and technology can be both a help and a hindrance.

Advances in science and technology have brought about a situation in which over 50,000 synthetic chemicals are produced and used in significant quantities, and close to 1,000 new chemicals are introduced each year.[12] Of the relatively small number tested, at least 25 of these have been shown to cause cancer or are suspected of causing cancer in humans; in laboratory tests several hundred have been shown to cause cancer in animals.[13] A much larger number has been shown to be toxic. At the same time, developments in the sciences of detection have enabled us to isolate substances in concentrations of less than one part per billion. However, here again, we have not always been pleased to discern these substances. For example, the dioxin TCDD, a contaminant of the notorious Agent Orange defoliant used in Vietnam and one of the most toxic substances known, has been shown to cause birth defects and assorted tumors in test animals when fed in concentrations below 100 parts per trillion. (There is also limited evidence of similar effects in humans.) A recent report claims to have isolated TCDD in soil samples around a chemical plant in concentrations up to 100,000 parts per trillion.[14] Our objective must be to direct new and sophisticated science and technology not only to discover the hazards that surround us, but also to solve the problems these hazards create. In particular, we must address the question: How can science and technology improve EHS regulations? Before tackling this question directly, we shall examine the goals of such regulations, their timing, and their form.

GOALS OF THE REGULATORY PROCESS

A great many methods could be used to achieve the objectives of a cleaner, safer environment. The choice is influenced by a multiplicity of goals, which are broad, complex, and often incompatible. One important objective of most governmental activities is economic efficiency. A simplified definition of economic efficiency is achieving some objective with the smallest expenditure of resources. Under certain assumptions it can be shown that a competitive market system will achieve such an optimal allocation of resources. General propositions have been derived about how efficiency can be achieved; these are usually stated in terms of the equality of small changes in the costs, inputs, and outputs.

A second important objective of governmental activities is equity in the distribution of the "proceeds" from the economic system. No governmental action takes place without some consideration of the distribution of costs and benefits across social, economic, occupational, ethnic, regional, or other important groups of society. For example, United States Steel Corporation has signed a compliance order to spend $400 million on pollution control equipment for nine plants to bring them into compliance with air and water pollution regulations by the end of 1982.[15] Equity in this case is generally interpreted to refer to the costs borne by U.S. Steel, the communities in which these plants are located, and the firms buying the steel. But virtually all of these costs will be shifted to steel customers and ultimately to final consumers. To the extent that this shifting takes place, the financing of the cleanup will take on aspects of a general sales tax because steel is a major component of a large number of products used in final consumption. One equity question is the proper balance between the pollution ills borne by the relatively few people living near the plants and the costs that will be borne by the general population (in the form of higher steel prices). Similar questions arise under almost any type of EHS regulation.

A special aspect of equity refers to competition among regions or other groups. For example, most pollution control legislation mandates national standards that are uniform across regions of the country. The basis for this geographic uniformity rests, in part, on the premise that without such provisions, certain areas would have a competitive advantage in attracting industrial plants. However, economic efficiency requires recognition of the fact that both the costs associated with meeting those standards and the benefits from their implementation are likely to vary widely across geographical areas.[16] Thus, efficiency goals are often in conflict with equity goals having to do with geographical or other distributional considerations.

Administrative simplicity is a third goal. Government programs should be easy to understand, administer, and enforce. Furthermore, the costs associated with this administration should represent a relatively small proportion of total expenditures.

A fourth goal of governmental activities relevant to the regulatory process is the preservation of individual freedom. For example, providing information on the consequences to the public of actions or exposures is universally regarded as a proper role of government regulations. Going beyond this role to set standards on prohibiting products puts the rights of one person in conflict with those of another.[17] Regulations that attempt to protect people against themselves are certain to be controversial—for example, requirements that helmets be worn by motorcyclists. The more direct and apparent the governmental constraints, the more contentious they are likely to be. Determining the best balance between the rights of an individual and the goals of society is often controversial, yet is central to many regulatory decisions.

Another possible goal of governmental action, often cited in the EHS regulatory area, is that of forcing advances in technology. Setting stringent standards with tight deadlines is alleged to create the proper inducements to shake off lethargy and meet desired objectives. Whether more good than harm occurs is unclear. For example, although it is evident that air pollution laws precipitated the discovery of new technologies for the control of automotive emissions, current policies appear to have created perverse incentives for manufacturers and a slackening of effort on the part of the industry to "clean up" the automobile in an economically efficient manner.[18]

An awareness of the various basic goals of governmental activities and the contradictions among them is necessary if society is to develop sound regulatory strategies. Resolving the controversies and determining the specific form and timing of regulations are still thorny issues.

THE FORM AND TIMING OF REGULATIONS

Regulatory strategies currently take various forms in our economy. For example, regulation may be based on direct controls (on selected wages and prices), minimum quality standards (for foods, drugs, air, and water), conditions for production (health and safety at the workplace), constraints on the price structure (of public utilities and airlines), conditions of distribution (days of business), and so on. Many of these examples involve the setting of standards to be met by specific industries. However, recent discussions on regulatory processes have given greater emphasis to use of the market rather than standard setting to accomplish desired goals.[19]

For a number of years, economists have suggested that the objectives of cleaner air and water could be attained more efficiently by imposing the "correct" fees on pollution emissions rather than by direct control of the polluters in the form of technology-based emissions standards.[20] If a fee were levied on each unit of pollution generated, those firms with the lowest costs of pollution abatement might achieve lower levels of emissions than the current standards, whereas other firms with high abatement costs might elect to pay the fee and pollute in excess of the current standards. By manipulating the magnitude of the fee, the regulatory agency could achieve the desired levels of ambient air and water quality. This, in turn, could lead to more cost-effective air and water pollution controls; more importantly, under this system the prices of firms' outputs (and thus individuals' consumption decisions) would better reflect the true costs to society of producing goods and services.

A recent study prepared for the Council on Environmental Quality examined alternative policies for attaining and maintaining a short-term nitrogen dioxide standard, based on data pertaining to Chicago.[21] In addition to conventional regulatory approaches, the use of emissions fees and marketable permits was evaluated. The study concluded that strategies making use of such economic incentives (that take account of differences in polluters' costs of control and differences in polluters' contributions to ambient conditions) are far less costly than those that do not. At the same time, it was noted that such incentive plans are not without implementation problems. For example, to implement an efficient fee system, the administering agency must have considerable knowledge of sources' control costs. Under a permit system, the agency must organize and oversee a "market," making decisions on such factors as the number of permits to issue and the rules that govern exchange. Despite these difficulties, the potential gains from such approaches certainly warrant further theoretical and empirical study so that their comparative merits can be determined more completely.

Another issue is the latitude given the regulatory agency. The Congress might do no more than state a general goal, or it may specify a detailed regulation. The classical notion is that Congress provides general guidelines, but delegates operating authority to a regulatory agency. The role of the agency is to interpret these guidelines on a case-by-case basis and promulgate standards. Unfortunately, the legislation associated with many EHS issues is notably deficient in specifying workable goals.[22] Instead, much of this legislation begins with high-sounding rhetoric about protecting the health of every American,[23] followed later by a caveat that the regulations cannot bankrupt the industry. The result is frustration for the regulators since they lack a clear mandate. In fact, they are often sued by both parties in a dispute, each of whom can point to language in the legislation that implies, on the one hand, the decision in question does not give absolute protection or, on the other hand, that it is extraordinarily expensive.[24]

It is important that regulations be timely and implemented quickly when threats to the environment, health, and safety are involved. However, when not only accidents and deaths are at stake, but also billions of dollars of compliance costs, proposed regulations will be the subject of intense controversy with long delays in implementation because of problems in rule-making, court challenges, and difficulties in enforcement. Thomas Jorling, EPA Assistant Administrator for Water and Hazardous Materials, told a congressional committee:

> One of the most acute frustrations I have come to experience is the immense difficulty associated with taking statutory mandates into implementation. Complexity, procedures, and shortages of resources all contribute; but there are also larger, more pervasive reasons related to institutional fears of changing or altering the system.[25]

One of the consequences of this situation is that regulatory agencies may adopt a passive role and not act until outside pressure is exerted. For example, according to one report, unions or public interest groups have been initiators of OSHA, EPA, or FDA action in 22 of the 26 instances through 1976 when the agencies regulated carcinogens.[26]

Another consequence is that it now takes years to promulgate many of these regulations. Again, an illustration is helpful. The National Institute of Occupational Safety and Health (NIOSH) has compiled a list of 28,000 toxic chemicals, 2,200 of which it has classified as suspected carcinogens. Although NIOSH recommended exposure limits for more than 100 of these found in the workplace, as of January 1979, OSHA had promulgated permanent standards for only 23 during the eight years it has been in existence, and some of these are in litigation or have been thrown out by the courts.[27]

A related aspect of this dilemma is the issue of due process. The legal question is: On whom is the burden of proof? For example, a drug manufacturer must prove to the FDA that a product is safe before the FDA will allow it to be distributed. However, the EPA must prove that an air pollutant is harmful before it can regulate its release into the environment. Not only does the question of who has the burden of proof affect who bears the costs of research and development (R&D), but it also can affect the timeliness of the regulations. One would expect the FDA to undertake more questioning and to use more rigorous criteria in setting standards than the EPA. As a consequence, one would also anticipate relatively greater R&D expenditures on substances regulated by the FDA and longer delays in their introduction.

The requirements in resources and time to promulgate new regulations mean that an agency will publish only a few each year.[28] It thus becomes crucial to ensure that the most important problems are singled out. An agency must look for the areas in which it can hope to make the greatest improvement in the environment, health, and safety. For example, OSHA concentrated initially on occupational accidents. This made sense in that the increase in accident rates during the 1960s was largely responsible for OSHA's creation.[29] However, with Worker's Compensation having been in effect since the turn of the century,[30] many preventive measures had already been taken. Indeed, OSHA's subsequent research suggests that it has had little or no effect on accident rates, even for the industries in which it has focused.[31] In contrast, occupational disease had been almost completely neglected, providing an important area for intervention.

With few exceptions, toxic substances are seldom present in concentrations that would cause acute health or environmental hazards. Rather, their concentrations are usually sufficiently small that there are questions as to whether such adverse effects are likely to occur. Even in cases where such effects are anticipated, it is exceedingly difficult to estimate their magnitudes. Thus, the regulatory decisions concerning which substances to regulate and at what level is akin to a generic statistical problem with possible "Type I" and "Type II" errors. For example, test results may indicate that a non-toxic chemical is toxic or that a toxic chemical is not toxic.[32] The former, false positive, is a Type I error, while the latter, false negative, is a Type II error. If additional data are collected and analyzed, the risk of incorrectly categorizing a substance as harmful or benign can be reduced. However, collecting such information is time consuming and expensive. To keep regulations timely and to ensure that data collecting is not exorbitant, decisions must be made while there is still risk of misclassification. This uncertainty is inherent, and more attention must be devoted to assessing the tradeoffs between the value (or expected contribution) of additional information and analysis, versus the losses and risks associated with further delay. Regulations should be recognized as being subject to change as desires evolve and should be responsive to improvements in scientific knowledge.

Setting good regulations requires a firm foundation of data and analysis. This begins at the stage of identifying whether there are substantial market failures (in the broadest sense) that would justify government intervention, and continues through to the setting and enforcement of specific regulations. Although values, judgments, and politics may play major roles in shaping these decisions, data collection and analysis must first determine what is known and what questions remain to be settled by the other aspects of regulatory decisionmaking.

INFORMATIONAL NEEDS

To make intelligent public policy decisions concerning EHS regulations, decisionmakers need information on the costs associated with alternative regulatory actions as well as the beneficial effects that are anticipated from such actions. Methods of analyzing some of these costs and benefits as well as data requirements are discussed, in turn, below. Detailed discussion of estimating the health benefits from EHS regulations is relegated to the next section.

ESTIMATING THE COSTS OF REGULATION

Under various executive orders and legislative mandates, agencies have been required to estimate the costs of implementing many of their regulations.[33] Even though no aspect of this process is simple, the theory behind this cost estimation is relatively straightforward and uses well-known techniques. Nevertheless, cost estimates for specific policies have exhibited wide ranges. For example, 1977 estimates of the cost of environmental control produced by the Council on Environmental Quality (CEQ) from EPA data were only $19.3 billion, of which at least $2 billion were public expenditures and $5 billion were the cost of controls on private vehicles. Yet, EPA estimates (made between 1975 and 1977) of the prospective control costs for just five major industries forecast 1977 costs for these industries to be $7.8 billion. Extrapolation of this estimate to all industries would place the total at $25 billion in 1977, substantially higher than the CEQ estimates.[34]

Perhaps one source of confusion is that in ordinary usage "cost" means the monetary expenditure required to purchase a good or service, for example, the money outlay for pollution control equipment. To economists, however, "costs" means opportunities foregone. Thus, in evaluating the costs of regulatory policies, the question to be answered is: What resources must society give up in order to devote more to implementing specific types of regulations? The value of these foregone opportunities represents the true costs of the regulations. For example, the costs of pollution-control include requiring new abatement equipment, modifying existing equipment, and operating the equipment. More subtle costs relate to the many current and prospective processes and designs that can no longer be used since they do not meet newer and stricter regulatory standards. There are also less direct costs associated with pollution control, including those incurred when plants of limited profitability shut down because they cannot cover the costs of new controls. Finally, there are the costs of administering the control program itself. In each case, society's scarce resources are being channeled away from other possible uses.

There is an inherent bias toward overestimating the costs associated with many regulatory actions. For example, one common way of measuring the costs of pollution abatement is to estimate the expenditures of adding "end-of-pipe" devices to "representative" facilities in an industry.[35] But adding such abatement devices to old facilities is often economically inefficient since it may also decrease operating efficiency. In the long run, it may be more economical to construct new facilities. One way to improve our estimates of the costs associated with regulatory policies would be to undertake some retrospective analyses of the difference between estimated and actual costs. For example, one might select the area of air pollution control and examine the difference between estimated and actual control expenditures for individual plants. Such a study could enable one to estimate not only the degree of bias in estimating costs, but also the quality of the estimates and sources of bias or dispersion.

Another important aspect influencing the costs of regulation is the role of new technology and technological improvements. Current cost estimates are based almost inevitably on existing technology. At best, only educated guesses can be made as to the effects on costs of advances in technology. Thus, by their very nature, such guesses are subject to considerable uncertainty and their usefulness in estimating costs is extremely limited; generally, analysts do not assume any advances in technology. Thus, to the extent that major technological advances occur in the future and other assumptions are not overly optimistic, current cost estimates will overstate the realized costs of certain regulatory actions. Again, retrospective studies are needed to look more closely at specific areas of regulatory policy in order to determine how changing technology has affected costs.[36] Questions to be answered include: How rapidly have solutions to specific problems been developed? How have resources dedicated to research and development led to less costly solutions?

ESTIMATING THE BENEFITS OF REGULATION

Estimating the benefits of EHS regulations is extremely difficult. For example, the benefits of air pollution abatement stem from reduced morbidity and mortality rates; decreased damage to animals, plants, and materials; less foul-smelling air; greater visibility; and a general increase in the quality of life.[37] Measuring these effects is extremely complex as detailed in the next section on estimating health effects. This section focuses on analyzing the other categories of benefits.

To explore the effects of air pollution on plants, animals, and materials, two basic types of approaches can be taken: experimental or statistical. For example, suppose one were interested in determining the effects of air pollution on the frequency of repainting. Experimentally, one might expose a painted surface to polluted air in a city and compare its useful life with that of a control surface exposed to relatively unpolluted air. Unfortunately, a number of other factors could be expected to influence the useful life, such as exposure to sunlight. In addition, it is necessary to determine when repainting is necessary. Is discoloration sufficient to require repainting? Whose judgment should determine when the useful life of the painted surface is over?

Statistically, one might analyze data for different areas of the country on the frequency of repainting, the levels of air pollution, and other relevant factors such as climate. The main advantage to this approach is that actual consumer behavior serves as a gauge in valuing the adverse effects, rather than experimental evidence. However, such statistical approaches are not without difficulties because they require comprehensive data banks. We are at a rather primitive stage in placing a value on less odors, improved visibility, more sunny days, and other such quality-of-life factors. Conceptually, economists value these types of effects by "willingness to pay." For example, residents of Pittsburgh might be asked how much they would be willing to pay to reduce the odors associated with the steelmaking facilities. While simple conceptually, it is very difficult to elicit reliable information since incentives to bias answers are overwhelming. Valuing intangibles is an area in which new ideas are needed and to which further research dollars should be devoted.

Even if one knew the desired ambient air quality, it would be extremely difficult to estimate the amount of emissions abatement required to achieve this goal. To do so, one would need an inventory of current emissions. Then diffusion modeling could be used to infer how these emissions were distributed geographically. Finally, atmospheric chemistry and physics could be used to predict the chemical transformation, creation of aerosols, long distance transport, and deposition rates. For example, sulfur dioxide emitted in New Mexico is subject to diffusion over a wide area, with some sulfates reaching the Eastern seaboard;[38] sulfur dioxide can be absorbed by plants or soil, can be converted to sulfates, can be adsorbed on small particles, and can remain in the atmosphere for weeks. Clearly, advances are needed in diffusion modeling as well as atmospheric chemistry and physics.

ESTIMATING THE HEALTH BENEFITS OF EHS REGULATIONS

As do environmental programs, health and safety regulations focus on protecting the public health.[39] Thus, even a cursory examination of such regulations would conclude that their primary benefits were in some way related to improving the health status of society. However, it is not enough to answer only qualitative questions such as: Does air pollution cause ill health? Is occupational exposure to benzene related to the incidence of leukemia? Will ingestion of small amounts of lead result in damage to the central nervous system or behavioral change? Policymakers also need information of a quantitative nature, answers to such basic questions as, how much?

But estimating the effects of pollutants on humans is extremely complicated. We are exposed to pollutants through a variety of pathways, including the respiratory system, the digestive system, and the skin.[40] Depending, in part, on the pathway and chemical form, the pollutant is taken up in the body, transported to other sites, metabolized, and excreted. When concentrations are high enough to produce acute effects, isolating the effect is relatively straightforward. However, when exposure is to low levels of interacting insults over a long period of time, it is much more difficult to infer the effect of exposure.

The chief aim of many EHS regulations is prevention of chronic disease effects. Because both exposure and effect may take place over decades, the concept of causality becomes blurred.[41] To observe an association (or correlation) is not to establish cause. Today many scientists are willing to agree that "smoking causes lung cancer." However, this was a long time in coming and there are still credible scientists who do not accept this hypothesis as proved.[42] As a further example, we see that a strong correlation exists between incidence of cancer of the colon and per capita meat consumption. However, the causal situation is ambiguous.[43] It may be the meat itself, the absence of high grain consumption in the diet, the way in which the meat is cooked, or some unrecognized factor. Much careful research is needed to investigate possible causal relationships in the EHS areas.

A National Academy of Sciences (NAS) committee has distinguished between three types of approaches to study health effects in these areas: investigations of the biochemical and physiological mechanisms, toxicology, and epidemiology.[44] Each has its strengths and its weaknesses. No one research approach alone is sufficient to provide adequate information on health effects. In the following discussion we begin with an examination of the possible contribution of laboratory experiments (covering the first two approaches identified by the NAS committee) and then turn to epidemiology, highlighting some of the difficulties associated with this method. Finally, we treat the difficult topic of monetizing estimated health effects.

LABORATORY EXPERIMENTATION

Laboratory experiments designed to reveal the "mechanisms of action" provide important information to aid in formulating regulatory policies. For example, in examining the mechanisms by which air pollutants affect health, such studies have suggested that sulfur dioxide is less harmful than acid sulfates and that particulates in the respirable range of 0.5 to 3.0 microns are more damaging to the lower respiratory tract than are larger particles.[45]

At the same time, such laboratory studies have their limitations. For example, even if it is hypothesized that high levels of air pollution harm people, it would not be ethical to conduct experiments that expose subjects to such levels. Even aside from the ethical issues, it is unlikely that long-term, low-level effects on humans could be uncovered in a laboratory setting.[46] As a result, researchers often turn to the use of small animals, microorganisms, or cultured human cells.

Animal bioassays can demonstrate qualitative effects such as cancer or birth defects that might occur in humans and can suggest quantitative, dose-response relationships. However, they, too, suffer from some serious limitations. First, a number of uncertainties are associated with attempts to relate data on one species, such as mice, to possible effects in other species, such as humans.[47] Small mammals have relatively short lifespans and differ physiologically from humans. Furthermore, in the laboratory it is virtually impossible to duplicate exposure conditions in the "real world," either in terms of the other factors and variables that are present,[48] or in terms of the dosages. In addition, such experiments are expensive. A study of a single chemical is likely to require at least 500 animals (including controls) and a period of more than three years, at a cost of at least \$250,000.[49] Finally, when effects of very low-level, long-term exposures are the object of investigation, statistically valid measurements require either that a very large population be exposed,[50] or that doses substantially higher than those encountered in the ambient environment be used.[51] Accordingly, researchers sometimes turn to dose-response extrapolation techniques to estimate exposure effects.

Any extrapolation technique, which can use one of several mathematical methods, should also be based on what is known about the underlying biological mechanisms. Unfortunately, this knowledge is rarely present.[52] The most commonly used methods involve the linear, nonthreshold model and the long-probit model.[53] The use of such techniques is generally believed to result in conservative regulatory actions, since they tend to overestimate the health effect of the pollutant in question.[54]

Because of all the problems noted above, there has been a great deal of interest in alternative approaches. In particular, the development of short-term bioassays to identify potentially hazardous chemical substances has generated increased attention. The most thoroughly studied of these, developed by Bruce Ames, is based on the relationship between the carcinogenic activity of some compounds and their production of certain types of mutations. Over the past 15 years Ames and his colleagues have shown that about 90 percent (158 out of 176) of the organic chemical carcinogens tested were also mutagens, and most mutagens appear to be carcinogens. Only a few (13 out of 108) "noncarcinogens" tested were mutagenic and many of these, argues Ames, are close relatives of carcinogens.[55] In its current form, the

Ames test costs between $300 and $1,000 per chemical and is used in as many as 1,000 laboratories.[56] Even though the knowledge gained from such short-term bioassays as the Ames test is very valuable, the NAS has concluded that "experience does not yet warrant reliance on mutagenicity as a substitute for carcinogenicity testing."[57] Furthermore, there remains the important question of the quantitative usefulness of such short-term tests.[58]

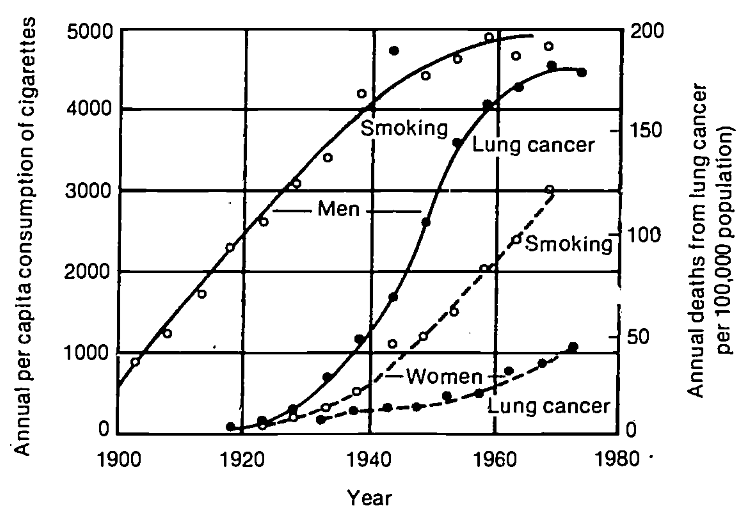
EPIDEMIOLOGY

Epidemiologists observe humans directly in their natural environment. When associations are discovered between patterns of disease and patterns of exposure, epidemiological studies can provide evidence about the quantitative (dose-response) relationship. Furthermore, when such epidemiological studies are supported by laboratory findings, a relatively sound basis for policymaking is assured. Unfortunately, as with the other approaches used to study health effects, epidemiology is not without its associated problems.

Spurious correlation is always a possibility. Many factors simultaneously affect human health, making it difficult to isolate specific causative agents; hence, any single, observed association may be an artifact of the particular data and population under investigation.[59] Furthermore, the use of epidemiological approaches to identify mutagens or carcinogens becomes even more problematic because of the long latency periods. This can be illustrated in the case of cigarette smoking.[60] Men started smoking cigarettes in large numbers in 1900, but the resulting increase in lung cancer did not appear until 20 to 25 years later. Similarly, women started smoking in appreciable numbers during World War II, and now their lung cancer rate is climbing rapidly (see figure 1).

Finally, there are serious deficiencies in the availability and quality of data for epidemiological studies. First, there are problems with the measures of health. For example, mortality data are not always recorded in consistent forms. Furthermore, because other factors affecting mortality must be controlled,[61] more extensive information than is currently being obtained should be collected, at least for a sample of death certificates. The need for a national registry of death certificates has been recognized for some time,[62] particularly in view of the mobility of the U.S. population. Because morbidity data are much more sensitive indicators of health status than mortality information, more effort should also be devoted to compiling illness records. What few large-scale morbidity samples exist contain inconsistencies and serious informational gaps.[63] Thus, there is a need for more comprehensive data including national registries for major diseases. Other useful morbidity information may be obtained from health maintenance organizations (HMOs) such as the Group Health Association and the

FIGURE 1. Relation between cigarette smoking and lung cancer.



Cigarette smoking and lung cancer are unmistakably related, but the nature of the relationship has remained obscure because of the long latent period between the increase in cigarette consumption and the increase in the incidence of lung cancer. The data are for England and Wales. In men (solid line) smoking began to increase at the beginning of the 20th century, but the corresponding trend in deaths from lung cancer did not begin until after 1920. In women (broken line) smoking began later, and lung cancers are only now appearing.

Source: "The Cancer Problem," by John Cairns. Copyright © 1975 by Scientific American, Inc. All rights reserved.

Kaiser Foundation Medical Care Plan. While these institutions represent a potentially rich source of large-scale morbidity data, current information is often quite limited in terms of specific diagnostic data and other factors that should be controlled (for example, smoking habits and socioeconomic status).

In addition to data pertaining to the general population, information on the health status of persons in particular occupations can be extremely useful. This follows from the fact that the best epidemiological results are obtained when there is a well-defined, relatively homogeneous group of people who have undergone similar exposures.[64] Nevertheless, no national occupational morbidity reporting system exists in the United States, although there are ongoing programs in several western states.[65] OSHA has now proposed rules that would require companies to keep careful records of exposure and health status for all employees exposed to any toxic substance.[66] Tens of thousands of workers are covered by these rules,[67] thus providing an invaluable data base if adequate information is collected on such factors as smoking habits, other personal factors, and residential history.

MONETIZING EFFECTS

Once the information is obtained to enable quantification of the relationship between exposure and health effects, then predictions can be made as to how reductions in exposure would change these health effects. The next step is translating the predicted changes into monetary equivalents. This is surely one of the most controversial areas in which economists intrude. Some say that putting a monetary value on human life is insensitive, crass, and even "inhuman." However, this view does not give sufficient weight to the fact that both individuals in their day-to-day actions and governments in their decisions about social policy do in fact make tradeoffs that involve the probability of death as well as goods with explicit dollar prices. These tradeoffs implicitly attach price tags to human life.[68]

Perhaps the controversy would abate somewhat and the dialogue would improve if the focus shifted from concern with placing the "correct" price tag on human life, to emphasis on estimating the value of lifesaving. The notion here is quite simple. Society has only limited resources to devote to lifesaving activities. The question is, where can those resources be used in the most cost-effective manner? Should we be spending one million dollars to prevent a cancer death by breast cancer screening or one hundred thousand dollars to save a life by getting people to buckle their seat belts?[69]

There is no denying that legitimate difficulties exist with placing monetary values on health effects or life itself. The problem is made even more difficult by the recognition that the values placed on such untoward events depend upon the probability of occurrence, the number of people affected, the specific type of effect, and people's general familiarity with the type of risk in question. Social psychologists have investigated the "psychology of risk" and found that a number of these dimensions influence whether individuals perceive a situation to be risky.[70] This, in turn, means that the values attached to such risks must often be situation-specific or, at the very least, tailored to the probability of occurrences, type of effect, and so on.

Economists have attempted to find clever ways of circumventing some of the problems of valuing life,[71] but their efforts are still regarded with suspicion, and often properly so. Nevertheless, government programs and other collective decisions involving the EHS issues can be analyzed carefully to determine the implied (or explicit) values placed on health effects, human life, or changes in the probability of death. This, in turn, could give insight into collective attitudes toward life valuations and, perhaps more importantly, could lead to greater consistency across public policies in these areas.

INSTITUTIONAL ARRANGEMENTS

In our review of "Informational Needs," we discussed briefly some of the difficulties associated with estimating the costs and benefits of EHS regulations. In the following section, we then detailed the complexities associated with estimating the health effects of such regulations, including problems of identifying causation, estimating dose-response relationships, and valuing the estimated health impacts. It is essential to have a variety of disciplines represented in the underlying research efforts characterized in these descriptions; to be successful, this work must involve adequate communication between fields. Furthermore, sophisticated methods of data collection are needed, including survey research. Finally, for much of the research outlined above, the large data banks must be manipulated by use of elaborate statistical programs and analysis.

These problems have implications for the institutions needed to carry out this research. The above requirements give rise to a picture of a sizable research organization that is interdisciplinary and that possesses excellent data collection facilities and computer hardware and software. If the interdisciplinary effort is to accomplish its goals, researchers must be willing to make the necessary investment to communicate with people from other disciplines.[72] At the same time, data collection and analysis procedures must be developed to permit extremely high capabilities. Together these factors suggest that such an institution requires long-term funding and support.

Several types of institutional "homes" are possible. Universities have the advantage of offering colleagues from a number of disciplines as well as graduate students who can contribute to the research. In addition, they tend to be relatively free of political influence. Government research facilities, such as the National Cancer Institute, offer a high degree of expertise in specific subject areas. At the same time, they are often in closer contact with policymakers, which can clearly be advantageous for undertaking research in support of regulations. Regional facilities, such as those found in Research Triangle Park, North Carolina, can forge simultaneous linkages to universities, business, and government, without necessarily developing intimate associations with any specific sector.

The prospect of progress in the scientific and technological areas that can enhance EHS regulations seems adequate justification for experimenting with several institutional settings in order to provide information on the pros and cons, suitability or unsuitability, of alternative ways to accomplish this necessary research.

NOTES

1. Economic regulation, to a large extent, had been confined to narrow objectives within single industries. For example, originally the Interstate Commerce Commission (ICC) regulated only railroads, then added other modes of transportation such as trucks and buses. Further examples can be seen in such regulatory agencies as the Civil Aeronautics Board (CAB), the Federal Communications Commission (FCC), and the Federal Energy Regulatory Commission (FERC).

2. Some of the background legislation, for example, the Clean Air Amendments of 1970 (P.L. 91-604), went so far as to specify exact emissions standards for pollution sources such as automobiles.

3. Some of these agencies (and the year in which they were established as opposed to the year of the establishing legislation) include: the Mining Enforcement and Safety Administration (1973), the Drug Enforcement Administration (1973), the Occupational Safety and Health Administration (1973), the National Highway Traffic Safety Administration (1970), and the Environmental Protection Agency (1970). See J. W. Allen, "Costs and Benefits of Federal Regulation: An Overview." Congressional Research Service, July 19, 1978, pp. 103–110.

4. For example, the automotive emissions standards (noted above) were clearly recognized as being beyond the then current technology. See E.P. Seskin, "Automobile Air Pollution Policy," in Portney, P.R., ed., *Current Issues in U.S. Environmental Policy*. Baltimore: Johns Hopkins University Press for Resources for the Future, 1978, pp. 74–75.

5. See, for example, "Has Environmental Regulation Gone Too Far?" *Chemical and Engineering News* 57:24, 1979.

6. DeFina, R., "Public and Private Expenditures for Federal Regulation of Business" (Working Paper No. 22), St. Louis, Washington University Center for the Study of American Business, November 1977, p. 3.

7. *Economic Report of the President*, February 1975, p. 159.

8. Denison, E. F., "Effects of Selected Changes in the Institutional and Human Environment Upon Output Per Unit of Input," *Survey of Current Business* 58:21, 1978.

9. "Economic Benefits of Air-Pollution Curbs Exceed Compliance Costs, Council Asserts." *The Wall Street Journal*, January 26, 1979, p. 20.

10. See Lave, L.B. and Seskin, E.P., *Air Pollution and Human Health*. Baltimore: Johns Hopkins University Press for Resources for the Future, 1977, chapter 10.

11. See Mitchell, R.C., "The Public Speaks Again: A New Environmental Survey." *Resources*, September-November 1978, p. 1.

12. See Ames, B.N., "Identifying Environmental Chemicals Causing Mutations and Cancer." *Science* 204:587, 1979.

13. Kendall, J. and Kriebel, D., "Carcinogen File: The Ames Test," *Environment* 21:15, 1978.

14. Smith, R.J., "Dioxins Have Been Present Since the Advent of Fires, Says Dow." *Science* 202:1166, 1978.

15. "Pollution Pact Sets a Record." *The Washington Post*, May 23, 1979, p. C1.

16. See, for example, Peskin, H.M., "Environmental Policy and the Distribution of Benefits and Costs," in Portney, P.R., ed., *Current Issues in U.S. Environmental Policy*. Baltimore: Johns Hopkins University Press for Resources for the Future, 1978, p. 144.

17. Smoking provides a good illustration. Labels on cigarette packages warn of the possible health hazards from smoking, and such warnings are generally accepted (except by manufacturers and other interested parties). However, recent attempts to pass local ordinances to restrict smokers in restaurants and other public areas have met with stiff opposition.

18. See Seskin, "Automobile Air Pollution Policy" (cited in note 4), pp. 83–87; and National Academy of Sciences, "Report of the Conference on Air Quality and Automobile Emissions to the Committee on Environmental Decision Making," mimeo (Washington, D.C., National Academy of Sciences, May 5, 1975).

19. See, for example, "White House Favors Economic 'Incentives' Over Pollution Regulations," *Environmental Health Letter*, August 1, 1977, p. 3; and the forthcoming Annual Report of the Council on Environmental Quality. In addition, section 405 of the Clean Air Amendments of 1977 calls for a study and report concerning economic approaches to controlling air pollution.

20. See, for example, Kneese, A.V. and Schultze, C.L., *Pollution, Prices and Public Policy*. Washington, D.C.: Brookings Institution, 1975.

21. Anderson, R.J., Jr., Reid, R.O., and Seskin, E.P., *An Analysis of Alternative Policies for Attaining and Maintaining a Short-Term NO_2 Standard*, prepared for the Council on Environmental Quality by Mathtech, Inc., Princeton, N.J., September 17, 1979.

22. Statutory deadlines present one of the clearest examples. The 1972 Clean Water Act, for instance, required every municipality to install secondary treatment and to meet water quality standards by July 1977. But even now that goal is years away from completion. See Quarles, J., "Runaway Regulation? Blame Congress." *The Washington Post*, May 20, 1979.

23. For example, the Clean Air Amendments of 1970 (P.L. 91-604) define the primary air quality standard to be one that protects the health of even the most sensitive individual.

24. For example, when the EPA Administrator announced a decision to suspend the pesticide aldrin/dieldrin, the Shell Oil Company, the manufacturer, and the Environmental Defense Fund, an environmentalist organization, each raced to be the first to file an appeal. See Smith, R.J., "Toxic Substances: EPA and OSHA Are Reluctant Regulators." *Science* 203:29, 1979.

25. Ibid., p. 28.

26. Ibid., p. 28.

27. Ibid., p. 29.

28. As of January 1979, the EPA had set permanent standards for exposure and effluent limits for only four hazardous air pollutants and six toxic water pollutants. Ibid., pp. 28–29.

29. See Smith, R.S., *The Occupational Safety and Health Act*, Washington, D.C., American Enterprise Institute, 1976, p. 5.

30. Oi, W.Y., "Safety at Any Price." *Regulation*, November-December 1977, p. 16.

31. Smith, R.S., *Occupational Safety*, chapter IV, cited in footnote 29.

32. The example is taken from Page, T., "A Generic View of Toxic Chemicals and Similar Risks," *Ecology Law Quarterly* 7:220, 1978 (Resources for the Future Reprint 159).

33. For example, the Clean Air Amendments of 1970 (P.L. 91-604) required the EPA to submit an annual report on the cost of air pollution abatement.

34. See Crandall, R.W., "Environmental Control is Out of Control," *Chemical & Engineering News* 57:30, 1979.

35. For a more thorough treatment of this issue, the reader is referred to: Gianessi, L.P. and Peskin, H.M., "Water Pollution Discharges: A Comparison of Recent National Estimates," Washington, D.C.: Resources for the Future Discussion Paper D-2, February 1977; and Hanke, S.H. and Gutmanis, I., "Estimates of Industrial Waterborne Residuals Control Costs: A Review of Concepts, Methodology, and Empirical Results," in Peskin, H.M. and Seskin, E.P., eds., *Cost-Benefit Analysis and Water Pollution Policy*, Washington, D.C.: The Urban Institute, 1975, p. 231.

36. For example, one area of importance relates to the abatement of sulfur dioxide emissions from stationary sources. Major improvements in stack gas scrubbers have resulted in more efficient and less costly pollution control; fluidized-bed combustion represents an even lower cost control alternative. See National Academy of Sciences, *Air Quality and Stationary Source Emission Control*, Report by the Commission on Natural Resources, National Academy of Sciences, National Academy of Engineering, serial no. 94-4 (Washington, GPO, 1975), especially part II. Another example, noted above, concerns the technological progress associated with meeting the automotive emissions standards; see footnote 18.

37. See Lave and Seskin, *Air Pollution*, chapter 10, cited in footnote 10.

38. National Academy of Sciences, Committee on Nuclear and Alternative Sources of Energy, *Risk Panel Report*, 1979, unpublished.

39. EHS areas that require further scientific evaluation include the health effects of air pollution, smoking, low-level ionizing radiation, cyclamates, saccharin, food preservatives such as nitrites, various kinds of diets, and various occupational hazards such as coal dust, cotton dust, benzene, and arsenic, just to name a few. See, for example, U.S. Occupational Safety and Health Administration, "Hearings on Generic Rulemaking Proceedings on Carcinogens," Washington, D.C.: U.S. Occupational Safety and Health Administration, 1978.

40. See, for example, National Academy of Sciences, *Effects of a Polluted Environment: Research and Development Needs*, Panel on Effects of Ambient Environmental Quality, Washington, D.C.: National Academy of Sciences, 1977, pp. 11, 41–42; and U.S. Department of Health, Education, and Welfare, *Human Health and the Environment—Some Research Needs*, Report of the Second Task Force for Research Planning in Environmental Health Science, Washington, D.C.: GPO, 1977, pp. 5–17.

41. Some kinds of human cancers have latent periods of twenty years or more. See Bingham, R., Niemeier, R.W. and Reid, J.B., "Multiple Factors in Carcinogenesis," Symposium on Occupational Carcinogenesis, *Annals of the New York Academy of Sciences* 271:13, 1976.

42. See Sterling, T.D., "A Critical Assessment of the Evidence Bearing on Smoking as the Cause of Lung Cancer," *American Journal of Public Health* 65:939, 1975.

43. See Harriss, R.C., Hohenemser, C., and Kates, R.W., "Our Hazardous Environment," *Environment* 20:9, 1978.

44. See National Academy of Sciences, *Effects of a Polluted Environment*, p. 41, cited in footnote 40.

45. See Amdur, M.O., "Toxicological Guidelines for Research on Sulfur Oxides and Particulates," in *Proceedings of the Fourth Symposium on Statistics and the Environment*, Washington, D.C.: American Statistical Association, 1977, p. 48.

46. For example, a slight increase in the mortality rate of subjects with life expectancies of seventy years would require hundreds of thousands of subject-years to ascertain the relationship.

47. In fact, some substances such as aflatoxin are found to be potent carcinogens in rats, but have little effect on mice. See U.S. Food and Drug Administration, "Assessment of Estimated Risk Resulting from Aflatoxins in Consumer Peanut Products and Other Food Commodities," January 19, 1978, unpublished.

48. For example, pollutants often interact to produce synergistic effects that are greater than the sum of the individual effects. An extreme case would be a chemical that is not itself a carcinogen, but acts as a promoter in the presence of one. See U.S. Department of Health, Education, and Welfare, *Human Health*, pp. 174–75, cited in footnote 40.

49. Maugh, T.H., II, "Chemical Carcinogens: The Scientific Basis for Regulation," *Science* 201:1200, 1978.

50. An environmental carcinogen causing cancer in one percent of 100 million people would result in a million cases of cancer. Detection of a chemical causing cancer in only one percent of test animals would require the use of 10,000 rats or mice and still one would have little confidence in so subtle a result. Furthermore, testing more than 50,000 chemicals with 10,000 rats or mice for each would require 50,000,000 animals, take many years and would be prohibitively expensive. See Ames, "Identifying Environmental Chemicals," p. 589, cited in footnote 12.

51. By exposing animals to as high dose as possible (the "maximum tolerated dose") statistical problems may be overcome; however, additional controversy is generated because such massive doses may overwhelm the defense mechanisms of the animal in question and lead to a pathway or disease that would never occur from lower doses. See Maugh, "Chemical Carcinogens," p. 1201, cited in footnote 49.

52. Because of this ignorance, some scientists refuse to extrapolate; see, for example, Carter, L.J., "How to Assess Cancer Risks," *Science* 204:811, 1979.

53. See, for example, Crump, K.S., Hoel, D.G., Langley, C.H., and Peto, R., "Fundamental Carcinogenic Processes and Their Implications for Low Dose Risk Assessment," *Cancer Research* 36:2973, 1976; and Hoel, D.G., Gaylor, D.W., Kirschstein, R.L., Saffiotti, U., and Schneiderman, M.A., "Estimation of Risks of Irreversible, Delayed Toxicity," *Journal of Toxicology and Environmental Health* 1:133, 1975.

54. See National Academy of Sciences, Advisory Committee on the Biological Effects of Ionizing Radiations, *The Effects on Populations of Exposure to Low Levels of Ionizing Radiation*, Washington, D.C.: National Academy of Sciences, 1972, and Marshall, E., "BEIR Report on Radiation Hazards Comes Unglued," *Science* 204:1062, 1979.

55. Ames, "Identifying Environmental Chemicals," pp. 589-90, cited in note 12.

56. Maugh, "Chemical Carcinogens," p. 1203, cited in note 49.

57. National Academy of Science, *Effects of a Polluted Environment*, p. 52, cited in note 40.

58. For example, there is more than a million-fold range in mutagenic potency in the *Salmonella* test used by Ames and there is also a similar range in carcinogenic potency. See Ames, "Identifying Environmental Chemicals," p. 592, cited in note 12.

59. For a discussion of these and other difficulties in epidemiological research, see Lave and Seskin, *Air Pollution*, especially chapter 2, cited in note 10.

60. This illustration is taken from Ames, "Identifying Environmental Chemicals," pp. 587-88, cited in note 12.

61. They include physical socioeconomic, and personal characteristics such as age, sex, race, income, smoking habits, exercise habits, genetic history, nutritional history, and medical care, as well as other

environmental factors such as climate.

62. See U.S. Department of Health, Education, and Welfare, *Man's Health and Environment—Some Research Needs*, Report of the Task Force on Research Planning in Environmental Health Sciences, National Institute of Environmental Health Sciences, Washington, D.C.: GPO, 1970; U.S. Department of Health, Education, and Welfare, *Statistics Needed for Determining the Effects of the Environment on Health*, Report of the Technical Consultant Panel to the United States National Committee on Vital and Health Statistics, Washington, D.C.: GPO, 1977; and National Academy of Sciences, *Effects of a Polluted Environment*, p. 42, cited in note 40.

63. Ibid., p. 41.

64. See Maugh, "Chemical Carcinogens," p. 1200, cited in note 49.

65. U.S. Department of Health, Education, and Welfare, *Statistics Needed*, p. 8.

66. See Abelson, P.H., "Regulating Exposure to Carcinogens," *Science* 202, editorial, 1978.

67. See Zeckhauser, R. and Nichols, A., "The Occupational Safety and Health Administration—An Overview," in *Study on Federal Regulation*, vol. VI, Committee on Governmental Affairs, U.S. Senate, Washington, D.C.: GPO, December 1978, p. 163.

68. For example, the Federal Aviation Administration (FAA) is charged with regulating air traffic to ensure public safety. By tightening controls, the FAA could reduce accidents further, probably with an accompanying increase in the cost of air transportation. Thus, each FAA decision on a safety issue implicitly contains a judgment concerning the value of life. See Fromm, G., "Civil Aviation Expenditures," in Dorfman, R., ed., *Measuring Benefits of Government Investments*, Washington, D.C., Brookings Institution, 1968, p. 172.

69. For more on this, see Raiffa, H., Schwartz, W.B., and Weinstein, C., "Evaluating Health Effects of Societal Decisions and Programs," in *Decision Making in the Environmental Protection Agency: Selected Working Papers*, vol. IIb, Washington, D.C.: National Academy of Sciences, 1977, p. 1.

70. For example, see Slovic, P., Fischoff, B., and Lichtenstein, S., "Issues in Hazard Management—7: Perceived Risk," *Environment*, 1979; and Tversky, A. and Kahneman, D., "Judgment Under Uncertainty: Heuristics and Biases," *Science* 185:1124, 1974.

71. For example, Thaler and Rosen have used information on risky occupations to impute the value of life. See Thaler, R. and Rosen, S., "The Value of Saving a Life: Evidence from the Labor Market," in Terleckyj, N., ed., *Household Production and Consumption*, New York: National Bureau of Research, 1976, p. 265.

72. We stress the need for dialogue between scientists (in the broadest sense) and policymakers in our recent paper; see Lave, L.B. and Seskin, E.P., "Epidemiology, Causality, and Public Policy," *American Scientist* 67:178, 1979.

5 Crime and Technology: The Role of Scientific Research and Technology in Crime Control

*by Peter K. Manning**

SUMMARY

Crime is a major social problem in the United States. It concerns politicians, the public, and social scientists, but most of all it concerns the police. Americans believe that progress more often than not is brought about by applying technology to problems. It is not surprising, then, that the urban police in America in the last 50 years or so have become both more concerned about crime and more technologically advanced. The police have differentially accepted innovations designed to improve their crime control capacity, but in the last 20 years especially, and since the founding of the Law Enforcement Assistance Administration in 1968, crime and technology have become intimately linked. In a sense, technology developed in other settings—computers, communications systems, and high-powered machines such as automobiles, helicopters, and fixed-wing aircraft—was introduced into a traditional, conservative, multicentered, or decentralized organization. Still remaining are the constraints on policing (including both procedural and statutory law); the lack of resources, information, and citizen cooperation; the private nature of much crime; and the problem of maintaining public order concepts that differ widely in a pluralistic society.

Technology developed outside of policing is seen as having great promise in crime control, even though there is not a deep understanding of its effects by the police nor, conversely, a deep understanding of police organization by many researchers involved in introducing new technologies to police work.

Technology, defined as the means by which an organization transforms or works on inputs to produce products, can be subdivided into knowledge technology or information used in the workflow; operations technology, which shapes the sequencing and patterning of the workflow; and material technology, which has to do with the physical properties of the objects used in work. The police use a human material technology and an operations technology that is highly individualized, nonroutine, and patterned by situations. They do not have a theory of policing that systematically organizes their work to produce maximal effectiveness. The impact of new technologies developed outside policing always must be seen in the context of present police technology, activities, focuses, and segments. In addition, new technology will interact with present or conventional technology.

Information and its flow through police organizations is perhaps the most important object on which policing works and which in turn can be considered a "product." The police are highly dependent on many kinds of information. The segments of police departments, their

* Professor of Sociology and Psychiatry, Michigan State University.

activities, and their knowledge resources each pattern the effects of technology on policing. Police deal with primary, or environment-gathered, information received by officers; with secondary, or once-processed information in investigative units; and with tertiary information or data about information and its organization, assembled at the administrative level of an organization. The bulk of information is not systematized in departments, links between levels of information are asymmetrical, and the effects of technology vary within a department.

There are three basic police operational strategies or patterns of relating to the environment. These are the *preventive approach*, which involves deterrence or prevention of crime; the *proactive approach*, which involves reconstruction of the crime, creation of the conditions under which it will occur, and active intervention; and the *reactive approach*, the traditional and most dominant strategy of responding to, investigating, and trying to solve or otherwise clear a crime already committed. In each, the role of information and of technology varies. In preventive policing, information of every sort (prospective, retrospective, and applied) is seldom used because the means to get it are not precise, definitive goals are ambiguous and often unstated, and resources are scarce. In proactive policing, prospective intelligence is used to identify and surveil possible villains, and retrospective intelligence may be used to link those arrested to others, but the limits of the knowledge available, as well as the great discretion of individual officers, means that shared knowledge is a very small part of the total knowledge held in the police unit. In reactive policing, significant technological innovations have been made: files of arrestees, outstanding warrants, and vehicle and license numbers have been computerized in many departments. These are largely retrospective uses; little if any prospective intelligence is gathered, and applied intelligence can only be used where officers have some understanding of the links between the crime, the suspects, and other individuals.

Technological innovations have been of two kinds, those based on primary and on secondary information. The first kind involves attempts to reduce response time and to increase the visibility and presence of the police. Research in the second area has focused on the impact of more professional juvenile units, more effective case screening for investigators, and reorganization of detective work. In general, research does not yield unequivocal results, and the precise consequences of technological innovation where primary information is involved are as yet not established. In the case of secondary information, some efficiency—meaning the use of resources to achieve an aim—is gained by case screening, but it is not clear that clearing cases should be the aim of detective work. Other values may be served by policing, and screening large numbers of cases may not be in the best interests of the department in terms of public contacts, information, victim assistance and reassurance, and police prestige. For example, police juvenile units using knowledge technology still leave enormous discretion to the officer and the unit.

The paradox of technology is that although it was seen as a clear and present means of producing more effective policing, especially in the area of crime control, it did not produce simple outcomes or effects. Many technological innovations have failed to take into account the salient features of the social organization of police work. Technology has had no narrowly predictable outcomes because its effects are patterned by the arena in which it works; the activities and strategies involved; the level, quality, and amount of information available; and the constraints within society, such as lack of citizen cooperation. Clearly, there are many factors that help form a crime pattern. The police have only marginal effects on those factors, and new technologies can make only limited increments in crime control, given this basic fact. Thus the precise effects of new technologies on crime cannot be established at this time.

INTRODUCTION

Crime is a major social problem in our nation. It concerns politicians, the public, and social scientists, but most of all it concerns the police. Americans believe that progress more often than not *is* achieved by applying technology to problems. It is not surprising, then, that the urban police in America in the last 50 years or so have become both more concerned about crime and more technologically advanced. The police have differentially accepted innovations designed to improve their crime control capacity, but in the last 20 years especially, and since the founding of the Law Enforcement Assistance Administration (LEAA) in 1968, crime and technology have become intimately linked. In a sense, technology developed in other settings—such as computers, communications systems, and high-powered machines like automobiles, helicopters, and fixed-wing aircraft—was introduced into a traditional, conservative, multicentered, or decentralized organization. Still present are the constraints on policing, including both procedural and statutory law; the lack of resources, information, and citizen cooperation; the private nature of much crime; and the problem of maintaining public order concepts that differ widely in a pluralistic society. This chapter will assess the contribution of technology to crime control by the police in the context of the social organization of police work. It will also review the research available on technological innovations in policing and their impact on crime.

THE CONTEMPORARY CRIME FOCUS OF URBAN POLICE

American police perform a variety of functions, only a small minority of which can be construed as crime focused.[1] Police claim and accept as central and most significant their law enforcement duties, while they routinely perform diverse other social services that they view as onerous, self-demeaning "dirty work." Thus, although they claim one central function, they are in fact obligated to carry out a number of others.

There are other major institutional contradictions in the police organization, and each contributes to the collective wish of police to define their role as thief takers and crime stoppers. Although they wish to seem fair and equitable, they must selectively enforce the law and must face public criticism for "bias," "repression," and other charges—even from the middle classes with whom they identify. In short, they must be aware of their political base, even as they assert their apoliticality. Although they are legally required to enforce the law and maintain public order, the police often have little legal power to do so (e.g., when crime is committed in private or among complying citizens). In all crime, they depend on the nature of the actions and groups involved: the more organized the criminal activities, the more effective they can be.[2] Procedural guarantees constrain their work. Standards of order differ among social classes and ethnic groups, and the police must judge what is acceptable. They are dependent upon witnesses, informers, and the public for information and cooperation, upon prosecutors for charges, and upon courts for convictions. They must conceal information from some public segments while revealing it to others, and this often leads to lying and deception.[3] They must rely on other organizations and police units to cooperate in providing information.

These constraints on the police are affected by technology to some degree, as we will see, and they are the primary reasons why the police dramatize and publicize their crime-related functions while making efforts to conceal and understate the non-crime-related functions, as they define them.[4] Similarly, they seize upon and display symbols of authority and power that are associated with the state, morality, and the law. The police tend to judge themselves with reference to their crime work or "real police work."

Administrators, while taking a somewhat broader view of policing and its responsibilities, tend also to adopt the crime-fighting rhetoric, even when it would appear to be a rather modest part of the role of the police—as in suburban or largely middle class communities where the crime rate is relatively low and of little public concern. The problem of assessing technological effects on crime thus hinges upon the definition of the key terms "crime" and "crime focused." The degree of crime focus varies from officer to officer,[5] from segment to segment within a department,[6] and from department to department.[7] The capacity of the police to be crime focused and the extent to which they realize that capacity have not been carefully studied until very recently, and most of the available studies use questionable official statistics generated by police departments. Most of the available measures assess the *means* of controlling crime, rather than the objectives or goals of crime control. When speaking of the crime focus of the police, one is speaking of an ideological justification, a distribution of attitudes and functions or roles, and of actual time spent. Each of these is subject to measurement and should not be seen as either a present or absent phenomenon in large city police departments. However, there remains little consensus on the definition of crime-focused policing.[8]

Although academic debate surrounds the definition of crime-focused police work and the most appropriate sort of data needed to measure crime and crime effectiveness, the police use their own statistics, standardized to a limited degree by FBI instructions, and they maintain control over the gathering, processing, dissemination, and analysis of those statistics. They do so for rather obvious reasons, in spite of repeated systematic criticism.[9] They are able to control what is accepted as "crime" because they receive the calls, process them, decide if a crime has actually happened, and shift the ecological bases of counting the calls (to show increases or decreases by areas). They can use the calls to show trends, even though there is no theory to account for changes, the explanations offered are usually ad hoc (the introduction of a new patrol strategy, changes in the weather, new beat allocations), and statistics have a scientific cast to them.

Although the actual amount of crime in a society is unknown and unknowable and every measure is only a partial one,[10] crime is a police construction. That is, the crime rate is based upon crimes known to the police, and crimes found by them, and it omits unreported, plus unfounded crimes, the so-called dark figure. Finally, the police present crime as if it were all of a piece, rather than an incredibly diverse set of events that are variably subject to police control and/or suppression. Perhaps it is the constraints on the police and their wish to control crime that has led them to accept, especially in the last few years, a variety of technological innovations designed to increase their capacity to control crime.

TECHNOLOGY AND THE POLICE[11]

The police developed as a decentralized, highly discretionary system of social control that was ecologically or territorially based, subject to rather limited command and control, and employing rather simple technology. There are a number of reasons why the police in America, in the last 50 years in particular, have adopted innovations from other systems of work processing. New forms of technology might lend prestige to the professionalization movement within policing and might satisfy the status aspirations of police administrators.[12] Newer and more powerful weapons systems, mobile computers, faster vehicles, and instant radio communication with centralized dispatch were viewed with caution by many officers, but they also appealed to their sense of adventure and excitement and confirmed the crime-fighting role. They provided distance, symbolized control, and produced respect from the public.

At the social organization level, the symbols associated with the new technologies decrease police accountability because they induce a heightened sense of police power and control among the policed. The attitudinal and social isolation of the police historically is amplified by the symbolism of technologies, and they use this symbolism to maintain a formal, almost mythical, claim to control crime. Internally, in terms of everyday functioning, it is not clear that the substance of police work has been radically modified in the last 150 years. At least the argument can be made that the structural, functional, and symbolic dimensions of policing have remained remarkably similar and that the power of the occupational culture is the single most important influence on officers and on policing.[13] The degree to which new technologies have changed these practices is a significant research question.

The control of the occupational culture over police practice is profound. In concert with the uncertain nature of the work, the non-routine nature of the problems faced, the uneven demands for time and effort across events, and the situational uncertainty faced by the officer, characteristics of the occupational culture mediate the internal rules and procedures and the external public or environmental demand features. The occupational culture of policing is the context into which any new innovations must be introduced; it includes the traditional devotion of the police at every level to a clinical, face-to-face view of their essential work, a correlate resistance to planning and to rational anticipation of problems, and a distrust of science and knowledge that is probably based in class origins and level of education, as well as in the enduring influence of socialization in the organization—the cultural, social, and political isolation—of the police and their local character.

An important determinant of the marginality of new technology in policing is the very traditional, almost patrimonial, nature of police bureaucracy. In contrast to the classic Weberian formulation of rational legal bureaucracy, which emphasizes formal rules and procedures, the structure of offices, and defined rules and attitudes as routine ways of achieving ends, police organizations resemble patrimonial bureaucracies.[14] Moreover, the patrimonial nature of police organizations means that loyalty, fealty, and ascribed attitudes toward persons vitiate alternative rational forms of work organization.[15] For example, the ecological proximity of units in policing (e.g., detective functions are clustered) is a structural substitute for formal coordination by rule and procedure.[16] Each unit functions in semiautonomy, and the organization acts as a set of loosely linked domains.[17]

In an echo of the Wickersham Report of 1931 (which emphasized the utility of new technologies and the need for scientific statistics), of the President's Crime Commission (1967), and of the National Commission on Productivity and Goals (1973), LEAA has recently funded a variety of innovations in technology. These include efforts to create better statistics on crime, more rapid response time, and better weapons, uniforms and other equipment. LEAA has focused especially on funding central computerized systems of command and control as well as information storage and retrieval. Many of these innovations, as Coulton has noted, were based on simple assumptions about the nature of the police role in crime control.[18] It was assumed that the problems were technical in nature. If they could be solved, innovations supported by a small group of administrators (regardless of their degree of understanding) and quickly implemented would rapidly bring positive results without unanticipated or negative corollaries. These assumptions were only partially true, and many innovations experimental or otherwise failed as a result.

The following sections define technology, identify its organizational effects, and assess the research based on technological innovations in policing. Organizational constraints, such as structure and information processing and strategies, along with external constraints, such as citizen cooperation, available information, and the variable nature of crime, exercise controlling effects on the police, and reduce the hoped-for effects of new technologies.

TECHNOLOGY AND POLICE ORGANIZATION

Techne, a Greek word, means an art, craft, or skill. Technology is defined in Webster's as (1) "a technical language," (2) "an applied science" or "a technical means of achieving a practical purpose," and (3) "the totality of means employed to provide objects necessary for human sustenance and comfort." Although these broad terms imply a very general method that is systematic and logical, technology is used in several specialized senses in the literature on organizational behavior. The assumption is that no social variable contributes more to the determination of organizational structure than technology. Work processes and flow are said to shape organizational structure, and work processes, to a greater or lesser degree, are patterned by technology. However, there is no agreement in the organizational literature about the phenomena to which technology refers:

> . . . the concept of technology has been operationalized in terms of the extent of task interdependence, automation of equipment, uniformity or complexity of the materials used, and the degree of routineness of the work . . . there seems to be general agreement that organizational technology involves either the mechanical or intellectual processes by which an organization transforms inputs, raw materials, into outputs.[19]

Important distinctions are made between types of technology. *Operations technology* involves sequencing and equipping the activities in the workflow, or the ways in which input is processed and output is produced and distributed. *Material technology* has to do with the physical characteristics of the materials used in the workflow. These materials are quite variable, from people to prunes, from gas to gastronomes. *Knowledge technology* refers to the characteristics of the knowledge used in the workflow.[20] The amount, quality, and specificity of the relevant knowledge and what is used in production may vary. These analytic types of technology can coexist in an organization, they may be in conflict or be compatible, or they may either increase or decrease integration of the various subsystems, producing structural differentiation and autonomy.

In police work, the officers have substantial discretion to shape tasks: the ordering of tasks, the kinds of tasks sought and available, and the resolution of them are not routine. Policing is individualized, and the sequencing of tasks is controlled and shaped by organizational decisions at every point, especially those made by the officer on the street. The material technology of policing processes human "units"—human interactions and group structures that are abstract, variable, and unpredictable. People have the choice to do other than what is mandated in the production process. Finally, knowledge technology, the standardized information available about the varieties of objects processed, is little developed. The police do not have clear standards to apply to people, their "raw material"; they do not have a set of routine procedures for processing that material, nor a systematic theory of causality explaining their work purposes.[21] The knowledge available is of a general sort, and it does not include consensually defined categories of information that is relevant to standard situations. The knowledge is more a "recipe knowledge" passed on through apprenticeship relations, by oral traditions, and through actual experience.[22]

Finally, there is a basic difficulty in defining the environment to which technology is directed, and how it shapes and determines the internal structure of the organization. Aside from the perceptions, strategies, understandings, and actions of the organizational members, it is virtually impossible to measure the environment to which such organizations are said to be responding and adjusting to their technology.[23] Among people-processing organizations, much of what the organization does, why it does it, to what and whom, and by what means, is patterned by tradition, culture, history, and the autonomy and power of various organizational segments. It is as rational to argue that organizations *create* a sense of the environment to which the organization adjusts over time, shaping its rules, procedures, and postulates, as to argue that environment, working through technology, creates and sustains the organization.

In summary then, the traditional technology of the police is non-routine, the material means used are limited and they are applied to people, and the knowledge available is nonsystematic situational knowledge rather than theoretically derived causal theories. Historically, the police have used very little of the formal technologies found in manufacturing; they have been content to restrict their technology to semi-routinized processing. Higher forms of technology—that is, rational routine types found in mass production, or material technology used in large batch work—are also associated with rational legal bureaucratic forms. In the police, bureaucratic organization of a special sort is combined with low routinization, limited amounts of material technology, and a quasi-logical or commonsense theory of policing.

Finally, the police have in the past used rather unsystematic files, records, and systems of intelligence gathering, processing, and storage. We follow here Zey-Ferrell's definition of technology as the means by which an organization transforms raw materials or inputs into processed outputs.[24] We make several further distinctions. The ways in which the police presently or traditionally work of course involve a technology—a nonroutinized, tacitly articulated technology. Police technology is the way police agencies work. Knowledge or information (defined below) both shapes the technology and is transmitted through or worked on by technology. Tasks are the subcomponents of the work process, accomplished by means, or technology, which can be modified. Whether a modification of a task by technology means that a new technology has been applied or the task has been redefined is difficult to answer. Workflow is the sequencing of tasks.

Finally, bureaucracy consists of the formal rules of procedure, offices, functions, and the associated structure of authority. The term technology refers both to traditional, rather disorderly, police technology and to "new technologies" associated with other work systems. Both the present police technology and new technologies are discussed here as they interrelate and as they affect crime. When one speaks of technological innovations in policing, one is referring to the attempt to introduce new means of defining and interfacing with the environment and of processing the product. Thus the innovations reviewed here are sometimes in operations technology (such as in the development of new processing schemes in detective work), are sometimes in material technology (such as the use of computers and centralized radio dispatching), and are sometimes in knowledge technology (such as the development of new models or strategies of enforcement and deployment of resources). To understand the introduction of these technologies into police work, the social organization of police work as traditionally practiced must be considered in concert with the innovations.

SEGMENTS, ACTIVITIES, AND RESOURCES

The role of technology in policing is patterned by the *segments* of police organizations (col. 1, table 1), the product of historical developments in policing; the *focal activities* of each segment (col. 2, table 1); and the *knowledge resources* of departments (col. 3, table 1). Information and its patterning further divide the substantial areas (see below). Three focuses of patrol are apprehension, deterrence, and intelligence gathering, although other functional classifications have been evolved.[25] All active units share these activities to

Table 1. Illustrative Outline of Police Segments, Activities, and Resources*

Objective Demands	Activities	Resources	
Substantive Area	Focus of Activity	Components	
		Hardware	Software
Patrol**	Apprehension, deterrence, intelligence	helicopters, planes, radios (personal and force), mobile computers/force computer, automobiles, weapons systems	Officers, non-sworn personnel, files, records, dispatch tapes
Traffic	Monitoring and facilitating apprehension	See Patrol	Officers, non-sworn personnel, files, court records, staff
Juvenile	Liaison, deterrence, apprehension	See Patrol	School records, court records, officers and staff, files: personnel, case, informant
Vice-Narcotics	Apprehension, intelligence, seizure regulation	Force computer, radios (personal and force), cameras, video equipment; body mikes, transmitters; transponders; automobiles; weapons; radio equipment—force computer, automobiles	Files, officers, non-sworn personnel, informants
Detective Work	Investigation, clearance, case-preparation	Weapons, tape recorders, cameras	Files, informants
Internal Affairs	Supervision, intelligence	See Vice-Narcotics	Case records, data files, informant files
Administration	Policy-setting, command control—supervision, budgeting—allocation	Information storage, processing and retrieval systems	Programming, recordkeeping files
Research and Planning	Intelligence, evaluation, planning	Information storage, processing and retrieval systems	Programming, staff, data, files

* These examples are illustrative and not exhaustive.

** The resources and technology of patrol, more than any other unit, are generally available to other segments in the organization.

greater or lesser degrees, with additional variations such as liaison and investigation. The resources of information technology, divided into software and hardware, are shown in the third column. The distinctive technology of internal affairs, narcotics-vice activities and, to some degree, detective work stands out. The administrative and planning and research units have formidable amounts of technology and staff. While the largest proportion of resources, personnel, and equipment is used by the patrol division, the capacity to synthesize information lies in administration.

INFORMATION SYSTEMS

The material in table 1 can be analyzed by noting the informational correlates of the first column. The first five substantive areas listed in col. 1 carry out line functions (although some departments consider vice and narcotics enforcement to be staff functions), the others staff functions. The range of tasks, although diverse, includes variants on the type of information processed, the degree of differentiation of the information storage system, and the extent to which the unit monitors or acts as a meta-integrative unit for other kinds of information.

Organizations are structures of information processing (as well as structures of material and operations processing), and they contain levels of information about information. The entire police department is dependent upon *primary information*, which flows in from citizens (according to one observer, approximately 87 percent of all requests for services are initiated by citizens), through central communication to patrol units.[26] Primary information is received to a lesser degree by units such as vice-narcotics, traffic, juvenile, and internal affairs. This information can be used to act immediately, or responded to as a *spontaneous demand*. It can be the basis for planned requests, such as providing officers to serve as crossing guards, at parades and demonstrations, or at athletic events. Police can also *create demand*, such as when they define a pattern of crimes, public concerns, or anticipated problems as being in need of control and then act to control it. This is the basis of much of the work initiated by vice-narcotics, but other units also act on created demand.[27]

Other segments of the organization, such as juvenile, vice, and detective, depend on *secondary information* that has been previously encoded and classified by other agencies, informants, or other officers (principally in patrol). Upon receiving primary information, each of these units must decode it and then encode it into terms useful for their practical activities. For example, patrol may forward to a narcotics unit a report that someone is "selling dope" in a particular neighborhood, or that an arrestee was found with heroin in his car. This initial bit of data, if perceived as useful, must be transformed into an information report, assigned to an investigator, and investigated as to veracity. Conditions for an arrest are then established if possible. (This process takes place in organization-centered drug units where such information is systematically decoded and encoded. In others, it is left to individual officers to decide whether to proceed further on any information received.) In the administrative, internal affairs, and research and planning units, information may be primary or secondary, but most of it is *tertiary*—that is, information about primary and secondary information in the system.

A distinction that to some extent elaborates upon the primary-secondary-tertiary classification is the degree of differentiation of the information system. Another term for this is *entropy*, or the equivalence of the amount of information across the segments of police work. Patrol systems are most information intensive, and most tied to the inputs flowing from the environment, and their exchanges with the environment are most likely to be written or recorded in the form of official paperwork such as activity logs and other reports, plus recorded data from radio communication. (These files and records are not cross-indexed except by the names of the arrestees, although computerized systems could be designed to do this.)

The traffic division shares these characteristics. Little new information is gathered once a call is answered or investigated. This task falls to other officers who deal in both primary and secondary information and are thus less environment dependent than the patrol and traffic segments, more autonomous in processing information, and more entrophic. Juvenile and detective officers are most dependent on primary information passed on to them, and narcotics-vice police are least dependent because a substantial minority of their information comes from their own efforts, or those of their informants. In this way, they resemble internal affairs and research and planning. Administration is most differentiated from the other police segments because the diversity of information is greater there. Although it possesses more general, abstract, and synthesized information, administration work depends upon the information-processing capacities of patrol officers and the realization of those capacities. It is very weakly linked to external information sources.

From an information perspective, primary information is tied most closely to tasks and demands in the environment, secondary information is encoded and related to the flow of knowledge technology within the system and, finally, tertiary information systems perform the integrative function. In some sense, then, the movement of the information through the organization from bottom to top involves a change in both the amount of information and what Donald MacKay has called the meaning of information:

> Information can now be defined as that which does logical work on the organism's orientation (whether correctly or not, and whether by adding to, replacing or confirming the functional linkages of the orienting system). Thus, we leave open the question whether the information is true or false, fresh, corrective or confirmatory, and so on . . . . The meaning of an indicative item of information to the organism may now be defined as its selective function on the range of the organism's possible states or orientation, or for short, its organizing function for the organism.[28]

INFORMATION TECHNOLOGY

MacKay's formulation suggests that the problem of applying new information technology to the police crime quest is not entirely a question of information levels, or of the gathering, processing, or application of information. The problem is rather one of *meaning*, or how information can be used to effect a change in the organizing function.

These distinctions concerning the quality, level, and kind of information are reflected in the final column on technology (table 1). Police technology includes both software and hardware. Perhaps the most common example of hardware is computer machinery and related input systems for processing and producing data such as card-reading machines, tape storage and associated memories, and printers or screens that generate or display output. The various trained personnel who maintain, run, adjust, and repair the machinery and the programs they write to put the data into usable format, the forms, flow charts, and the like, are viewed as software. In practice, software and hardware are inextricable.

This technological capacity is limited by certain aspects of police work. Information that is worked on by members, or processed, is not of the same quality, specificity, and amount across units and persons. Aspects of the social organization of police work ensure that the information available remains located primarily in the heads of officers, and secondarily in their own case files, private notes, or log books. Only under certain specified conditions does it become universally understood, generally shared, and reproducible knowledge held collectively by the organization. The fragmented and personalized state of the knowledge among officers, designed in part to protect them from supervision and discipline, is nowhere systematized into a uniform, comprehensible set of data files that are cross-referenced, retrievable, shared, in universal format, and available to all officers. Thus:

> One consequence of this state of affairs is that even the most advanced among our police departments are not anywhere near the objective of developing adequate information storage and retrieval systems. Even if the present efforts to make use of modern electronic technology would succeed in coding existing information, this would not encompass the knowledge that is currently neither shared nor recorded.[29]

The social organization of police work adds to the asymmetrical nature of primary, secondary, and tertiary information flow. It does so for the following reasons. The basic and fundamental knowledge of policing is believed to be the facts at hand gathered on the scene that only the person there can understand fully and in depth. All valid police knowledge is thought to be contextual knowledge, surrounded by unexplicated, tacit assumptions and meanings. The essential knowledge of police work is thought to be substantive, detailed, concrete, temporally bounded, and particularistic in nature. The validity of this type of knowledge is derived from a police belief that police officers should learn and know social life in precisely this detailed and grounded fashion. In fact, the evidence suggests that most officers do not possess this rich knowledge, in part because they do not seek it; in part because they are not trained to do so; in part because they are not systematically rewarded for being active, observant, and interested; and in part because they rarely—at least in large cities—have the opportunity to spend a great deal of time in the same area, with the same partner, during the same period of the day.[30]

Often it is worthwhile for the officer to retain information for personal use. Information is rarely shared: ". . . police departments accommodate a colossally complicated network of secret sharing, combined with systematic information denial . . . the overriding rule (is) that no one tells anybody else more than he absolutely has to."[31] The reporting system thus becomes a parody of its intended controlling function, and it tends to include material that is not vulnerable to misinterpretation or is nonproblematic, rather than items that are rich in information, problematic, and interpretable when combined with other information. Wilson observed that the reporting system in police departments is "not designed to insure that the problem will be solved (often it cannot be) but to protect the department against a charge that it 'did nothing.' "[32] The several kinds of files listed in table 1 are not cross-referenced, except in the heads of patrol officers. No system articulates, for example, detective or juvenile records; the narcotics information files on nicknames, modus operandi, informants, or cases; departmental arrests and charges; police dispatch tapes; phone messages files; and detectives' case files.

THE POLICE AND THE ENVIRONMENT: OPERATIONAL STRATEGIES

These previous conceptualizations are largely based upon the ways in which information flows into the police department, through various segments, creating and shaping demand. If one turns the analytic focus and asks how the police using this information interact with the environment, one has to draw other distinctions. Although we have argued that the determination of organizational structure by the environment is unproven, some patterning comes as a result of the nature of the events with which the police are routinely asked to deal. Primary information has as its root source citizens, and thus citizen cooperation is the fundamental basis of all police work, and of most of their interactions with the environment. The mandate also has been defined in such a way that the police are required also to anticipate or deter some kinds of events, and this is one of the bases for the accumulation of secondary and tertiary information. When the police react to citizen demand, they are directed in large part by citizen information; when they create demand, or act on the basis of secondary or tertiary information, they are engaging in *proactive* policing.[33] They may also act preventively. Each of these can be considered a strategy by which the police attempt to control or interact with the environment. These types of policing are considered *operational strategies*, and can be classified by the time at which the police intervention takes place, the aim of the intervention, the place of the agent in the process, and the relevant technology (see table 2). While table 1 involves a general classification of policing by functional units, the strategies shown in table 2 are analytic modes that cut across organizational segments. The two presentations should be considered complementary ways of tracing the possible relationships between police functions, information, and technology.

The *preventive strategies* of the police are considered tangential to achieving their unspecified aims, and police community relations units and crime prevention units are thought of as low status, marginal, and ancillary to policing. They function without clear goals, direction, purpose, and supervision.[34] Little if any *prospective intelligence* is gathered by community relations and crime prevention officers because there is no theoretic conception of the causes of crime and therefore the type of information required is not defined. *Applied intelligence* may be used to mount advertising campaigns, to carry out Operation Identification campaigns (where police offer to mark valuable property with an identification number on file with the police), or to mobilize burglary prevention operations in areas where there are perceived crime problems. *Retrospective intelligence* stands in isolation from operations. In effect, crime prevention units provide no new primary or secondary information.

The *proactive strategies* of the police differ from the preventive and reactive strategies in the time of the crime, the aim of the strategy, the place of the agent, and the role of technology and information. In proactive policing, the police actively generate information and attempt to create the conditions of arrest (as in making drug buys that are later used as the basis for making an arrest for the sale of an illicit substance) or to intercept crime via "stakeouts," saturation operations, or road blocks. Prospective intelligence gathering is used prior to the commission of a crime, and usually after information is received from informants or undercover operatives. Retrospective intelligence is also used, in that once a crime is known to have been committed and is either written up in case files or arrest records, it becomes a data file for identifying potential perpetrators of newly discovered crime. Applied intelligence, although used in proactive policing, has its most frequent use in reactive work.

In the case of crimes detected or viewed by the police as a result of reactive strategies, or brought to their attention by citizens, technology serves different functions and has quite different capacities. The crimes at issue come to the attention of the police after their commission is known: this includes the vast majority of crime, both personal and property offenses, as well as much traffic and juvenile crime. The purpose of technology in these kinds of crimes is to organize and systematize the data received into a workable format. These sorts of operations are typically constituted by statistics gathering: mapping and charting crimes as to time, location, amount of property involved, and type of crime by district and by unit or person(s) responsible for investigation. Little attempt is made to use the information in a predictive, preventive, or evaluative fashion.

Dominating reactive strategies is the role that technology plays in organizing information. Once a possible crime has been identified and the police have information as to its content, personnel, location, and consequences, both prospective and retrospective intelligence are used. Typically, prospective intelligence is gathered through informants, confessions, or clearance of someone who admits to previous known crimes, or by detectives' interviews of persons involved. The use of latent fingerprint or voiceprint information to establish suspects is largely a media myth, given the enormous number of data points that would have to be checked to match a given print against a file of prints. More sophisticated data processing can contribute little to the transforming processes necessary to link passive information (files, records, documents, messages) with a crime and thus convert it into active, or suspect-relevant, information.[35]

A second use of technology in reactive police strategies is retrospective in the sense that it makes available to patrol officers information on other past crimes by suspects currently under investigation—e.g., outstanding warrants, the most important of this sort of information. Stolen vehicles, license plates, and driver's licenses and criminal records are among the systematized types of data.[36] For situations where an officer intervenes and is indecisive about the status of the suspect with regard to arrest or detention, such computerized systems may assist decisionmaking. They will provide the necessary information that will allow the police to hold and perhaps charge a person who might otherwise not be questioned or arrested and charged with a crime or violation for which a warrant has been issued. These arrests represent small, but important, increments for police departments. Traffic warrants, once computerized, can be used effectively in making arrests and in generating income and have thus gained popularity in some cities.[37]

APPLICATIONS OF TECHNOLOGY

The effects of technology on the police efforts at crime control are patterned by policing segments, activities, dominant strategies, and information types, as well as by the kind and level of technology. Table 2 shows that technological innovations in units where primary information predominates are found primarily in patrol. In segments where secondary information is predominant, technological innovations are less common and more difficult to characterize. In units where tertiary information predominates, no published social research suggests that increases or decreases in crime control are the result of technological innovations. These generalizations are tentative, as the review below will show, in part because the measurement of crime itself is indirect and sometimes unmeasured, and the impact of police action must be inferred from other data.

Primary Information

In research on units dependent on primary information, several assumptions have been made about the relationship between crime and police activity. Three of these assumptions are that reduced police response time, once a call is received, will decrease citizens' fears; that increases in the visibility of vehicles will increase apprehension rates; and that police presence will increase the probability of apprehension. These assumptions have been addressed in recent research using semiexperimental methods.

TABLE 2 "TYPES OF POLICE OPERATIONAL STRATEGIES" REMOVED DUE TO COPYRIGHT RESTRICTIONS.

RESPONSE TIME

An assumption of the task force report on crime and technology, issued by the President's Crime Commission (1967), was that there were at least three ways to deter crime: rehabilitating identified offenders, removing such offenders from society, and deterring potential offenders. The report argued that increasing the apprehension capability of the police increased the risk/benefits ratio for criminals, which served to deter crime. It was further argued, on the basis of preliminary data from Los Angeles, that "the chance of arrest was appreciably higher when the police responded more rapidly—62 percent for response of one minute and only 44 percent for all cases of 14 minutes or less."[38] This suggests that the means of reducing response time could be a significant factor in increasing the apprehension capacity and, by inference, the deterrence capacity of the police. A number of alternative means to reduce response time have been considered, including setting priorities for types of calls,[39] using civilian community service officers to respond to calls not thought to require sworn officers;[40] and trying to equalize workload, miles traveled within precincts, patrol frequency, average travel time to calls, and time devoted to crimes occurring outside precincts.[41]

Research has shown some results that link response time to increased apprehension.[42] According to Chaiken, "the fraction of responses that resulted in an arrest is a constant plus an exponentially decreasing function of response time."[43] Other studies did not yield these results.[44] Bieck's studies of response time set out to analyze the contributions of each of three components of response time to the total: time between event and citizen call, processing time within the department, and time between dispatch and arrival on the crime scene. This researcher found in a series of careful studies in Kansas City, Mo., that more time elapsed between the recognition of events by citizens (according to their recollections) and a call to the police than between the receipt of the call and processing and between the processing time and arrival of the vehicle on the scene. The citizen decision and the time elapsed *prior* to notifying the police are much more important in determining the total time elapsed than the police response. Using observers, citizen interviews, and dispatch tapes, Bieck found that rapid response time produced an arrest in only 3.7 percent of the Part I (serious crimes). This study concludes: "Because of the time citizens take to report crimes, the application of technological innovations and human resources to reduce police response time will have negligible impact on crime outcomes."[45] Further, the number of arrests effected by more rapid travel time is small. The costs of attempting to reduce response time, including accident risks, computer costs, vehicle maintenance, and public dissatisfaction, were they to raise public expectations and fail, might not be balanced by benefits. Thus there is no evidence that rapid response acts as a deterrent to future or contemplated crime.

There are other research problems. As Chaiken writes, officers may race to calls knowing a suspect is being held or is on the scene, thus making shorter time for arrests spurious.[46] These findings also are supported by those of Larson, Colton, and Larson from research that tested the effect on police operations of computer dispatch, prioritized calls and automatic vehicle locators (AVMs), which transmit a signal shown on a central dispatch screen and facilitate on-line tracking of vehicles. These researchers asked if such systems would reduce response time, improve command and control capacities, increase officer safety, and reduce voiceband congestion. The reduction in response time amounted to about 6 percent overall, and they conclude: "If the St. Louis M.P.D. is interested in average response time reductions, it should concentrate on aspects of the police response system that are not directly related to A.V.M." In regard to command and control, no clear evidence was derived, but the study showed that AVM did not produce variation in the response time for given priority calls of levels 1, 2, and 3. The study concluded overall that "AVM may potentially be useful in improving the efficiency and effectiveness of the patrol force, but this aspect of the system must still be tested."[47]

When these findings are combined with the data from victimization surveys[48] that show roughly 50 to 70 percent of all crime goes unreported, and with data showing that because of legal limits and the invisible or private nature of much crime, about half of all crime is highly resistant to any effect of policing,[49] some limits on the police impact on crime suppression are revealed. Control of the police through calls, control of information given, space and time limits, and the tendency of the police to ratify the citizens' expressed desires in crime incidents[50] means that lowering response time may not yield significant crime control results.

PRESENCE AND VISIBILITY

As James Q. Wilson once wrote, the police are not clear whether their aim in crime control is to concentrate crime in one area, or to disperse it equally across all areas.[51] Models of allocation have been of primary concern to researchers who have explicitly attempted to disperse crime, at least insofar as police presence serves to deter or increase the probabilities of apprehension by fortuitous placement.[52] These experiments, or models of allocation, were known by the acronyms LEMRAS (Law Enforcement Manpower Resource Allocation System, upon which other models were based) and PCAM (Patrol Car Allocation Model), the latter designed by Chaiken and Dormant. After noting the use of these models to determine staffing levels, the allocation of officers to geographic areas and shifts, and the hours at which shifts should begin, Chaiken and Dormant note that such models are used by only a handful of departments and that no department uses all their capacities on a regular basis. Further, they note that for simulation models which examine the effects of changing dispatch rules, or setting up a car locator system, of specializing vehicles to functions, and of reallocating across precincts, "no example of sustained use of simulation models was found, nor did we find a single instance of important decisions made based on the output from such a model."[53] Beat queuing of calls, linear programming for scheduling patrol cars, and manpower scheduling models are rarely used, and when used (as in the case of beat design), their aim is workload equalization.

Other experiments with police presence can be examined historically, beginning with the Kansas City Preventive Patrol study carried out by the Kansas City, Mo., Police Department and the Police Foundation.[54] The importance of this study is that it specifically tested the assumption of technological innovation, made in the technology task force report of the President's Crime Commission (1967), that increased police presence would increase arrests (apprehension capacity) and, by implication, deterrence. Using an experimental design, three conditions were employed. One condition, the control, maintained patrol at previous levels; the second had patrol set at two to three times the previous level, while a final condition permitted cars to enter the area only on calls for assistance. The last condition was intended to have no preventive patrol. The results showed that citizen satisfaction and victimization rates, measured by survey interviews, were not statistically different across conditions. The amount of reported crime was also essentially equivalent. It was found that in 60 percent of the time in three conditions, officers were "driving around," and in only 19+ percent of the time were they acting in a strictly "crime-related" or proactive capacity, such as checking suspicious circumstances or investigating crime.[55]

Several analysts conclude that routine preventive patrol (the time left over after crime-related activities and administrative duties are subtracted and a proportion of time that varies from study to study of police workload), is largely "unproductive" by the measures employed.[56] Cordner's excellent review of the content of workload studies suggests that 30 to 70 percent of all calls are *not dispatched* but are handled internally by dispatchers and operators, thus showing the enormous internal discretion of the department. Only a small proportion (about 13 percent) is narrowly defined, crime related, or even related to law enforcement. A small amount (4+ percent) is narrowly defined and not crime related; most of the calls are "ambiguous in nature." Half of patrol time is "driving around time" and another 10 to 20 percent is spent on nonduty activities such as "girl watching" and "taking breaks." Finally, a highly variable but important proportion of time is spent in administrative duties (10 to 50 percent).[57]

These data show that if there is to be any increment in the effect of police on crime, it must result from direction or *structuring* of the time that presently comes under driving around, administrative duties, and nonwork-related times. Although it is a substantial proportion of time, the discretion of the officer is such that only innovations of a major sort, such as the Wilmington, Del., split-force experiment or the Rochester, N.Y., team policing idea (see below) will serve to constrain officers.[58]

The Wilmington split-force experiment in policing divided the patrol force into two units, a *structured*, or directed, patrol and a *basic* patrol force. The first was to be crime focused, would make immediate crime-related responses, and would do directed, problem-oriented patrol, while basic patrol handled other calls. The results of a year-long experiment in Wilmington showed that overall reported crime decreased in the city slightly (6.1 percent), but the number of crimes cleared also decreased, in spite of an increase in patrol clearances (−105.5 percent). The reason was that detective clearances dropped dramatically (−61.4 percent), causing an overall drop of some −28 percent in the total department's clearance rate. Further, the actual number of arrests overall dropped in spite of these new techniques, in part because the monthly average number of arrests made by detectives decreased more than 50 percent, while patrol arrests increased very slightly on the average of fewer than one a month.[59]

This study illustrates the dilemma of many studies of police performance and the crime-related effects of technological innovations. As the authors state, the aim of the study was not to test the effects of the experiment on crime reduction, changes in clearance rates, or citizen satisfaction (although these data were gathered), but to test a concept of allocation. The irony is that crime was reduced during this time. Clearance rates and arrests dropped overall, as did charges made on the basis of these arrests. These changes cannot be attributed to experimental effects.

Other deployment experiments using knowledge technology suggest that (a) police deterrence of crime is not clearly established, although there are suggestions that more aggressive patrol tactics may produce increases in the apprehension rate;[60] (b) some "tipping effects" occur and arrest rates diminish, given increases in manpower;[61] (c) displacement effects, or the movement of crime from one locale to another as a result of police pressure, do seem to occur for residential burglary;[62] (d) saturation patrols may produce temporary crime-repression effects;[63] and (e) personnel increases per se may produce increased police presence. However, not all officers are on patrol, deployment patterns differ, and arrest rates are differentially produced by police segments and officers.[64] Research suggests that a deterrent effect, as measured by arrests made, is produced by personnel increases.[65]

These experiments involving software modifications in technology are fraught with methodological and logical errors, making claims for police effects questionable. The first error is the confounding of the arrest rate and the crime rate, and the subsequent attributing of deterrent effects to arrests. Since arrests seem to be at least partially a response to resources and resources rise following a rise in crime, there are few examples of resources producing an increase in arrests. This means that "the effect of total police resources on crime rates is not therefore understood."[66] Further, even research such as Wilson and Boland's, which used simultaneous equations for comparing arrest rates and victimization rates, is cross-sectional and does not show longitudinal effects.[67] Secondly, the units of analysis vary and city size seems to pattern the effect of resources on crime.[68] The actual proportion of time spent on crime work varies from city to city, and research has not controlled for this fact, nor have the arguments for a deterrent effect investigated the actual circumstances and events that lead to an arrest. Clearly, as Black, Clark and Sykes, and Pepinsky have shown, the interactional context is perhaps the single most important basis for arrest.[69]

The methodological problems of such studies are a function of their strengths in other respects. By aggregating the statistics of a number of departments, researchers may obscure the difference in strategies in different departments so nicely inferred by Reppetto.[70] On the other hand, detailed case studies may be too sensitive to idiosyncratic features of given departments, their history and traditions. Studies based on crime-related performance measures use official statistics, and treat the internal processing of information and the actual occurrence or situation of arrest as a given. As Wilson and Boland write, for example: "None of these studies of the impact of police presence made a significant attempt to monitor what the police actually did."[71] From this perspective and a general methodological review of findings, one can conclude that the precise effect of technology on police behavior, given primary information, is still unclear.

Studies that attend to the introduction of knowledge technology which might increase police effectiveness—primarily in terms of retrospective and prospective intelligence gathered and made available by crime analysis units in support of patrol—show that these units are not well used. Reinier et al., on the basis of a large study of 23 departments' use of crime analysis units, found that the equipment was viewed with suspicion by officers, and though they generally thought that crime analysis might be useful, police felt that crime analysis does not "contribute significantly to meeting on-going problems of allocation and deployment of patrol resources."[72] Furthermore, the more formalized the modes of analysis (e.g., involving computer analyses), the less they are used. In effect, the more likely that the analysis is precise, abstract, general, formal, and conceptual, the greater potential it has for shaping effective decisions according to information theory, and the less likely it will be used by officers. These conclusions are consistent with the previous review of models by Chaiken.[73]

Secondary Information

The most important technological innovations bearing on processing and using information are models of case screening employed by detective and juvenile units. These are basically software changes in operations technology, although they also have knowledge technology implications.

Three recent studies focus on the effectiveness of detectives, and technology is one of the variables studied. The most significant of these studies is the work of Greenwood, Chaiken, and Petersilia, which attempted to describe the detective function and to evaluate the effectiveness of it, as measured by case clearances.[74] A clearance is a crime for which the police have decided that sufficient reason has been established to state that a case is cleared.[75] More specifically, once a case is referred to the detective division and screened for investigation, an officer can dispose of the case in a number of ways—e.g., by arrest, by confession, by issuing an outstanding warrant, by transferring the case to another jurisdiction, etc.[76]

The number of cases cleared is one of the common measures of the worth of a detective. Using a questionnaire sent to 300 large urban departments (153 returned the questionnaire) and observations in 25 sites, the three studies concluded that differences in training, staffing levels, workload, or investigative procedures had no significant effects on arrest or clearance rates; that specialized units had little effect on arrest rates; and that the organization of units had little impact on arrest or clearance rates. They found that 64.8 percent of serious crimes received only superficial attention, and that only about half of all cases were investigated at all. Cases were solved primarily as a result of information given to officers on the scene by the victim; this was the single most important determinant of whether an arrest will be made.

A study of what has come to be called the Rochester System was undertaken at about the same time, and used the idea of a team of patrol officers (around 30), four investigators, patrol sergeants, and a patrol lieutenant. Although before-and-after measures are not available, team and nonteam groups were compared, and the teams made more arrests; cleared a larger proportion of burglaries, robberies, and larcenies; and made more on-scene arrests than did nonteam personnel.[77]

Using screening approaches to assign investigators to cases, two recent studies have shown that if elements of a case (e.g., whether a witness is present, whether physical evidence was found on the scene, how long it has been since the crime was committed) are weighted and a cutoff point is established, much time can be saved by screening out cases, without significant loss of cases that might be cleared.[78] In general, this model will serve to eliminate cases that experience shows will not be solved. It should be noted, however, that all these studies use clearance rate as a dependent variable, and that they are retrospective studies of cases already processed by officers. This means that the definition of what a clearance is varies from department to department studied and that other measures of police effect are not gathered. As Kelling has noted, ". . . with the exception of the Rand study, which did attempt a modest analysis of how investigators spend their time, the studies provide no linkages between what investigators do and the ultimate outcome of the case. In essence, all are 'black box' studies which look at the relationships between input (detectives' time) and output (clearances)."[79] We have no idea how those outcomes are produced, nor how the cases are themselves founded in the various departments (i.e., legitimated as crimes to be investigated). It can only be inferred on the basis of the Greenwood, Chaiken, Petersilia study that, in general, knowledge (e.g., investigative strategies and resources technology) does little to increase the clearance rate but may reduce time invested in cases that will not produce clearance results.[80]

There is evidence that where police action in secondary information-dependent units depends on transactions across organizations, the police attitude toward the agency involved and the number and kinds of formal linking mechanisms available more strongly determine information flow than the technology used to gather, assemble, or analyze the data involved.[81] It is likely that closer links between subsystems within the criminal justice system tend to increase prosecutions for vice and narcotics crime, at least on the basis of preliminary evidence.[82] Other studies of juvenile units do not explicate the links that produce additional arrests for juvenile crime but suggest that a professional sort of police department, with higher levels of formal education and training, increases the level of arrest when two socially comparable cities are studied.[83]

Research on technological innovations in police work carried out by systems theorists and engineers very often has relied on a model of police organization that uses the analogy of the computer. They may therefore underestimate the effects of segmented work, the different focuses of activities, and the primary, secondary, and tertiary information distinction. New technology can only be used to transform or systematize available information; strategies of policing set constraints on the impact of innovation. New technology must function within the previous social organization, meanings, praxis, and values of the groups into which it is introduced. New technology produces no narrowly predictable effects because it is grafted onto established lines of social organization and information flow. There is a pattern of resistance to new innovations, as the perceptive Larson report on the St. Louis AVM project candidly revealed.[84]

There is little awareness in police research of parallel research on the introduction, for example, of automation into banks and assembly lines.[85] From the perspective of public opinion and the consequences of increased technology, there is serious doubt whether, looking at the wide range of police performance measures—citizen satisfaction, citizen cooperation, reduction in charges of brutality, etc.—technology will have any appreciable effect.[86] It would appear, in fact, that motorized patrol and computerized dispatch may reduce the quality of police citizen contact.[87] If one examines conventional measures of police performance, such as crime control or suppression indexes, the picture is not very clear. As Colton has remarked:

> Not unlike other non-technical innovations, it is very difficult if not impossible to evaluate a police technological innovation on crime impact alone. The failure of allocation and command and control projects to demonstrate clear improvements in police patrol performance have already been pointed out. Perhaps a greater failure was the original expectation that was set up in the 1960's, that we might be able to establish such linkages. *Criminal activities are based on a wide range of factors, only a small portion of which are influenced by police activity* (emphasis added). Changes in deployment patterns or response rates may have some modest influence, but criminal statistics are far too imprecise to measure these differences or isolate the portion of the changes attributed to police allocation or technology as opposed to changes, for example, in the weather or in the unemployment rate.[88]

Simple random variation in rates, as well as variations in crime reporting patterns, are greater than the amount of variation explained by any known police practice, with or without technological innovations.[89] The maximal effect of police is still minuscule when compared to the effects of social values, personal attitudes, class interests, and informal modes of social control.[90] In other words, the precise effects of technology on crime cannot be measured accurately at this time.

CONCLUSIONS

Assessing the general problems and prospects of technological innovations in crime control, one can draw several tentative conclusions. The work on crime and its relationship to police activity is not theoretically grounded.

The various functions of the police must be seen in concert and in conflict, rather than seeing all operations as complementary. They are not. For example, Los Angeles instituted a computerized dispatch that would send the nearest car to the scene of a crime, regardless of the precinct boundaries, whereas another system of assignment emphasized the importance of community knowledge in crime prevention. The two issues worked at cross-purposes, one emphasizing rapid response and the other community contact.[91] Logically, other innovations work in the same fashion. For example, drug enforcement drives involving a large number of hand-to-hand buys and sweeps to "eyeball" and arrest persons in a selected area work against attempts to create good community relations because they stir up anger and resentment and suggest repressive, or at least discretionary, enforcement tactics. Similarly, case screening may focus attention on cases viewed as "solvable," while other positive consequences of investigation, such as victim reassurance, information on other activities in the neighborhood, and establishing police presence, are lost. The resultant trade-offs, since they are complex and value laden, are not weighed by changes in procedures. The research on technological innovation and crime control uses rather crude indexes of crime control and may induce a rather narrow view of the implications of the work.

The argument that general technological advances should be reflected in the practice of contemporary policing is unassailable. However, for a number of reasons, the police are both vulnerable to those innovations and unable to judge their value themselves. Some research designed to test the effects of technology is not sociologically informed. The performance measures chosen in that research were based on official police data on crime. The relationship between such data as arrests and charges and the crime rate or victim rate is clouded by interactions between independent and dependent variables, shadow effects (rate changes not produced by the police), time-sequence problems, and the inability of researchers to control exogenous variables such as weather, changes in insurance practices, and the reporting behavior of citizens.[92]

Police organization is treated isomorphically with predictable and nonproblematic information flows, and with the nonadditive effects of changes in information levels within police segments. External constraints on police actions to control crime—such as legal and procedural guarantees, citizen reporting practices, the diversity and private nature of much crime, the nonrational or unpredictable nature of some crime events—all act to make police effects marginal, and technological effects are a small but undetermined portion of those effects. Social research will meet resistance, and police resistance does not indicate that the innovations are inappropriate or should be abandoned. Looked at in the perspective of the coming decades, the bare outlines of research needs are just being set. Only in the future will the precise effects of technology on crime be established.

NOTES

1. Goldstein, H., *Policing A Free Society*, Cambridge, Mass.: Ballinger, 1977, Chapter 2.

2. Reiss, A. J., and Bordua, D., "Organization and environment: a perspective on the municipal police," In Bordua, D., ed., *The Police*, New York: Wiley, 1967, pp. 25-55.

3. Manning, P. K., "Police lying," *Urban Life and Culture* 3:283-306, 1974. See also Marx, G., "Thoughts on a neglected category of social movement participant: the agent provocateur and the informant," *American Journal of Sociology* 80:402-442, 1974.

4. Manning, P. K., "Dramatic aspects of policing," *Sociology and Social Research* 59:21-29, 1974. See also Manning, P. K., *Police Work*, Cambridge, Mass.: Massachusetts Institute of Technology Press, 1977.

5. Muir, W. K., Jr., *Police: Street Corner Politicians*, Chicago: University of Chicago Press, 1977. See also Furst, B., Lucianovic, J., and Cox, S., *What Happens After Arrest?* Washington, D.C.: Institute for Law and Social Research, 1977.

6. Manning, P. K. and Van Maanen, J., eds., *Policing: A View From The Street*, Santa Monica, Calif.: Goodyear Publishing Co., 1978.

7. Wilson, J., *Varieties of Police Behavior*, Cambridge, Mass.: Harvard University Press, 1968. See also Wilson, J., and Boland, B., "The effect of the police on crime," *Law and Society Review* 12:367-390, 1978; Reppetto, T. A., "The influence of police organizational style on crime control effectiveness," *Journal of Police Science and Administration* 3:274-279, 1975.

8. There are at least 10 different meanings of the term "crime focused" as it refers to policing. One of the purposes of future research would be to clarify the definitions and then select appropriate measures.

9. Mandel, J., "Problems with official drug statistics," *Stanford Law Review* 21:991-1040, 1969. See also Robinson, S., "A critical view of the Uniform Crime Reports," *Michigan Law Review* 64:1031-1054, 1966; Gould, L., "Crime and its impact on an affluent society," In Douglas, Jack D., ed., *Crime and Justice in American Society*, Indianapolis, Ind.: Bobbs-Merrill: 81-118, 1971: Kitsuse, J., and Cicourel, A., "A note on the uses of official statistics," *Social Problems* 11:132-139, 1963; Hood, R., and Sparks, R., *Key Issues in Criminology*, New York: McGraw-Hill, 1970; Seidman, D., and Couzens, M., "Getting the crime rate down: political pressure and crime reporting," *Law and Society Review* 8:457-493, 1974; Milakovich, M., and Weiss, K., "Politics and measures of success in the war on crime," *Crime and Delinquency* 21:1-10, 1975; Defleur, L., "Biasing influences on drug arrest records: implications for deviance research,"

American Sociological Review 40:88-103, 1975; Manning, P. K., *The Narcs' Game: Organizational and Informational Constraints on Drug Law Enforcement*, Cambridge, Mass.: Massachusetts Institute of Technology Press, forthcoming.

10. Reiss, A. J., Jr., and Biderman, A. D., "On exploring the 'dark figure' of crime." *Annals* 374:1-15, 1967.

11. Many of these summary points are made in the works of Wilson, J. (1968 and 1978, see refs. 7 and 21). See also Bittner, E. (1970, see ref. 29); Reiss, A. J., Jr. (1971, see ref. 26); Clark, J. and Sykes, R. (1974, see ref. 69); and Manning, P. K. (1977, forthcoming, see ref. 9) on the American police. Similar themes are emphasized in comparative and British studies by Banton, M. (1964, see ref. 90); Cain, M., *Society and the Policeman's Role*, London: Routledge and Kegan Paul, 1973; Holdaway, S., ed., *British Police*, London: Edward Arnold, 1979; Punch, M., *Policing the Inner City*, London: Macmillan, 1979.

12. Fogelson, R., *Big City Police*, Cambridge, Mass.: Harvard University Press, 1977. See also Reppetto, T. A., *The Blue Parade*, New York: Free Press, 1978.

13. Manning, P. K., "The social control of police work: observations on the occupational culture of policing." In Holdaway, S., ed., *British Police*, London: Edward Arnold, Ltd., 1979.

14. Delany, W., "The development and decline of patrimonial and bureaucratic administrations," *Administrative Science Quarterly*: 458-501, 1966. See also Manning, P. K. (forthcoming, see ref. 9).

15. Peabody, R. L., "Perceptions of organizational authority: a comparative analysis," *Administrative Science Quarterly* 6:463-482, 1962.

16. Stinchcombe, A., *Creating Efficient Industrial Organizations*, New York: Academic Press, 1974.

17. Weick, K., "Educational Organizations As Loosely Coupled Systems," *Administrative Science Quarterly* 21:1-19.

18. Colton, K. W., "The diffusion of police computer technology." In Colton, K. W., ed., *Police Computer Technology*, Lexington, Mass.: D.C. Heath, 1978, pp. 273-280.

19. Zey-Ferrell, M., *Dimensions of Organizations*, Santa Monica, Calif.: Goodyear, 1979.

20. Perrow, C., *Organization Analysis*, Belmont, Calif.: Brooks/Cole, 1970.

21. Manning, P. K. (forthcoming, see ref. 9). See also Thompson, J., *Organizations in Action*, New York: McGraw-Hill, 1967; Wilson, J. (1968, see ref. 7); Wilson, J., *The Investigators*, New York: Basic Books, 1978.

22. Van Maanen, J., "Observations on the making of policemen," *Human Organization* 32:407-418, 1973.

23. Weick, K., *The Social Psychology of Organizing*, Reading, Mass.: Addison-Wesley, 1969. See also Weick, K. (1976, see ref. 17); Child, J., "Organizational structure and performance: the role of strategic choice." *Sociology* 6:1-22, 1972; Aldrich, H., and Pfeffer, J., "Environments of organizations," In Inkeles, A., Coleman, J., and Smelser, N., eds., *Annual Review of Sociology* 2:79-106, 1976; Manning, P. K. (forthcoming, see ref. 9).

24. Zey-Farrell, M. (1979, see ref. 19), p. 108.

25. Goldstein, H. (1977, see ref. 1), p. 35.

26. Reiss, A. J., Jr., *The Police and the Public*, New Haven, Conn.: Yale University Press, 1971, p. 11.

27. It is important to note in this context that other more "invisible" factors might create demand. For example, in some cities the mayor's office or the city council has dictated the priority system for handling calls.

28. MacKay, D., "The informational analysis of questions and commands." In Cherry, C., ed., *Information Theory: Fourth London Symposium*, London: Butterworth and Company, 1961, pp. 470-471.

29. Bittner, E., *The Functions of the Police in Modern Society*, Washington, D.C.: U.S. Government Printing Office, 1970, p. 67.

30. Reiss, A. J., Jr. (1971, see ref. 26), p. 5.

31. Bittner, E. (1970, see ref. 29), p. 64.

32. Wilson, J. (1968, see ref. 7), p. 70.

33. Black, D., "Police encounters and social organization." Unpublished Ph.D. dissertation, Department of Sociology, University of Michigan, 1968.

34. Norris, D., *Police-Community Relations: A Program that Failed*, Lexington, Mass.: D.C. Heath, 1973. See also Kreps, G., and Weller, J., "The police-community relations movement," *American Behavioral Scientists* 16:402-412, 1973.

35. Willmer, M. A. P., *Crime and Information Theory*, Edinburgh: Edinburgh University Press, 1970.

36. J. Murphy in 1974 estimated that there were 2 500 computer terminals in police cars and that by 1983, if present trends continue, half the 75,000 police cars in the United States will be so equipped. In addition, there will be growing use of hand-held or small computer terminals by police officers on scooters, on foot, and in vehicles.

37. Colton, K. W. (1978, see ref. 18).

38. Blumstein, A., "Systems analysis and the criminal justice system," *Annals* 374:92-100, 1967, p. 96.

39. The Birmingham, Alabama, Police Department, with the assistance of the Police Executive Research Forum, is developing models of types of calls and appropriate strategies for response to them. This report to the National Institute of Law Enforcement and Criminal Justice, LEAA, should be completed by late 1979.

40. These CSOs are being used in a number of cities, including Madison, Wis., New Rochelle, N. Y., and Birmingham, Ala.

41. Ferreira, J., Jr., "Comparing patrol unit allocation methods." In Larson, R. C., ed., *Police Deployment*, L.R.P. Vol. 1, Lexington, Mass.: D.C. Heath, 1978, pp. 183-217.

42. Isaacs, H., "A study of communication, crimes, and arrests in a metropolitan police department." In *Task Force Report: Science and Technology*, President's Crime Commission, Washington, D.C.: U.S. Government Printing Office, 1967. See also Clawson, C., and Chang, S., "The relationship of response delays and arrest rates," *Journal of Police Science and Administration* 3:53-68, 1977.

43. Chaiken, J., "Deterrent effects of police activities." In Cramer, J. A., ed., *Preventing Crime*, Beverly Hills, Calif.: Sage Publications, pp. 109-135.

44. Brown, W., "The evaluation of police patrol operations." Unpublished thesis, University of Ottawa, 1974 (cited in Chaiken, 1978, see ref. 43).

45. Bieck, W., *Response Time Analysis: Executive Summary*, Kansas City, Mo.: Kansas City Police Department, 1977, p. 25.

46. Chaiken, J. (1978, see ref. 43), p. 130.

47. Larson, R., Colton, K., and Larson, C., "Evaluation of a phase I implementation of an automatic vehicle monitoring (AVM) system in St. Louis," In Colton, K. W., ed., *Police Computer Technology*, Lexington, Mass.: D.C. Heath, 1978, pp. 243-267.

48. Skogan, W., "Citizen reporting of crime: some national panel data," In Guenther, A. L., ed., *Criminal Behavior and Social Systems*, 2nd ed., Chicago: Rand McNally, pp. 131-148.

49. Skogan, W., "The promise of policing: evaluating the performance, productivity, and potential of local law enforcement." Presented to Workshop in Policy Analysis in State and Local Government, SUNY-Stony Brook, N.Y., May 1977.

50. Black, D., "The production of crime rates," *American Sociological Review* 34:733-747, 1970.

51. Wilson, J., "Dilemmas of police administration," *Public Administration* 25:407-417, 1968.

52. Larson, R. (1978, see ref. 41).

53. Chaiken, J., *Criminal Justice Models: An Overview*, LEAA/NILECJ, Washington, D.C.: U.S. Government Printing Office, 1976.

54. Kelling, G. et al., *The Kansas City Preventive Patrol Experiment, Summary Report*, Washington, D.C.: The Police Foundation, 1974.

55. This research has been extensively criticized and continues to

be controversial in the police world. See Larson, R., "What happened to patrol operations in Kansas City?" *Journal of Criminal Justice* 3: 267-298, 1975.

56. Cordner, G., "Police patrol work load studies: a review and critique," Unpublished paper, School of Criminal Justice, Michigan State University, 1979. Reviews of the relevant literature on studies of types of patrol using various measures of productivity (most of the studies were so inadequate as research that the findings were, strictly speaking, noncomparable) show that routine patrol tactics do not produce benefits equivalent to costs involved. (See Gay, W., *Improving Patrol Productivity*, Vol. I: *Routine Patrol*, Washington, D.C.: U.S. Government Printing Office, 1977; Schack, S., *Improving Patrol Productivity*, Vol. II: *Specialized Patrol*, Washington, D.C.: U.S. Government Printing Office, 1977; Farmer, D., "The future of law enforcement in the United States: the Federal role," *Police Studies* 1: 31-38, 1978.) These studies, since they are focused almost exclusively on crime figures, did not measure such matters as the reassurance function of patrol or noncrime services. (See Bahn, C., "The reassurance factor in police patrol," *Criminology* 12:338-345, 1974; Meyer, J. C., Jr., "Patterns of reporting noncriminal behavior to the police," *Criminology* 12:70-83.)

57. Cordner, G. (1979, see ref. 56).

58. Pate, T., Bowers, R., and Parks, B., *Three Approaches to Criminal Apprehension in Kansas City*, Washington, D.C.: The Police Foundation, 1976.

59. Tien, J., Simon, J., and Larson, R., *An Alternative Approach in Police Patrol: The Wilmington Split-Force Experiment*, LEAA/NILECJ, Washington, D.C.: U.S. Government Printing Office, 1977.

60. Boydstun, J., *Community Profiling and Police Patrol*, San Diego, Calif.: San Diego Police Department, 1974. See also Chaiken, J., Lawless, M., and Stevenson, K., "The impact of police activity on crime: robberies on the New York City subway system," New York: Rand Institute, 1974; Wilson, J., and Boland, B., (1978, see ref. 7).

61. Tittle, C., and Rowe, A., "Certainty of arrest and crime rates: a further test of the deterrence hypothesis," *Social Forces* 52:455-462, 1974.

62. Reppetto, T. A., "Crime prevention and the displacement phenomenon," *Crime and Delinquency* 22:166-177, 1976.

63. Wilson, J., *Thinking About Crime*, New York: Basic Books, 1975.

64. Furst, B. et al. (1977, see ref. 5).

65. O'Connor, J., and Gilman, B., "The police role in deterring crime," In Cramer, James A., ed., *Preventing Crime*, Beverly Hills, Calif.: Sage Publications, 1978, pp. 75-108.

66. Chaiken, J. (1978, see ref. 43).

67. Wilson, J., and Boland, B. (1978, see ref. 7).

68. O'Connor J., and Gilman, B. (1978, see ref. 65).

69. Black, D. (1970, see ref. 50); Clark, J., and Sykes, R., "Some determinants of police organization and practice in a modern industrial democracy," In Glaser, D., ed., *Handbook of Criminology*, Chicago: Rand-McNally, 1974, pp. 455-495; Pepinsky, H., "Police decision-making," In Gottfredson, D., ed., *Decision-Making in the Criminal Justice System: Reviews and Essays*, Washington, D.C.: U.S. Government Printing Office, 1975, pp. 21-52.

70. O'Connor, J., and Gilman, B. (1978, see ref. 65); Reppetto, T. A. (1975, see ref. 7).

71. Wilson, J., and Boland, B. (1978, see ref. 7).

72. Reinier, H. et al., *Crime Analysis in Support of Patrol*, NEP Phase I Summary Report, Washington, D.C.: U.S. Government Printing Office, 1977.

73. Chaiken, J. (1976, see ref. 53).

74. Greenwood, P., Chaiken, J., and Petersilia, J., *The Criminal Investigation Process*, Lexington, Mass.: D.C. Heath, 1977.

75. Wilson, J., and Boland, B. (1978, see ref. 7).

76. Skolnick, J., *Justice Without Trial*, New York: Wiley, 1966, (2nd Edition 1975, with new epilogue), pp. 167-181.

77. Bloch, P., and Weidman, D., *Managing Criminal Investigations*, Washington, D.C.: The Urban Institute, 1976.

78. Greenberg, B. et al., *Enhancement of the Investigative Function*, Menlo Park, Calif: Stanford Research Institute, 1975. See also Eck, J., "Burglary investigation decision model replication: a multisite evaluation," presented to the LEAA National Workshop on Criminal Justice Evaluation, November 1978.

79. Kelling, G., "Working paper on criminal investigation," unpublished, The Police Foundation, Washington, D.C., 1978.

80. Greenwood, P. et al., (1977, see ref. 74).

81. Johnson, K., *Police Inter-Agency Relations: Some Research Findings*, Sage professional paper, Beverly Hills, Calif.: Sage Publications, 1978.

82. Skolnick, J., and Woodworth, J., "Bureaucracy, information and social control: a study of a morals detail," In Bordua, D. J., ed., *The Police*, New York: Wiley, 1967, pp. 99-136.

83. Wilson, J., "The police and the delinquent in two cities," In Wheeler, S., ed., *Controlling Delinquents*, New York: Wiley, 1968.

84. Larson, R. et al. (1978, see ref. 47).

85. Shepard, J., *Automation and Alienation*, Cambridge, Mass.: Massachusetts Institute of Technology Press, 1971. See also Faunce, W., *Problems of an Industrial Society*, New York: McGraw-Hill, 1968.

86. Marx, G., "Alternative measures of police performance," In Larson, R.C., ed., *Police Accountability*, Lexington, Mass.: D.C. Heath, 1978, pp. 15-32. See also Goldstein, H. (1977, see ref. 1), chapter 3.

87. Kelling, G., "The quality of urban life and the police," In Conrad, J. P., ed., *The Evaluation of Criminal Justice*, Beverly Hills, Calif.: Sage Publications, 1978, pp. 101-122. See also Murphy, P., *Commissioner*, New York: Simon and Schuster, 1977.

88. Colton, K. W., "A decade of experience since the crime commission: conclusions and recommendations," In Colton, K. W., ed., *Police Computer Technology*, Lexington Books: D.C. Heath, 1978, pp. 281-289.

89. Gould, L. (1971, see ref. 9). See also Skogan, W. (1976, see ref. 48).

90. Banton, M., *The Policeman in the Community*, New York: Basic Books, 1964.

91. Colton, K. W. (1978, see ref. 88).

92. Chaiken, J. (1978, see ref. 43).

6 Crime Control: Science, Technology, and the Institutional Framework

*by Richard A. Myren**

SUMMARY

Discussion of the role of scientific research and technology in crime control requires an understanding of the context in which our substantive criminal law is created and administered. The most important single factor in control of crime is citizen support of the criminal justice system. A second important factor is congruence between the criminal law prohibitions and the value systems and ideologies of the residents of a jurisdiction. Without congruence, there will be little citizen support. Without citizen support, the most diligent and efficient crime control agencies using the latest results of scientific research and technology will find their efforts thwarted.

There are definite limits on what science, particularly social science, can do for governments. Social scientists must be careful to limit their work to description and to avoid prescription. Despite that limitation, the potential contribution of scientific research to crime control is great. In an attempt to realize that potential, the Federal Government has established the National Institute of Law Enforcement and Criminal Justice (NILECJ). In its first decade, the Institute spent $212,597,000, but not all of that was for research; NILECJ also has a mandate from Congress to carry out related activities. An educated guess is that about $80 million of the budget has gone for research.

Critics maintain that much of the NILECJ budget has been spent unwisely. The problems cited have been primarily in policy and management. Many of these have now been corrected, and Congress seems to be moving toward making NILECJ more independent of other Law Enforcement Assistance Administration (LEAA) operations, another recommendation of the critics. Although NILECJ has the largest Federal budget, other Government agencies support some crime-related research relevant to their primary missions. Private agencies also make their contribution to support of criminal justice research.

Scientific research can contribute to crime control in a number of ways. It is useful in defining the problem of crime and determining its causes; in planning programs believed capable of preventing significant amounts of crime; in establishing policies for the formal governmental reaction of society to crime; in deciding how individuals will be proved to have violated the criminal law; and in planning what to do with persons convicted of crime. Law, philosophy, and the social sciences (anthropology, economics, political science, psychology, and sociology) contribute most heavily to most of these processes. The natural and life sciences and modern technology are useful primarily in demonstrating that crimes have been committed and in identifying the perpetrators.

A society's investment in crime research should be broad based, long range, and cumulative. Support should also be given to studies on research utilization.

* Dean, School of Justice, American University.

Special efforts should be made to increase the research capacities of university programs, which can aid state, regional, and local criminal justice planning units. Such arrangements will assure that research continues as an institution in our society, that the scientific method remains a part of our cultural heritage, and that its potential contribution to crime control is realized.

INTRODUCTION

This essay considers the role of scientific research and technology in crime control, a role that must be viewed in its institutional context. Crime is a complex concept and an even more complex reality. Significant factors that both create the potential for and set limits on major contributions by scientific research and technology to crime control include community support of efforts to contain crime and an appreciation of the role in and impact of ideology and value structure on the success of crime control methods. A third factor is the difficulty in effectively using the knowledge created through scientific research. But first there must be agreement on what scientific research and technology mean in this setting.

DEFINITION OF RESEARCH, DEVELOPMENT, AND TECHNOLOGY

The Task Force on Criminal Justice Research and Development of the National Advisory Committee on Criminal Justice Standards and Goals defined research as "an activity that involves the generation of knowledge that will lead to significant solutions to criminal justice problems and is based on replicable scientific procedures." Research was contrasted with development, which was defined as "an activity that involves the translation of new knowledge into improved criminal justice practices or techniques." The task force found technology to require a more lengthy treatment:

> Hard and soft technologies are labels commonly used in the applied sciences. Hard technologies refer to tangible materials, though the idea also encompasses the physical embodiment of any idea. Soft technologies are more conceptual and analytical, though they can have profound effects on people's lives when implemented. Often innovations contain a mix of both hard and soft technologies. . . . The use of these terms should not be interpreted in the same sense as in hard and soft science, or technology R&D that is analytically hard (rigorous) or analytically soft (nonrigorous). The terms do not connote the same values, and care should be exercised in their use.[1]

Extending the example, both the soft and hard technology in an information system will have been developed by scientific research; the adaptation of the technology to the needs of the criminal justice agency is development. In a very real sense, all technology is the product of applying the results of scientific research.

In discussing the role of scientific research and technology in crime control, it is first necessary to describe the context in which research and technology are used in that effort. First, there is today general agreement among criminologists that there is no single, simple cause of crime. Causation is pluralistic, with elements in both the individual and the environment.[2] Second, perhaps the most important element in crime control in any community is the support of area residents. That support can take a number of forms, and the foremost is a general respect and support for the local government. If the residents hold government or any part of it in contempt, they are not likely to support or cooperate with crime control efforts by criminal justice officials.

IMPORTANCE OF CITIZEN SUPPORT

One important, specific kind of support is allocation of sufficient financial resources to allow justice systems to function effectively. The extent to which local citizens are willing to tax themselves, to support bond issues, and to vote for elective officials who will propose anti-crime actions determines in part the adequacy of justice agency staffs in terms of not only their numbers and initial level of competence but also specialized education and training later.

Another way in which citizen support for the local criminal justice system shows is in the reporting of crimes. Recent national surveys indicate that less than half of committed crimes are ever reported to the authorities, which means that our justice systems actually deal with only part of their potential business.[3] There are a number of reasons for failure to report offenses:

> In the national survey of households those victims saying that they had not notified the police of their victimization were asked why. The reason most frequently given for all offenses was that the police could not do anything. . . . The next most frequent reason was that the offense was a private matter or that the victim did not want to harm the offender. . . . Fear of reprisal, though least often cited, was strongest in the case of assaults and family crimes.

Other reasons listed were "did not want to take the time" and "too confused or did not know how to report."[4]

Still another way in which community residents can demonstrate cooperation or lack of it is by serving as witnesses when they have knowledge about a particular case. Here again, there is a wide range of attitudes among potential witnesses. There have been recent concentrated efforts by the National Institute of Law Enforcement and Criminal Justice (NILECJ) of the Law Enforcement Assistance Administration (LEAA) to make those in the justice system aware of practices that discourage people from reporting offenses and coming forward as witnesses.[5] It is hoped that these efforts will result in greater citizen support for justice agencies.

It is also possible for citizens to participate directly in the justice system as volunteers and auxiliaries. They can work as members of citizens' groups in continuing evaluation of the performance of justice agencies. And they can seek to create mechanisms for direct citizen participation in setting policy for agency operations.[6] An interested and actively participating citizenry is the basic element in successful crime control. Without that base, the most diligent and efficient agencies using the latest technology and other results of scientific research will find their efforts thwarted.

IMPACT OF VALUE AND IDEOLOGICAL SYSTEMS

Another very important element in crime control is congruence between the prohibitions of criminal law and the value systems and ideologies of area residents. A striking example of that premise was the prohibition experiment. In recent years, the lesson has been relearned with laws against the possession and use of marihuana.[7] Another issue on which the scales seem to have tipped recently is the criminal law prohibiting abortion.[8] In both of these areas, controversy still rages. Controversy has also developed over the role of our economic system in the genesis of crime. A group sometimes known as critical/Marxist criminologists argues that insight into crime can be obtained only through "understanding social life in relation to the underlying mode of production and the class struggle. This criminology mainly focuses on the relation of the economic system to the production of crime. A materialist analysis of criminality and crime control is required."[9] At the moment, the question of whether this critical/Marxist approach will gain headway is still very much uncertain.[10]

TRANSLATING RESEARCH INTO PRACTICE

Yet another caveat to be kept in mind is that research results are not automatically transferred or transferable into practice. Ideally, research findings should be tested promptly in real-life situations. But in carrying out this process of developing theory through testing in governmental and other social settings, research personnel must recognize and work within the constraints of political systems. Social and behavioral scientists must strive to be amoral and value free while generating new knowledge but must then, when testing their models as scientists in a democratic society, consider the morals and values of the system in which that testing is done. The difficulty of playing this role has been well delineated by Kalman H. Silvert. He stated that under the very best of conditions, the social scientist can do the following for governments with his special skills:

- generate and make available new data.
- order these data.
- indicate relevant theoretical patterns for the interpretation of the data.
- indicate the probabilities of effectiveness of various selected courses of action.
- indicate which choices are foreclosed.
- indicate which new choices will be made available.

Silvert noted that very few if any scholarly documents submitted to any government have satisfied these difficult requirements and that temptation to take the easy path from description to prescription is great.[11] This means that there must always be some distance between research personnel and practitioners, and the result is the almost inevitable dynamic tension between the two groups. The mutual obligation of each is to insure that this tension is constructive rather than destructive.

BACKGROUND AND CURRENT STATUS OF RESEARCH

ROLE OF NILECJ AND OF THE LEAA

Modern official concern with the potential for using scientific research and technology in crime control began with the appointment of a Science and Technology Task Force in 1966 by the President's Commission on Law Enforcement and Administration of Justice. The task force produced a report that led to inclusion in the Omnibus Crime Control and Safe Streets Act of 1968 (Public Law 90-351) of provisions creating the National Institute of Law Enforcement and Criminal Justice within LEAA. NILECJ has now completed its first decade of funding scientific research and technology projects designed to help control crime.

Through the fiscal 1969-79 years, LEAA has been authorized by Congress to spend $10,620,111,000. Actual appropriations for these 11 years amounted to

$7,213,145,000, about 68 percent of the authorization. Of the LEAA appropriation, $265,625,000 (3.7 percent) has been allocated to NILECJ; table 1 shows the year-by-year record of Institute funding. Note, however, that not all of the NILECJ budget is used to fund research.[12] Although estimates vary, an educated guess is that about $8 million is spent annually on projects that qualify as research.

Table 1. NILECJ Funding, in thousands of dollars

Year	Appropriation	Obligated	Expended
1968-69	$ 3,000	$ 2,824	$ 3
1969-70	7,500	7,348	2,626
1970-71	7,500	6,024	6,081
1971-72	21,000	19,825	9,558
1972-73	31,598	30,365	19,900
1973-74	40,098	29,961	34,500
1974-75	42,500	40,925	39,201
1975-76	32,400	35,743	41,672
Transition Quarter	7,000	7,001	8,906
1976-77	27,029	19,644	27,410
1977-78	21,000	39,164	22,740
Subtotals	240,625	238,824	212,597
1978-79	25,000	—	—
Totals	265,625	238,824	212,597

Source: *NILECJ/LEAA: LEAA Appropriations History Record*, Office of the LEAA Budget Director

CRITIQUE OF NATIONAL INSTITUTE OF LAW ENFORCEMENT AND CRIMINAL JUSTICE

Perhaps the most comprehensive critique of the Institute's performance has been that of the National Academy of Sciences (NAS), under an award made by NILECJ itself. In a 1977 report, *Understanding Crime: An Evaluation of the National Institute of Law Enforcement and Criminal Justice*, the Academy's Committee on Research on Law Enforcement and Criminal Justice summarized the findings of its 18-month study as follows:

> We conclude that the Institute has not been the catalyst or sponsor of a first-rate and significant research program commensurate with either its task or resources. It has clearly had some successes with individual projects and has begun to develop some basic and vital data and a research community . . . , but structural and political constraints have all too often deflected the Institute from its true mission. . . . Furthermore, we conclude that . . . the Institute in its present form is not likely to become a significant and quality-oriented research agency. We also conclude, however, that there is a need for a program of research on crime problems that is national in scope. . . . Consequently, we recommend both structural and conceptual reordering of the Institute itself and of its research agenda.[13]

The Academy report made 19 specific recommendations of which some of the most significant are highlighted here:

- The Institute should develop more programs that are cumulative in nature.
- The Institute should use a long range set of priorities.
- The Institute should use devices for making funding choices that would force it to take deliberate and systematic stock of what related research has already been undertaken, to tighten research designs, and to determine appropriate grantees and contractors.
- Upon completion of their projects, all NILECJ grantees and contractors should make their data available for secondary analysis, replication, and verification.
- The Institute should use announcements of areas of interest as the primary means of generating concept papers and proposals.
- The presumption should be in favor of granting rather than contracting as the Institute method for obtaining research.
- The Institute should use a variety of mechanisms to establish more positive relationships with a broadly defined research community and to enrich the dialogue between staff and quality researchers.
- The Institute's budget should not be increased in the near future. The Institute should change its emphasis to smaller proposals.

* * * * * *

- The Institute should employ a less obtrusive monitoring system that would allow more flexibility to grantees.
- Substantive program areas . . . should be the basis for creating the framework for program administration and budget allocation.

- Funding levels should not be rigidly fixed within substantive areas.

- Strict funding cycles—two or three a year—should be established and adhered to.

- The structure of NILECJ's research program should have appropriate evaluation, dissemination, and technology development functions integrated into the major research effort.

- . . . LEAA's domination over the Institute must be eliminated.

- . . . overall program priorities should be set by a statutorily authorized criminal justice research advisory board.

* * * * * *

- . . . the National Criminal Justice Statistical Service, the National Institute of Juvenile Justice and Delinquency Prevention, and Project SEARCH should all be in the NILECJ structure. We endorse the idea of a bureau of criminal justice statistics; the ideal arrangement would be to locate this bureau within an independent NILECJ.[14]

REORGANIZATION OF LAW ENFORCEMENT ASSISTANCE ADMINISTRATION

In response to the NAS report, the NILECJ published a statement that "major efforts have been made to implement all fifteen recommendations within the authority of LEAA and the Institute to implement."[15] Those requiring reorganization are now before Congress. The Senate Judiciary Committee recommended creation of an Office of Justice Assistance, Research, and Statistics (OJARS) in the Department of Justice. It would supervise and coordinate the activities of LEAA and two new agencies, a National Institute of Justice and a Bureau of Justice Statistics. In the House, a Kennedy/Administration bill has also been passed, but with somewhat different provisions. The two are in a conference committee at this writing.[16]

NILECJ's administrators (7 in 11 years) have developed, through trial and error and as a result of such constructive criticism as that in the NAS report, a method of funding that includes a long range (five-year) research agenda. Considered in setting that agenda were the Institute's congressional mandate; priorities set by the Attorney General and the LEAA Administrator; judgments of the Institute's professional staff; recommendations of the Institute's advisory committee; and the reactions of a large group of criminal justice researchers, practitioners, and Federal, state, and local officials.[17]

Items on the agenda for fiscal 1979 are correlates and determinates of criminal behavior; violent crime and the violent offender; community crime prevention; career criminals and habitual offenders; use and deployment of police resources; pretrial process (consistency, fairness, and delay reduction); sentencing, rehabilitation, and deterrence; and performance standards and measures for criminal justice. This set of priorities corresponds largely to that recommended in the NAS critique.

In addition to this basic and applied solicited research agenda, NILECJ operates an unsolicited research program "to widen the range of creative responses to criminal justice research issues." Two funding cycles are underway this fiscal year (1978-79), with a total anticipated commitment of $750,000.

Legislatively assigned objectives have also resulted in a number of Institute-administered programs in addition to research and development. These include evaluation of criminal justice programs; field tests of promising research findings and advanced criminal justice practices; training for criminal justice practitioners and assistance to the research community through fellowships and special workshops; service as an international clearinghouse for criminal justice information; and support for a science and technology program that tests and develops standards for equipment used by criminal justice agencies.

ROLE OF THE CENTER FOR RESEARCH ON CRIME AND DELINQUENCY

Second most active in funding crime-related research among Federal agencies has been the Center for Research on Crime and Delinquency (CRCD) of the National Institute of Mental Health (NIMH). This center was established because of the frequency with which mental health problems are involved in crime and delinquency. It sees itself as the focal point in NIMH for research, training, and related activities in the areas of crime and delinquency, individual violent behavior, and law and mental health interactions. Most of its grants go to scholars working in traditional university and research institute settings, although the center does fund demonstration projects that involve the careful development, articulation, testing, and evaluation of program models which can then be of benefit to other agencies and jurisdictions with similar needs. Its annual budget of about $5 million is small compared to that of NILECJ.[18]

ROLE OF OTHER FEDERAL AGENCIES

Among other Federal agencies, the National Science

Foundation, the Department of Labor, and the Department of Health, Education, and Welfare also fund an occasional significant crime-related research project because of the direct relevance of crime to their primary concerns. As a quasi-governmental agency, the National Academy of Sciences has also become active in the criminal justice area. In addition to its evaluation of the performance of NILECJ, the Academy has had substantive projects dealing with deterrence, rehabilitative techniques, and legislative impact on the courts.[19]

In 1975, the LEAA established the National Advisory Committee on Criminal Justice Standards and Goals (NAC) to continue the work of the earlier National Advisory Commission on the same subject. One of the subunits established by the new Committee was the Task Force on Criminal Justice Research and Development. The task force set three major purposes for its final report:

> First, the report addresses issues of the support and conduct of R&D in criminal justice, rather than the topics that might be researched.
>
> Second, this report is intended to serve the needs of criminal justice R&D policymakers over the middle range (e.g., 5 to 10 years).
>
> Third, the report attempts to draw together for the criminal justice community many issues on the conduct and management of R&D that have frequently been raised, but usually outside of the criminal justice field.[20]

MANAGEMENT OF FEDERAL RESEARCH EFFORT

The report contains 50 recommendations dealing with the support of research and development, conducting criminal justice research and development, research and development utilization, technology research and development, research on criminal justice organizations, and research on new criminal justice problems. These six topics are divided into two sections. The first three deal with the planning, management, and evaluation of funded criminal justice research and the last three demonstrate how the principles discussed can be applied in three subject matter areas. Note that much of the NAS critique of NILECJ also dealt with the management of research. Concern about the effectiveness of that effort, particularly at the Federal level, is not unique to the criminal justice area; the conclusion from the literature seems to be that there is no comprehensive U.S. science policy. In fact, it is suggested that science and management are essentially opposing ideals, although "they can nonetheless nourish one another."[21]

SETTING CRIMINAL JUSTICE RESEARCH PRIORITIES

This brief sketch of the background and current status of the role of scientific research and technology in crime control leads the author to conclude that the research agenda for criminal justice is set largely by LEAA/NILECJ because that agency has the largest share of the funds currently being expended. As indicated before, NILECJ has evolved a formal procedure for setting annual priorities. There is little doubt that the prime mover in establishing those priorities is the will of Congress, as perceived by LEAA administrators. Those administrators divine congressional will from the statutory provisions creating LEAA, as modified since original passage; from the questions and statements of Congressmen during hearings on LEAA's budget and periodic reauthorization; from the content of bills introduced in both houses to modify LEAA and its programs; and from correspondence and other personal communications with individual Congressmen. But congressional will sets priorities only in the broadest of terms, leaving the details to other mechanisms.

Within those general parameters, the LEAA professional staff assesses the input of the research community, practitioners, and public officials at all levels—both directly through personal contact and through the representatives of those groups on NILECJ's Advisory Committee. Other Federal agencies and private philanthropy will fund only projects not on the NILECJ priority list that advance the special interest of the agency or foundation. Although documentation cannot be obtained within the time and financial resources available for this essay, it is the considered judgment of the author that the research and technology budget allocations for crime control projects of both private foundations and other Government agencies vary inversely with those of NILECJ. As the expenditures of NILECJ have increased, their expenditures for criminal justice have decreased. The author believes that there will be a corresponding increase as NILECJ outlays decrease.

INFLUENCE OF PUBLIC INTEREST ON RESEARCH BUDGET

It also seems obvious that public interest and concern about crime as a social problem dictate the level of congressional appropriations for criminal justice research. When crime was first mentioned in the polls of citizen concern, LEAA's budget increased. Now that inflation, energy shortages, and environmental pollution have captured citizen interest, budgets for criminal justice research have fallen and will probably continue to fall.

KNOWLEDGE ABOUT THE CRIME CONTROL RESEARCH COMMUNITY

To the best of my knowledge, there has been no published study of the nature of the research community that has responded to the program of NILECJ. There is some discussion in the NAS critique, reflected in recommendation seven of that report (see above). Again my impression is that there has been heavy reliance, particularly for programs involving very large sums of money, on nonuniversity/commercial and nonprofit research organizations. However, the smaller, more theoretically oriented projects do go to colleges and universities. Among the latter, there also seems to have been a shift as LEAA/NILECJ matured toward support of projects in universities with solidly established reputations for sound research, generally in the social and behavioral sciences. An aversion has definitely developed to the risk inherent in making awards to untried younger scholars in institutions without an established reputation for excellence in research.

BENEFIT TO CRIME CONTROL FROM RESEARCH AND TECHNOLOGY

Crime control benefits from current scientific research and technology are uneven and difficult to assess. Many of the resulting reports go unheeded. However, there have also been dramatic successes, one of which is the Prosecutor's Management Information System (PROMIS), originally developed by the United States Attorney for the District of Columbia. With the aid of the Institute for Law and Social Research, the computer version of PROMIS is now in operation in 21 jurisdictions and the manual version in another 18. Transfer to the system is underway in 37 additional jurisdictions. Four states (Colorado, Michigan, New Jersey, and New York) are now in the process of implementing PROMIS on a statewide basis, using grant money obtained under LEAA's Incentive Fund Program. The Colorado system will be regional. Planning for implementation of PROMIS is underway in 36 additional jurisdictions, including 6 more entire states.[22]

Five principal objectives are listed for PROMIS:

- to give visibility to the differences in the importance of cases and to assure even-handed and consistent justice.
- to give special intensive pretrial preparation to the more serious cases, to which PROMIS assigns high numerical ratings.
- to know immediately when accused persons have multiple criminal cases pending at the same time, and to have instantaneous access to case status information on any pending case.
- to develop statistical reports on prosecution and court activity that illustrate bottlenecks, training needs, crime trends, legal problems, and social problems.
- to aggregate empirical data about the court process, crime trends, and socio-legal problems in order to discover what is successful and to devise ways to improve and sustain success.

Results claimed for PROMIS include conviction rates for cases involving serious matters, reduction in time lag before indictment, establishment of legal paraprofessional positions to assist attorneys in certain tasks revealed through PROMIS to be performed unsatisfactorily, provision of comprehensive training programs for attorneys to combat deficiencies in prosecutive skills revealed by PROMIS, development of comprehensive charging manuals to document policies and procedures regarding the prosecutor's filing of criminal charges, and the initiation of criminology research projects to analyze the PROMIS data base. One of the latest research reports is *A Cross-City Comparison of Felony Case Processing*, in which 13 jurisdictions that have operational PROMIS systems are compared.

Although development and spread of the PROMIS system is not a typical case, it does illustrate the potential of scientific research and technology in crime control.

POSSIBLE CONTRIBUTIONS OF RESEARCH TO CRIME CONTROL

Any scheme for categorizing the contributions of science and technology to crime control depends on the mission of the person making the categorization. For purposes of this essay, the contributions are classified according to their usefulness in defining the problem of crime and determining its causes; planning programs believed capable of preventing significant amounts of crime; establishing policies on the formal governmental reaction of society to crime; determining how individuals will be proved subject to governmental crime control processes; and planning what to do with persons who have become subject to those processes. To facilitate further discussion, the criminal justice system is defined as those units of government that create and administer the criminal law. This includes legislative bodies of many kinds; planning agencies; police and other investigative units; the offices of prosecutors; courts and other adjudicative bodies; and institutions and other agencies given custody of selected persons processed by the system.

IN DEFINITION OF CRIME

Defining crime is a process of translating value judgments and ideological preferences into the statutes and judicial decisions that make up our substantive criminal law. Particularly useful in that process are the scientific research findings of lawyers, philosophers, and all kinds of social scientists, anthropologists, economists, political scientists, psychologists, and sociologists. Research from these disciplines can identify and clarify competing values and ideologies, provide insight into public attitudes on those concepts, and illuminate such issues as whether the substantive criminal law is an effective way to control a particular behavior problem.

Promulgation of the American Law Institute's Model Penal Code[23] is an example of the collaborative process of lawyers and persons from these disciplines in producing a code that has had a great deal of impact on the substantive criminal law of many states. But there is still far from complete agreement that our definition of crime in the United States is fair and just. Typical of a view that is being voiced with ever increasing clarity—even though it has not yet had the impact on actual criminal codes that more traditional views have achieved—is that of philosopher Jeffrey H. Reiman, who states in a recent book:

> The dangerous acts and crimes unique to the wealthy are either ignored or treated lightly, while for the so-called common crimes, the poor are far more likely than the well-off to be arrested, if arrested charged, if charged convicted, and if convicted sent to prison. Hence the failure of the criminal justice (system) has a pattern: *the rich get richer and the poor get prison*. (Emphasis in the original.)[24]

Crimes that Reiman would formalize and emphasize include creating and maintaining industrial hazards, causing death and serious injury through improper emergency medical care, contaminating people through chemicals (air pollution, cigarette smoking, and food additives), and institutionalizing poverty. This debate on the values and ideologies to be protected and furthered by the substantive criminal law is an important one. Scientific research can contribute light where heat has been more characteristic in the past.

IN DETERMINATION OF THE CAUSES OF CRIME

Determining the causes of crime is a process that has been going on for centuries. It is now generally accepted that there is no single simple cause of crime—that crime causation is pluralistic. The search goes on for criminogenic factors both within individuals and in the social environment. Both medical science and psychology are useful in the search for causes within human beings, although the search for organic biological causes has been both relatively unsuccessful to date (except in a negative sense) and discredited by generalizations that tend to brand categories of persons as "criminal types," a practice that favors racism and genocide.

Research on biological factors as possible causes of crime does continue, however. Recent studies explore the possibilities that being less physically attractive can influence young persons toward delinquency; that menstrual and premenstrual tension may lead to commission of crime; that the children of female offenders may be characterized by a high rate of antisocial personalities even when removed from the cultural influence of the mother through adoption; that the presence or absence of XYY chromosomes may be a genetic influence on disposition toward human aggression; and that twins might show the same behavioral traits because of a common hereditary endowment.[25] Considerable insight has been gained into the pathologies that invade mental health through the contributions of psychologists and psychiatrists in such areas as learning theory, understanding the elements of human motivation, and understanding the impact of stress on individuals. Specific areas investigated recently have included intelligence and juvenile delinquency; learning disabilities and juvenile delinquency; lunar cycles and homicides; extroversion, neuroticism, and delinquency; personality and criminality; and pornography and its effects.[26]

Anthropologists and sociologists contribute to a greater understanding of humans as social beings, of the formation of social institutions, and of the formation of both conforming and deviant (including criminal) relationships by individuals and groups with social institutions. Titles in a recent collection of such studies include investigation of dishonesty in the Soapbox Derby; the broken home and delinquency, American women and crime; race, social status, and arrest; status origins, track position, and delinquency; delinquency and the school dropout; the influence of church attendance on delinquency; television violence and children; gang violence; disparate treatment of delinquent gangs of upper-middle-class and lower-class youths; girls in gang violence; and population density and crime.[27]

Through their research, economists give us greater insight into economic crimes, particularly organized and white collar crime.[28] Political scientists help us understand the corruption of government that is a frequent concomitant of economic crime.[29] Thus it is the behavioral and social sciences that are most useful in determining causes of crime.

IN PLANNING PROGRAMS TO PREVENT CRIME

Planning programs believed capable of preventing significant amounts of crime again draws upon the research primarily of behavioral and social scientists, with lawyers and social workers also becoming involved with program implementation. Most social planning that leads to crime prevention is a positive development that leads to healthier social institutions and eventually to healthier communities for all. Some of these programs are designed specifically to serve groups of individuals convicted of violating the criminal law, processed by the criminal justice system, and then released back into the community. Such programs are designed to continue any rehabilitative efforts begun in the system by helping the releasee become reintegrated into the community.

Other planning programs will use technological advances for "target hardening"—making it more difficult to commit a crime. An example of that kind of crime prevention is developing better door and window locks and intrusion alarms to frustrate burglars. To further the search for new ways to prevent crime and to publicize better what is already known, LEAA has funded a National Crime Prevention Institute at the University of Louisville (Kentucky).

IN ESTABLISHING POLICIES

Establishing policies for the formal governmental reaction of society to crime is a process that calls once more on lawyers, philosophers, and all types of social scientists. Involved here is the familiar decision of how much restraint should be imposed on individual freedom in order to increase social order. There is general agreement today that there should be governmental rather than individual reaction to antisocial conduct defined as criminal. The question of what that reaction should be and how it should vary with the kind and seriousness of the offense is essentially political, involving, as does the definition of crime itself, the application of value systems and ideologies. For this reason, the research results discussed before in terms of defining crime are useful here, too.

IN DETERMINING GOVERNMENTAL CRIME CONTROL PROCESSES

Determining how individuals will be proved subject to governmental crime control processes gets us into the design of police, prosecution, and adjudication agencies and the definition of their procedures. The detection of covert crime raises issues of invasion of privacy. For example, requirements that all individuals should carry identity cards perhaps bearing one or more of their fingerprints, that their daily movements be recorded and reported, that homes and workplaces should be open to inspection at all times, that the mobility of the population be severely limited, and that guns and other implements frequently used in committing crimes be registered could no doubt result in more effective crime control. But such measures also restrict privacy and other freedoms and would not be tolerated in the United States. Determining what procedures will be allowed is again a political process that involves weighing values and ideologies. Philosophers, political scientists, and lawyers produce research useful in that process. They in turn will frequently apply the work of anthropologists, economists, psychologists and sociologists. The American Law Institute Model Prearraignment Code[30] is the result of such collaboration.

Once permissible procedures have been decided upon, lawyers and those political scientists whose specialty is public administration will be useful in setting up police departments, prosecutors' offices, and court systems. In so doing, they will call on the research of sociologists and others into organizational theory and behavior. They will also find most useful the computerized information systems, one of the modern technological advances most useful in crime control.

As the police and other investigators begin gathering evidence tending to prove or disprove the participation of a given person in a violation of the criminal law, they will use the research results of forensic scientists. Here the natural and life sciences can assist the investigator with chemical analyses, microscopic comparisons, and other procedures useful in determining the facts to be presented in court. It is here that the hard sciences and the modern technologies they have produced are most helpful in crime control. Their contribution is appreciable, but it comes largely after the tasks of defining the substantive criminal law and designing the criminal justice systems necessary for its administration.

IN PLANNING RESPONSE TO CONVICTIONS

Planning what to do with persons convicted of crime again involves the very difficult task of applying social and behavioral science knowledge to the process of making value and ideological judgments. The decisions to be made involve not only the extent to which it is possible to modify human attitudes, motivations, and value systems but the extent to which it is ethical, moral, and in keeping with democratic ideals of government to apply what we know in modifying, punishing, and incapacitating our fellow human beings. Here once more the best work of philosophy, jurisprudence, and the behavioral and social sciences is needed to make wise decisions.

One very difficult area in which the guidance of these disciplines is sorely needed is in the exercise of discretion in criminal justice matters. Indeed, a great deal of discretion is built into the administration of our substantive criminal law. Some of that discretion is openly authorized and used; some of it is unauthorized and hidden. As the work of Leslie Wilkins, Don Gottfredson, and others has demonstrated,[31] criteria used in the exercise of discretion, whether open or covert, can be identified, recorded, and systematized. When surfaced in that way, the criteria can be evaluated both by those who are using them and by the public at large.

When discretion is mandated by statute, there is no problem with issuing standards for its use. When the statutes are silent, discretion is usually required by the exigencies of a situation. Thus, guidelines officially adopted as a basis for making necessary choices are desirable. An open statement by an official stating the need for discretion and the guidelines for its exercise is a kind of accounting for stewardship. It makes possible a public evaluation both of the case for discretion and the criteria guiding its use. The pertinent policy-setting entity to which the official is responsible can then make its own decision, both as to the need for discretion and the fitness of the criteria being used to guide decisionmaking.

Where independence on the part of the decisionmaking official may seem desirable (as may well be the case, for example, with judges sentencing persons convicted of crime within limits permitted by statute), voluntary adoption of carefully drafted and recommended standards should not be a problem. Refusal by such an official to adopt reasonable standards should be a matter of public record that can be considered by those responsible for ongoing evaluation of the official's performance. Independence does not mean freedom from accountability. There is no official, appointed or elected, who is not ultimately responsible to the people and who cannot be removed by some method, even though it might require the heroic and seldom used process of impeachment.

Without a doubt, the most difficult case for adopting standards to guide the exercise of discretion occurs when discretion appears to be forbidden by law. An example is the frequent provision in statutes requiring police officers to enforce all of the law all of the time. However, there is probably no large city police department in the United States today that has the resources necessary to meet such a mandate. Thus, setting priorities—making decisions as to where enforcement resources will be concentrated—is a practical necessity. And, it is particularly important that both the process for setting priorities and the end result expected be formally publicized and made known to the policy-setting entity to which an official is responsible. Having been so identified, both the process and its result can then be evaluated.[32]

In sum, this very brief sketch of possible scientific and technological contributions to crime control indicates that the natural and life sciences and the wonders of modern technology are useful primarily in demonstrating that crimes have been committed and in arriving at the identity of the perpetrators. Law and the social sciences, on the other hand, are the sources of knowledge that is useful in defining crime, in gaining insight into crime causation, in planning crime prevention programs, and in deciding what to do with persons adjudged guilty of committing criminal acts.

THOUGHTS ON THE FUTURE OF RESEARCH FOR CRIME CONTROL

In considering the future role of scientific research and technology in crime control, one can only hope that the current move toward balanced budgets in government at all levels in the United States will not lessen that role. A strong argument can be made that this generation has an obligation to those yet to come to invest in the development of new insights into crime and the more general problems of social justice. As John Rawls has said:

> Each generation must not only preserve the gains of culture and civilization, and maintain intact those just institutions that have been established, but it must also put aside in each period of time a suitable amount of real capital accumulation.[33]

Investment in scientific research is a very real intellectual capital accumulation, a responsibility that rests not only on government but on the private sector as well. The very first recommendation of the Task Force on Criminal Justice Research and Development begins with the statement that "Criminal justice R&D should be supported by a variety of sources—Federal, state, local, and private."[34] Not only does such a broad base of support guarantee a "suitable amount" of research, but it allows greater diversity in the approach to problems. One hopes that it also will lead to the expenditure of funds to establish creative environments that can lead to surprise results in criminal justice research. Leslie Wilkins argues often that the amount of surprise generated by a research project—results neither planned nor anticipated—may well be a measure of the worth of that project.

SUPPORT BASE FOR CRIME-RELATED RESEARCH

Assuming that criminal justice research not only survives but broadens its base of support, the testing, validation, and technical assistance programs of LEAA now underway should also be continued. In fact, a new subspecialty of research on knowledge utilization has come into being and is now beginning to be applied to criminal justice.[35] This process recognizes that there are a number of knowledge transfer issues related to criminal justice that are significantly different from those found in such fields as agriculture, medicine, and education. A recent research proposal lists six issues that "hold promise for profitable analysis: (1) the impact of values and moral judgments on the determination of criminal justice policies, (2) the frequency with which major policy is made within the confines of the political arena, (3) the diversity of points of view, (4) the characteristics of the actors who play major policy roles, (5) the limited nature of the Federal role, and (6) the lack of continuity and consistency in Federal policy."[36] The research being funded in this subspecialty should help produce profits from the capital investments now being made in substantive research on crime control.

LONG-RANGE PRIORITIES

Also welcome is the acceptance by NILECJ of the principal recommendations of the NAS critique. A long range set of priorities should yield more cumulative research. Studies on research design and methodology should be encouraged, and research project data should be made available for secondary analysis, replication, and verification. In addition, more attention should be paid to development of the criminal justice research community. A rudimentary infrastructure is now in place to support such a community: College and university teaching and research programs that are transdisciplinary in nature and have identity as separately structured and administered units now exist and are flourishing. Textbook, monograph, and treatise series in criminal justice are now favored lines with many publishers; journals devoted to criminal justice research and policy matters abound. Professional and scholarly associations have sprung up; in addition, the specialized criminal justice developments on our campuses, together with the availability of increased financial support for crime-related research, is leading to the more frequent choice of crime problems as vehicles for research by scholars from various disciplines. The last development is shown by the increasing number of reports in the disciplinary research journals dealing with crime-related matters.

UNIVERSITY JUSTICE STUDIES PROGRAMS

One of the characteristics of the new justice studies programs is that most of them are in state-supported colleges and universities. That fact makes it particularly easy for such programs to establish long-term relationships similar to that recommended by the NAS report between a new independent NILECJ and LEAA, with state, regional, and local planning units being responsible for crime control in their areas. Investment of funds from all sources—Federal, state, local, and private—in the nurturing of such relationships is well warranted. Longstanding Institutes of (Local) Government in many state universities exemplify the many values inherent in such collaboration. Some of the more important benefits are that research capacity is built into a local institution, that research data are archived locally, and that actual use of research results is made easier.

It has been true in the past that only a small handful of justice studies programs have had a sophisticated research capacity in their faculties, but that situation is rapidly changing. There are now at least a half dozen programs that are as good as any in other fields, and the addition of young research-oriented members to their faculties (which is the order of the day) is generating research nuclei in many other areas. The development of at least another 20 research-oriented justice studies faculties, building on current potential wherever it may be found, deserves a subsidy. The result will be an invaluable national resource capable of bringing about a much more significant role than before for scientific research and technology in future crime control.

UNDERPINNING VALUE JUDGMENTS

Because of the recognized importance of the generation of new knowledge about the control of crime and because many of the issues are moral and political, it is imperative that they be addressed by scientific research. As one observer points out:

> Where there are important value choices, there would seem to be good grounds for invoking public input of information. Furthermore, if we are asking the public to make value choices . . . it follows that we have a duty to inform the public.
>
> The position of the research scientist may be argued in the same general terms. Research is an institution in society. Some might think a highly prestigious institution. The scientific method is a part of our cultural heritage. If the general public have (sic) a stake in accountability of officials in the public sector, the research community has a special stake.[37]

This plea for consideration of the research function was a reaction to a rigid and unthinking application to social science research of principles derived from medical research and designed to protect human subjects where potential serious physical harm and even death are involved. This application was made in a way that threatens to stifle social science research. But the application also can be read as an eloquent statement of the particular importance of the role of scientific research and technology in crime control. That role should not only be continued but expanded.

REFERENCES

1. These definitions are from the National Advisory Committee on

Criminal Justice Standards and Goals, *Criminal Justice Research and Development*, Washington, D.C.: U.S. Government Printing Office, 1976, pp. 1 and 87.

2. See generally Sutherland, Edwin H. and Cressey, Donald R., "A Sociological Theory of Criminal Behavior," chapter 4 of *Criminology*, Philadelphia: J.B. Lippincott Company, 10th Edition, 1978.

3. For a critical comment on the use of victim surveys, see Levine, James P., "A Critique of Criminal Victimization Surveys," chapter 10 in Savitz, Leonard D. and Johnston, Norman, *Crime in Society*, New York: Wiley, 1978.

4. The President's Commission on Law Enforcement and Administration of Justice, Task Force on Assessment, *Crime and Its Impact: An Assessment*, Washington: U.S. Government Printing Office, 1967, 0-239-113, p. 18.

5. See Cain, Anthony A. and Kravitz, Marjorie, *Victim/Witness Assistance: A Selected Bibliography*, Washington: U.S. Government Printing Office, 1978, 260-992/4506.

6. For an extended discussion, see Myren, Richard A., "Decentralization and Citizen Participation in Criminal Justice Systems," *Public Administration Review* 32:718-38, special issue, October 1972.

7. Weissman, James C., *Drug Abuse: The Law and Treatment Alternatives*, Cincinnati: Anderson Publishing Co., 1978.

8. See discussion in the chapter on "The Enforcement of Morals," In Sykes, Gresham M., *Criminology*, New York: Harcourt, Brace, Jovanovich, Inc., 1978, pp. 165-166.

9. Quinney, Richard, *Criminology*, 2nd Edition, Boston: Little, Brown and Company, 1979, p. 26.

10. Sykes, pp. 231-233.

11. Silvert, Kalman H., "American Academic Ethics and Social Research Abroad," *American Universities Field Staff Reports Service: West Coast South America Series*, Vol. 12, No. 3, July 1965. This issue was raised before in Myren, Richard A., "Nature of the Criminal Justice Doctorate," *Criminal Justice Graduate Curriculum Development*, National Criminal Justice Education Consortium Reports, Vol. 4, 1976.

12. National Academy of Sciences, *Understanding Crime: An Evaluation of the National Institute of Law Enforcement and Criminal Justice*, Washington: National Academy of Sciences, 1977, pp. 52-65.

13. Ibid., p. 3.

14. Ibid., pp. 96-110.

15. National Institute of Law Enforcement and Criminal Justice, *The Implementation of the Recommendations Made by the Committee on Research on Law Enforcement and Criminal Justice of the National Institute of Law Enforcement and Criminal Justice*, Mimeograph, 10 pages, 1977.

16. National Center for State Courts, "Hearings Set to Begin on Bills Reauthorizing LEAA: Supporters Plan Major Fight to Bar Cut in Appropriation," *Washington Memorandum*, Vol. 5, No. 1, 1979. This report summarizes the interests involved.

17. National Institute of Law Enforcement and Criminal Justice, *Program Plan: Fiscal Year 1979*, Washington: U.S. Government Printing Office, 1978, p. vii. The information in this paper on the current Institute agenda is also from this plan.

18. For general information and a description of current awards, see current mimeographed statements available from NIMH describing its Center for Research on Crime and Delinquency and listing the active research grants of the center.

19. For a discussion of current criminal justice research funding, see President's Commission on Law Enforcement, pp. 11-22.

20. National Advisory Committee on Criminal Justice Standards and Goals, *Criminal Justice Research and Development*, Washington: U.S. Government Printing Office, 1976, 0-219-286, pp. 5.

21. Bozeman, Barry and Mitroff, Ian, symposium eds., introductory comment to "A Symposium: Managing National Science Policy," *Public Administration Review* 39:111, March/April 1979, referring to Mason, Richard O., "The Role of Management in Science." Other articles in the symposium are Bozeman, Barry, "Straight Arrow Science Policy and Its Dangers"; Levine, Arthur L., "The Role of the Technoscience Administrator in Managing National Science Policy"; Mitroff, Ian I., "On Managing Science Holistically, or Is the Management of Science Becoming More Important Than the Philosophy of Science?"; Gordon, Gerald, and Neumann, Yoram, "A Three Facet Model of Scientific Development"; and Lambright, W. Henry, and Teich, Albert H., "Policy Innovation in Federal R&D: The Case of Energy." See also "U.S. Science Policy Earns Mixed Review," *Public Administration Times*, 15 March 1979, p. 7, c. 1, and Greenberg, Daniel S., "Scientific Research: Wheat and Chaff," *The Washington Post*, 17 April 1979, p. A21, c. 6.

22. The section on PROMIS is taken from the *PROMIS NEWSLETTER*, Vol. 4, No. 1, May 1979, and other materials available from the Institute for Law and Social Research, 1125-15th Street, N.W., Washington, D.C. 20005. *A Cross-City Comparison of Felony Case Processing* by Kathleen B. Brosi of the INSLAW staff was published in 1979 for LEAA by the U.S. Government Printing Office (1979-281-380/1622).

23. American Law Institute, *Model Penal Code*, Philadelphia: The American Law Institute, 1962. An updated, annotated edition is in preparation at this writing. See also the National Commission on Reform of Federal Criminal Laws, *Study Draft of a New Federal Criminal Code*, Washington: U.S. Government Printing Office, 1970.

24. Reiman, Jeffrey H., *The Rich Get Richer and the Poor Get Prison: Ideology, Class, and Criminal Justice*, New York: Wiley, 1979, p. viii. See also Black, Donald, *The Behavior of Law*, New York: Academic Press, 1976.

25. These examples are selected because the reports are readily available in a section on biological factors as causes of crime in Savitz and Johnson, pp. 249-308. See also Moran, R., "Biomedical Research and the Politics of Crime Control: A Historical Perspective," *Contemporary Crises* 2:335-357, 1978.

26. Reports of these studies can also be found in a special section of Savitz and Johnson, pp. 309-392.

27. These studies are collected in Savitz and Johnson, pp. 393-536.

28. Savitz and Johnson, pp. 537-572.

29. See Gardiner, John A., "Appendix B—*Wincanton: The Politics of Corruption*," pp. 61-79, in The President's Commission on Law Enforcement and Administration of Justice, *Task Force Report: Organized Crime*, Washington: U.S. Government Printing Office, 1967, 0-256-177.

30. American Law Institute, *A Model Code of Pre-arraignment Procedure*, Philadelphia: American Law Institute, 1975.

31. See for example Gottfredson, Don M., Cosgrove, Colleen A., Wilkins, Leslie T., Wallerstein, Jane, and Rauh, Carol, *Classification for Parole Decision Policy*, Final report to NILECJ, Washington: U.S. Government Printing Office, 1978.

32. See also Davis, Kenneth Culp, *Discretionary Justice: A Preliminary Inquiry*, Chicago: University of Illinois Press, 1976.

33. Rawls, John, *A Theory of Justice*, Cambridge, Mass.: Harvard University Press, 1971, p. 285.

34. National Advisory Committee, p. 22.

35. See Havelock, Ronald G. (in collaboration with Alan Gaskin et al.), *Planning for Innovation*, Ann Arbor: Center for Research on Utilization of Scientific Knowledge, Institute for Social Research, 1969. See also Rogers, Everett M., and Shoemaker, F. Floyd, *Communication of Innovations*, New York: The Free Press, 1971; Yin, Robert K. et al., *Changing Urban Bureaucracies: How New Practices Become Routinized*, Santa Monica: RAND Corporation, 1978.

36. "Knowledge Utilization in Criminal Justice," proposal submitted to NILECJ by Dr. Ronald G. Havelock, Knowledge Transfer Institute, The American University, Washington, D.C.

37. Wilkins, Leslie T., "Human Subjects—Who's Subject?" Unpublished manuscript, January 1979, pp. 32-33.

III Science and Technology for the State and City

1 Science and Technology in State and Local Governments: Problems and Opportunities

*by Irwin Feller**

SUMMARY

This paper addresses several policy issues relating to the formulation of Federal policies to promote a more effective utilization of scientific and technical (S&T) knowledge by state and local governments. In particular, it is concerned with the issues of (1) how state and local governments perceive their needs for S&T knowledge; (2) the appropriateness of a Federal role in providing assistance to state and local governments in this area; (3) an assessment of Federal efforts to provide S&T assistance to state and local governments; and (4) needed additions to any such future Federal programs.[1]

The paper draws heavily upon an emerging body of research findings on the processes of adoption, diffusion and utilization of S&T knowledge by state and local governments.[2] These findings point to a more complex set of relationships that affect the utilization of S&T knowledge by these jurisdictions than is currently accounted for in many Federal programs. For purposes of this paper, the central importance of these findings is that they point to basic incompatibilities between the way in which the Federal Government is organized to provide S&T assistance and the elements which influence the effective utilization of this assistance by state and local governments. These incompatibilities relate to differences between the product orientation of Federal agencies and the process-orientation needs of state and local governments: to different perspectives on what constitutes success, both in terms of the measurement of output and the time period over which activities are evaluated; and to a lack of incentives for Federal agencies to support the type of activities that are increasingly viewed as the essential elements in developing effective capabilities to utilize S&T knowledge within state and local governments. The range of activities that Federal agencies conduct that relate to S&T assistance to state and local governments is broad, transitory, and devoid of systematic evaluation.[3] This paper thus provides an assessment—in part well documented and in part impressionistic—of patterns, tendencies, and trends rather than categorical judgments.

THE USE OF SCIENTIFIC AND TECHNOLOGICAL KNOWLEDGE BY STATE AND LOCAL GOVERNMENTS

State and local governments employ S&T knowledge in much the same way as the American populace employs the English language—on a daily basis, unquestioningly, and at less than technically attainable standards of performance. The utilization of S&T knowledge by state and local governments as a discrete public policy issue is of relatively recent origin, dating from the mid-1960s. The emergence of this issue involved the convergence of several streams. These included concerns

* Director, Institute for Policy Research and Evaluation and Professor of Economics, Pennsylvania State University.

over the increased technical complexity of public policy issues and the deteriorating financial conditions of state and local jurisdictions. The latter were, in part, attributed to low levels of productivity found to characterize public sector operations. Also contributing were the efforts of Federal agencies to promote the utilization and transfer of findings from research undertaken with their support. This Federal push arose from the belief that the technical and managerial know-how that had successfully placed a man on the moon could readily be applied to the solution of many domestic problems.

The euphoria of the mid-1960s, symbolized by the transition of the moon-ghetto metaphor from a hortatory slogan for action to the analyst's description of the innocence of an earlier period, has dissipated.[4] There have also been changes in the forms that specific policy issues have taken. For example, the urban fiscal crisis of the 1960s has given way to the Proposition 13 mentality of today and the then pressing concerns with air and water pollution are now being supplanted by requirements upon states to develop policies on energy and toxic materials. In the face of these changes, the root needs of state and local governments for the more effective utilization of S&T knowledge remain the same: (1) improving the quality and efficiency of the delivery of services by state and local governments; and (2) improving the quality of decisions (formulation, enactment, administration, and review) made by these levels of government. They remain the same for a number of reasons. The social forces that gave rise to these problems have not changed much. The assistance offered by the Federal Government to strengthen the S&T capacities of state and local governments has been both modest in scale and not always employed successfully. Finally, the needs remain the same because specific assistance projects have yet to evolve from the vulnerable and halting status of demonstration projects or quasi-experiments to permanent programs.

A recurrent problem in the formulation of Federal intergovernmental science and public sector technology transfer programs is the tendency to treat the above-cited concerns of state and local governments as a mirror image of Federal endeavors to achieve other valid and overlapping policy objectives. Prominent among these other subsidiary objectives are (1) augmentation of the rate of technological (and presumably productivity) change within state and local governments; (2) development of strengthened policy management capabilities—the identification of needs, analysis of options, selection of programs, and allocation of resources—on public issues containing S&T components; and (3) substantive application of S&T knowledge in specific policy fields, such as health care, energy or environmental protection. The complementary aspects between the two core objectives and these other objectives are readily discerned. The conflicts are less obvious; nevertheless they exist.

In the late 1960s, for example, low levels of economic activity in defense and aerospace industries gave rise to proposals to "couple" or otherwise improve the "interface" between these industries and state and local governments. Meshing of the technical capabilities of these industries with what were perceived to be the technically based needs of state and local governments had a surface logic (as well as obvious political appeal). Individual projects directed at promoting these linkages have received favorable evaluations (e.g., California Four Cities program and Baltimore Applications Project).[5] But these projects have also uncovered problems relating both to the characteristics of the technologies being developed and to the organizational and individual incentive systems that have served to limit the replicability of such demonstrations.

Conceptualization of the use of S&T knowledge by state and local governments has undergone several changes over the past decade. Chief among these changes are the recognition that the processes which shape such use are more complex than initially articulated in such documents as *Public Technology* or *The Struggle to Bring Technology to the Cities*, and that solutions are more difficult tnan originally envisioned. It is now recognized, at least by some, that there are several different elements contained within the broad headings of S&T knowledge for state and local government or of intergovernmental science relationships. There are also several different processes at work that affect the utilization of S&T knowledge (defined to include both the adoption and incorporation of new technologies and the search for and utilization of S&T information). Furthermore, these processes are only imperfectly understood.

State and local governments already employ a variety of approaches in their efforts to maintain more systematic utilization of S&T knowledge.[6] Governors, for example, have designated science advisors, established science advisory boards, and built a scientific and technical capacity into central management agencies, such as budgeting or planning. State legislatures have employed designated staff scientists; they have added individuals with training in the physical and biological sciences to standing committees; and they have established liaison arrangements with the faculties of colleges and universities. Cities have established municipal science advisory bodies; they have employed technology transfer agents, sometimes as part of a separate initiative, other times as part of larger projects designed to foster the spread of innovations among a cluster of cities. National organizations (e.g., National Governors' Association, National Conference of State Legislatures, International City Management Association) have similarly become involved in projects designed to foster the dissemination of S&T information among their constituent members. These and other national organizations have likewise become involved in efforts to generate "user-based" research and development (R&D) agendas. These agendas are designed to induce an R&D project selection pattern by Federal agencies that is consistent with the needs of state and local governments.

Several general observations concerning these approaches are more useful here than detailed descriptions. The utilization of S&T knowledge involves several different elements; I have classified the elements within the S&T label as follows: (1) R&D (which includes both state and local R&D activities and state and local involvement in R&D agenda setting at the Federal level); (2) Technology Utilization (which includes the processes and procedures employed by state and local governments to search for, assess, and "routinize" new practices, including those offered by both the private sector and the public sector); and (3) Scientific Information and Advice (which includes the processes and procedures employed by state and local governments to search for and assess policy-relevant S&T information).

The importance of these elements appears to vary across levels of government. Within the executive branch of state government, the governor's office is apparently concerned principally with access to S&T information and advice[7] and relatively unconcerned with the processes by which line agencies acquire new technologies. Line agencies evince a high interest in obtaining access both to information and to new technologies. The concerns of the state legislatures parallel those of governors.[8] State legislators want access to S&T information; they are relatively unconcerned with issues of technological change. Most critically, they want an independent access to S&T information. Legislatures have been found to employ line agencies as a major source of their information. However, with the political resurgence of this branch of state government has come an increased number of statements to legislators' needs for sources who are responsive and/or accountable to their special working (and political) environments. City governments have generally been more active in direct technology transfer programs than in strengthening their access to S&T information. In many ways, this relative emphasis reflects the division of responsibility among levels of government. In most states, city governments are the principal suppliers of most of the public sector services (e.g., fire protection, police protection, solid waste collection and disposal) about which efforts to develop new public sector technologies have centered.

The dominant view, expressed by users (practitioners) and researchers alike, is that "local" conditions determine which of the set of alternative approaches (models, mechanisms) is likely to be most effective in a given site. The delineation of discrete models is useful mainly at the planning stage to assess alternatives. In practice, elements of several models can readily be combined. Finally, the emphasis on alternative approaches derives largely from the project format through which Federal support has been provided for state and local activities. Federal support for these approaches has tended to be couched in terms of "tests" or "demonstrations" of alternative models (or, more formally, of the hypotheses, often implicit, contained within proposals) concerning the organizational determinants of the flow and utilization of S&T knowledge. These approaches are conceptually distinct from one another, and could indeed remain so on an operational level. However, there is also reason to believe that there is a convergence (or broadening) of both objectives and tactics of approaches over time so that initial distinctions lose significance.

RELATIONSHIPS BETWEEN FEDERAL AGENCIES AND STATE AND LOCAL GOVERNMENTS

Expectations for the consequences of Federal policy initiatives in the field of intergovernmental science derive from the same mood that currently affects all Federal programs. As has been noted by others, prescriptions for future Federal initiatives are affected by the dampening of expectations caused by the set of experiences and disappointments associated with the presumptive faith in large-scale Federal interventions of the 1960s.[9] More than offsetting this dampening of expectations, however, have been the positive changes that have occurred both among state and local jurisdictions and between the Federal Government and these jurisdictions: (1) There has been an increased awareness and articulation within segments of state and local jurisdictions of their needs for more systematic access to S&T knowledge. In part, this increased attention derives from self-initiated efforts; in part, it derives from participation in federally funded projects in the field of intergovernmental science. (2) Federal support of discrete projects designed to foster the utilization of S&T knowledge has produced a sufficient number of "successes" (however roughly this judgment is reached), to maintain Federal, state, and local commitments to the concept that Federal assistance can contribute to a more effective utilization of S&T knowledge by these jurisdictions. (3) Finally, a slow (and still only partial) acceptance has apparently occurred on the part of Federal officials that Federal support to state and local governments to utilize S&T knowledge more effectively is consistent with the realization of Federal objectives and, indeed, may even be necessary for their realization.

The Federal Government can and should play a major role in assisting state and local governments to improve their capabilities to utilize S&T knowledge. The positive contributions of Federal involvement have spanned several sequences specified in the prototypical diffusion paradigm. Federal assistance has helped to foster an awareness among the potential adopters—state and local governments—of the latent gains to be obtained through more effective utilization of S&T knowledge. Activities (conferences, research projects) supported by the Federal Government have helped to clarify the concepts of S&T knowledge for state and local governments.

Federal support has also created channels through which state and local governments can communicate with one another on topics relating to science and technology. This communication frequently deals with the attributes of specific innovations (the role of information channels in traditional diffusion paradigms). In addition, communication takes place concerning both the political aspects of generating change (whether it be the introduction of a new technology or the introduction of new types of information into legislative debate). It also occurs in regard to the positions that state and local governments should take as interest groups with the Federal Government in discussing modifications in the existing, federally dominated system of public support for domestically oriented R&D. The establishment of the Intergovernmental Science, Engineering and Technology Advisory Panel constitutes recognition by the Federal Government of the role of state and local governments in addressing issues relating to national science policies.

Federal assistance has been a major source of the funding for the initial adoptions of the "innovative" projects which are designed to "test" new arrangements for improving the S&T capabilities of state and local government. This assistance has been a necessary element in inducing state and local governments to undertake discrete projects directed at the utilization of S&T knowledge. Assistance of this type is likely to continue to be necessary in the near future because current approaches (a) do not adequately address the range of needs of all potential users (e.g., S&T mechanisms for "citizen" legislators), (b) do not fully address potentially significant levers for promoting a more effective utilization of S&T knowledge (e.g., changes in state regulatory and procurement procedures that might facilitate the acquisition of new technologies by local governments), (c) do not appear to have much promise of overcoming specific "barriers" to the generation of new public technologies (e.g., market aggregation), and (d) may not be cost effective.

More fundamentally, Federal assistance is necessary because state and local governments will not, and probably cannot, provide the resources required to create awareness, to test and to implement new approaches, and to communicate findings to comparable jurisdictions.

Federal assistance, however, may not always constitute an augmentation to self-initiated efforts by state and local governments. Federal assistance is generally offered on terms that require the recipients to meet certain Federal objectives. Some state and local organizations that provide S&T assistance to their constituents apparently now hold the view that their ongoing efforts were "torqued" to meet Federal agency objectives once they began to accept Federal grants. A consequence of Federal assistance was that over time these efforts became less useful to their constituencies. This torquing may result from Federal agency pressure to push specific ("high-") technologies or types of information through a delivery system in lieu of the items that were the early staples of the project. It may result from having the energies of the lead individuals ("entrepreneurs") in state or locally initiated projects redirected from development and provision of services to the user communities, to the pursuit and administration of Federal grants.

AN ASSESSMENT OF FEDERAL EFFORTS TO PROVIDE SCIENCE AND TECHNOLOGY ASSISTANCE

Major problems exist in the current ability of the Federal Government to provide the type of assistance that is needed by state and local governments. These problems relate to the limited Federal understanding of the processes of adoption, incorporation and utilization of S&T information in state and local governments, to the range of approaches typically supported by Federal agencies, to the systems of accountability and incentives confronted by Federal agencies, and to the organizational placement of S&T assistance programs at the Federal level.[10]

a. Federal S&T assistance to state and local governments is most often oriented either to the transfer of discrete technologies (hardware or software) developed under agency auspices (either an agency's own R&D efforts or those conducted under financial support from an agency), or to the dissemination of information relevant to an agency's mission objective. Federal agencies tend to relate to their functional counterparts within state and local government. There is considerable logic to this system of vertical networking, particularly as it facilitates the technical evaluation of new approaches and serves to develop stable channels of communication over time. The effectiveness of this approach is buttressed by recent findings which indicate that mission agencies—indeed, frequently only a few individuals within these agencies—are typically the locus for decisions to adopt new technologies within state and local governments.

This approach, however, contains a built-in barrier when efforts are made to shift support for the continuation of new approaches from Federal to state or local funding. The successful transfer of a specific innovation program from the status of a *demonstration project* to a *routine part* of the ongoing operations of state or local organizations frequently involves the creation of a broader coalition of participants than is necessary to obtain initial participation in the project. Thus, at the state level, successful utilization of S&T knowledge may involve not only a line agency which has been the direct recipient of Federal assistance, but also the legislature, which may be called upon to underwrite the costs of a technological system after a Federal demonstration grant expires or to enact legislation embodying state-of-the-art S&T knowledge initially conveyed to an agency. Strategies which foster adoption, such as limiting the number of participants, may work against implementation. Moreover, the leverage of Federal agencies at this transition juncture may be small. As Robert Yin has noted: "The major conditions that lead an innovation to become routinized all appear to be internal to the specific local agency."[11]

b. Federal agencies are organized and evaluated in terms of the conduct of activities that are consistent with their legislative mandates. Within an agency, alternative approaches to providing S&T assistance to state and local governments will likely be ranked in terms of bureaucratic imperatives concerning mandates and control over resources. These imperatives are reinforced by increasing demands for accountability and evaluation. Agencies must be able to show tangible products (outputs, realization of managerial objectives, or "nuggets") for their expenditures. These internal pressures tend to place a premium on programmatic approaches that lead to discrete, product-oriented activities, such as promotion of the use of a specific technology in a large number of jurisdictions. In terms of the criteria by which a Federal agency is evaluated, the contribution that this innovation makes to the performance of end users may be of secondary consequence.

From a bureaucratic perspective, the process-oriented programs that are most needed by state and local governments are inherently messier activities. Such programs do not lend themselves as readily to uniform project activities, to easily identifiable project milestones, or to evaluation efforts, since they will often reflect the unique set of circumstances surrounding each site. They tend to involve a greater degree of delegation of resources to the users. Thus, from the perspective of a Federal agency, such programs tend to be less attractive than alternative approaches (e.g., field agents) for pursuing similar objectives which contribute to its growth.

For example, a principal cause of the increased need by state and local governments for access to S&T information is the catalogue of Federal legislation passed within the past decade which requires state implementation. Federal agencies have indeed assisted state governments in developing S&T information systems (e.g., the MISTIC project operated by the National Conference of State Legislatures). However, because state governments do not always perceive Federal agencies to be neutral or objective transmitters of S&T information, they have sought to improve their own capabilities to obtain the information. Thus a situation is created in which a Federal agency has an incentive to promote the dissemination of information arising out of its activities, while the user communities want Federal support to develop information systems that will permit them to reach out to a broader set of sources of information. There is little incentive for single Federal agencies to support broad-based, multi-purpose information systems among the user communities. (Indeed, as suggested above, development of such an information capability among the user communities may tend to increase their abilities to critically assess and thus possibly to reject or disagree with information and advice generated by agency-specific information dissemination programs.)

A similar set of conflicts exists over the role of the Federal laboratories. Despite the several reports and policy statements that recommend increased utilization of these laboratories in an intergovernmental context and despite a number of examples of the successful utilization of their services, the commitment of the laboratories to provision of S&T assistance to state and local governments will likely remain marginal in the absence of explicit mandates or explicit modifications of the incentive systems confronted by both the laboratories and their employees.

The consequence of these patterns is that major gaps exist between the type of support needed by state and local governments and the incentives of Federal agencies to provide this support. Federal agencies will continue to relate to state and local governments along functional lines. The impact of this assistance will depend upon the scope of the intergovernmental element within each agency and the effectiveness of the related S&T programs (e.g., extent of user input into the development of R&D agenda; identification of relevant decisionmakers; content and format of "tech briefs"). These programs, however, will leave unmet the needs for programs which: (a) develop within state and local governments an understanding of the potential gains to be achieved through continuous and systematic access to and utilization of S&T information; (b) cross-cut functional areas; (c) are sufficiently flexible to accommodate the multiplicity of cases in which applications of low level technologies can improve public sector efficiency;

and (d) are designed to facilitate the *incorporation* as well as the adoption of discrete, federally supported endeavors into the ongoing operations of the user communities.

THE ROLE OF A CAPACITY BUILDING APPROACH TO INTERGOVERNMENTAL SCIENCE RELATIONSHIPS

In an intergovernmental context, capacity building refers to those Federal policies and programs that are intended to strengthen the capabilities of state and local governments in the range of activities which are required for improved public management, e.g., policy management, program management, and resource management.

Capacity building and linkage building are "putty clay" characteristics of such a wide variety of Federal programs that considerable difficulties exist in defining the relevant set of such activities and even more in assessing the effects of these activities. For example, the Advisory Commission on Intergovernmental Relations (ACIR) classification of all Federal grant-in-aid programs for fiscal year 1975 lists 30 programs whose primary objectives were held to be capacity building.[12] ACIR also notes that many other categorical grants may have contributed to capacity building at the state and local levels.

A capacity building approach implies that an objective of Federal assistance programs should be a strengthening of the capabilities of state and local governments to identify the S&T parameters of policy issues, to search for and secure the appropriate solutions and to implement these solutions in an effective manner. This approach implies that the Federal interest extends beyond simply making state and local governments more receptive customers for technologies or information developed under Federal sponsorship. Under this latter perspective, an offer of technical assistance by a Federal agency to a state or local agency may make the recipient better off in the context of the single problem being addressed, but it will tend to leave the agency in a state of continued dependency on Federal assistance when future problems arise. Alternatively, the Federal agency may offer assistance in such a way that the state or local agency becomes increasingly self-sufficient over time.

By its very nature, a capacity building program will be more open-ended and flexible than a categorical or competitive assistance program, and even more so than a "demonstration" project. Moreover, since it is the users who will determine what works best, the final products of such a program, given the diversity in initial capabilities among state and local governments, are likely to be quite varied, thus complicating efforts to assess or evaluate outcomes.

Recent research findings of innovation and implementation processes in state and local governments have highlighted the important influence of jurisdiction-specific "selection environments" or diffusion milieus on the extent to which new technologies are adopted and incorporated. In light of these findings, it would be difficult to generalize from the effects of capacity building programs, broadly defined, to those which are intended to promote the effective utilization of S&T information. More to the point, little systematic evaluation has been made of those capacity building programs which are "close" to the S&T area. For example, in its evaluation of the State Science, Engineering and Technology Program, the Stanford Research Institute (SRI) cites four Federal programs as having strong capacity building elements: Department of Housing and Urban Development's 701 program (Comprehensive Planning Assistance), Economic Development Administration's 302 program (State and Local Economic Development Planning), Environmental Protection Agency's 208 program (State and Areas Water Quality Management Planning Program), and the Coastal Zone Management Program.[13] Yet for these programs, little systematic evaluation exists. The few evaluations that have been conducted have pointed to the problems associated with such programs, notably uncertain political support accorded them within the recipient organization, overlapping jurisdictional issues, and perhaps most important of all, lack of technical expertise necessary to implement the programs.

Assesing the Impact of Capacity Building Programs

Assessment activities of the type needed to gauge the impact of capacity building programs are only now being conducted, so the judgments presented here are both tentative and clearly based more on my own analysis and observations than on formal studies. Moreover, these statements relate to strategies or to broad programmatic thrusts, rather than to the contributions of individual projects.

My summary assessment, organized about levels of government, is as follows:

STATE EXECUTIVE

Capacity building at this level of government entails the greatest degree of complexity of any of the three levels of government considered here. This complexity derives from three sources. (1) The state executive branch is involved more extensively in the separate elements of the intergovernmental science system (e.g., R&D priority setting, technology transfer) than are other jurisdictions, and is involved with issues that are not major concerns of other branches (e.g., R&D funding). (2) The number and size of the different organizations involved in the generation and utilization of S&T knowledge are larger than in other branches, so that issues of coordination and control (for example, among a line agency, the budget office, and a governor's office), are relatively more important in determining whether an increase in the capability of a subunit is translated into a general improvement in the capability of the total organization. (3) The executive branch has a more complex history of previous efforts to utilize S&T knowledge—including an early (misplaced) emphasis on state science advisory boards and state science and engineering foundations, which now may constitute impediments to renewed executive interest in S&T capacity building programs.

The principal achievements in capacity building at the executive level are the *recognition* that: this capacity must be integrated into the policymaking process within a given state; the formal organizational characteristics of such a capability will vary with the policymaking process selected by a governor; and the involvement today of senior-level officials within the executive branch in efforts to develop this capability. There has been an important (if largely unrecognized and unappreciated) shift in the formulation of the meaning and role of an executive level S&T capability and of the participants in state level efforts to secure this capability.

STATE LEGISLATURES

Significant improvements have been made by state legislatures to address the S&T components of policy issues. Ten years ago it would have been difficult to identify a state legislature which had a discrete organizational unit or liaison arrangement specifically directed at obtaining S&T information or which employed staff personnel with advanced degrees in the physical or biological sciences. Today such units, arrangements, and individuals exist in approximately 17 states, albeit primarily in those states with "professional" legislatures. Moreover, while a substantial portion of the current legislative involvement is underwritten by Federal grants, a transition (institutionalization) to state support has occurred in some of the legislatures. The state legislative community is also becoming increasingly sophisticated in its understanding of an S&T capability, in its ability to evaluate the suitability of alternative capacity building mechanisms in the context of different legislative environments, and in its ability to articulate its position in discussions with Federal representatives concerning the design of any new system of intergovernmental science relationships.

LOCAL GOVERNMENTS

Capacity building endeavors for local governments now center about the establishment of a series of innovation networks and groups. A major effort along these lines, the Urban Technology System, is currently being evaluated. Multiple objectives are being pursued under the various local government networking projects so it is not possible to provide a single assessment concerning their impacts. The networks appear to have great potential utility as channels for bringing problem-solving expertise to municipal problems; they may also contribute to an awareness within city governments that acquisition of a comparable internal capability is a productive expenditure of local revenues. Insufficient evidence has been generated from existing network projects to permit a conclusion concerning their effectiveness as channels for the lateral transfer of information or of innovation. In general, the diffusion/technology transfer potential of networks is probably overstated. Indeed, the emphasis on this aspect of the networks appears to detract from their accomplishments and potential in other roles. Networking appears to be an effective method for generating R&D agendas. The issue here is less on the input side than on the likely effect these agendas will have on the behavior of Federal agencies. The potential of the networks as a means of aggregating municipal markets to stimulate private sector R&D has yet to be demonstrated, and appears quite limited.

Capacity building programs inevitably encounter two principal barriers at the Federal level. First, the programs are designed to assist state and local governments to improve their internal operations. Self-interest (or a sense of responsibility) on the part of these organizations would appear to dictate that they provide themselves with these very same capabilities from their own resources. (This argument is given extra force today by the perspective which juxtaposes the Federal Government's attempt to reduce the size of its budget deficit with the surplus budget position of some state governments.) Moreover, even if aggregated broadly across Federal agencies, the sums proposed for S&T capacity building programs appear insignificant relative to total Federal grants to the states (and to the cities via pass-through provisions), again raising the question of why current grants are not adequate to develop the sought-after capabilities. Second, the capacity building concept is a fuzzy one, both programmatically and administratively. It is difficult to specify parameters for such a program (e.g., how much capacity is needed? what is the dollar outlay necessary to obtain this level of capacity? how does one define a list of eligible recipients? and how does one know when the sought-after level of capacity has been obtained?). More to the point, little effort has been given to providing such a specification.

These questions are not new ones (recall the debates over the contours of the poverty program—how poor is poor?). However, they take on a special importance in the S&T field today because the meaning of capacity in this context is qualitatively less precise (i.e., less quantifiable) than are other Federal objectives (e.g., adequate health, adequate income) and because, except for the recommendations contained in the SRI's evaluation of the SSET program,[13] there are few specifics of what the parameters for such a program might look like.

FEDERAL GOVERNMENT

The main issue is the degree to which the Federal Government should support capacity building in state and local governments for utilization of science and technology.

Despite the strictures mentioned above, a Federal capacity building approach to the use of S&T knowledge by state and local governments is necessary for the following reasons:

1. Productivity improvement in state and local government frequently involves the application of "low-technology" or "good housekeeping" solutions. These solutions are too mundane and too varied to be the subject of specific Federal technology transfer programs, yet they may constitute the "appropriate technologies" that are most frequently needed by state and local governments.
2. An emerging corpus of diffusion research points to the importance of the processes of "reinvention," "specification," and "adaptation" in the sequences from adoption to routinization. States and local governments use technologies to achieve ends not foreseen by the suppliers (either private manufacturers or Federal agencies) and in altered forms. A necessary condition for this "matching" to occur is that these jurisdictions possess the requisite technical or managerial skills to exploit the potential efficiencies latent within new technologies.
3. The range, heterogeneity and urgent, but often temporary, quality of state and local government needs for S&T information—air pollution in the late 1960s, toxic wastes at present—give an added value to approaches in which these users can define the work agendas of the suppliers of S&T knowledge and can influence the format and time periods in which answers are returned.
4. Capacity building may be a necessary complement for the successful implementation of more conventional Federal technology transfer and information dissemination programs. It must not be forgotten that a substantial portion of current state and local needs for improved access to S&T knowledge derives from the passage of Federal legislation which requires state and local actions at the implementation stage. This pattern is rife with conflict. Federal offers of technical assistance may be rejected by state and local governments, not because these organizations make decisions solely on "political" as compared with "technical" criteria, but because they do not view Federal agencies as sources of "objective," "technical" solutions to the tasks mandated for them. For many reasons, including the view that they have technical capabilities comparable to those of Federal agencies, state governments may seek to pursue their own solutions.

S&T capacity building for state and local government is itself a programmatic innovation. The failures and limitations of efforts undertaken along these lines during the past decade cannot be ignored. Yet these shortcomings are in part the unavoidable experiences associated with novel arrangements, regardless of whether they are technologies or social institutions. In the conceptual framework within which Federal policies are formulated, there is a danger of comparing the muddled, inconclusive, and often disappointing course of development of new programs with the seeming inexorability of well-established programs. For example, from the very beginning of S&T capacity building endeavors, agriculture has been held up as a model for emulation. The agricultural system has several of the features which are considered objectives of capacity building endeavors: user-oriented R&D, user input into the orientation of Federal R&D, etc. The congressional debates, statements of agricultural leaders, scientists, university presidents, faculty, state officials, and other participants involved in the evolution of the current system, document disputes and controversies concerning institutional roles, missions, relationships, and funding. Approximately 70 years were required to institutionalize the system. In brief, the entire set of issues confronted today in S&T capacity building had to be thrashed out.[14]

NOTES

1. For additional perspectives on these same questions, see Roessner, J. David, "Federal Policy and the Application of Technology to State and Local Government Programs," *Policy Analysis* 5 (Spring, 1979), pp. 181-200. See also Carey, William D., "The Relationship Between Federal, State and Local Government Support for Research and Development," in *Priorities and Efficiency in Federal Research and Development*, papers submitted to the Subcommittee on Priorities in Government, Joint Economic Committee, 94th Congress, 2nd Session.

2. Summaries of these research findings are presented in Feller, I., "Research Findings and Issues in the Design of an Intergovernmental Science System," paper prepared for the Office of Science and Technology, Executive Office of the President, November 1978, and Roes-

sner J., David, "Technological Diffusion Research and National Policy Issues," paper prepared for the ORSA/TIMS, Joint National Meeting, Los Angeles, November 1978.

3. For a description of these programs, see Federal Coordinating Council for Science, Engineering and Technology, *Directory of Federal Technology Transfer*, Washington, D.C.: U.S. Government Printing Office, 1977; also Roessner, J. David, *Federal Technology Transfer: An Analysis of Current Program Characteristics and Practices*, a report prepared for the Committee on Domestic Technology Transfer, Federal Council for Science and Technology, NSF-76-400, 1975.

4. Yin, Robert, "Contemporary Issues in Domestic Technology Transfer," in *Domestic Technology Transfer: Issues and Options*, U.S. House of Representatives, Committee on Science and Technology, 95th Congress, Second Session, Serial CCC, Vol. I (November 1978), pp. 3-40.

5. U.S. General Accounting Office, *Technology Transfer and Innovation Can Help Cities Identify Problems and Solutions*, Report to the Congress, August 6, 1975. National Academy of Public Administration, Report of a Panel, *The Baltimore Applications Project: A New Look at Technology Transfer*, March 1977, prepared for Goddard Space Flight Center, NASA.

6. Feller, I. and Flanary, Patricia, *Diffusion and Utilization of Scientific and Technological Knowledge within State and Local Governments*, State College, Pa.(The Pennsylvania State University: Institute for Policy Research and Evaluation, 1979), Report to the National Aeronautics and Space Administration under Contract NAS5-24329.

7. Helminski, Edward and Muchmore, Lynn, *SSET at Mid-Stream, An Interim Assessment*, Washington, D.C.: National Governors' Association, 1978, Report to National Science Foundation, Contract No. C-ISP-20657.

8. Menzel, D., Friedman, R., and Feller, I., *Development of a Science and Technology Capability in State Legislatures*, State College, Pa.: The Pennsylvania State University: Center for the Study of Science Policy, 1973.

9. See, for example, *Proceedings of A Symposium on the Institutionalization of Federal Programs at the Local Level*, McLean, Virginia: MITRE Corporation, 1979.

10. Portions of this section are taken from Feller, I., *Diffusion and Utilization*, op. cit.

11. Yin, R., *Changing Urban Bureaucracies: How New Practices Become Routinized*, Santa Monica, California: Rand, Report to the National Science Foundation, Grant No. PRA76-15207.

12. Advisory Commission on Intergovernmental Relations, *Categorical Grants: Their Role and Design*, Washington, D.C.: ACIR, 1977.

13. Waldhorn, Steven, Golub, James, King, James, and Malek, Roger, *Increasing the Capacity of State Governments to Access and Use Scientific, Engineering and Technical Resources*, Menlo Park, California: SRI International, 1978.

14. Feller, I., "The Science and Technology Experience at the State Level: An Overview," paper prepared for the National Governors' Conference's Workshop on the State Science, Engineering and Technology Program, Atlanta, Georgia, November 1977.

2 Science and Technology in State and Local Governments: The Federal Role

by Robert K. Yin*

SUMMARY

The purpose of this paper is to review existing Federal efforts in supporting R&D for state and local governments, and to suggest the ways in which such support can be improved in the next five years.

Federal efforts, while fragmented among many different mission agencies, nevertheless tend to focus on either of two strategies: technology-push efforts, and demand-pull efforts. In the former, support has been provided to promote a wide variety of technological innovations, each of which is presumed to improve local service practices. In the latter, support has been provided to increase the organizational capabilities of state and local governments, so that their staffs are more oriented toward using science and technology. In addition, a now common demand-pull practice at the Federal level is to engage state and local officials in setting Federal priorities, with the recent efforts by the Intergovernmental Science, Engineering, and Technology Advisory Panel (ISETAP) being among the more prominent attempts.

As for future support over the next five years, R&D activities must first be distinguished from two other types of activities—capacity building for state and local organizations, and program analysis by such units. Greater emphasis is recommended, especially for such Federal R&D agencies as the National Science Foundation, on the support of R&D rather than the support of capacity-building or program analysis activities. The paper proposes several topics in urban social R&D that are worthy of further investigation, including the operation of neighborhood housing markets, industrial revitalization, and demographic trends that are affecting urban lifestyles. Similarly, several topics in urban technological R&D should be pursued in the future, mainly focusing on the factors associated with successful innovation (rather than the more traditional orientation toward identifying barriers to innovation).

Federal policies also include the way that R&D is to be supported, and not merely the substance of the R&D agenda. First, there is a need for developing and assessing *programmatic* strategies and relying less heavily on the success of individual R&D *projects*. Second, there is a need for new policies to improve the overall quality of R&D that is conducted on state and local problems.

THE CONTEMPORARY FEDERAL SCENE

Using Federal Science and Technology to Serve State and Local Needs

A major aim of Federal policy in science and technology (S&T)[1] is to serve the needs of state and local governments. Current S&T policy has been oriented toward the support of "public" technology programs, first

* Visiting Associate Professor of Urban Studies and Planning, Massachusetts Institute of Technology.

highlighted by the President's message of 1972 and reinforced by the March 1979 message.[2] In the earlier message, the following goals were established for Federal support for public technology:

- To develop systematic ways of communicating to the appropriate Federal agencies the priority needs of state and local governments;

- To assure state and local governments adequate access to the technical resources of major Federal R&D centers; and

- To develop methods whereby the Federal Government can encourage the aggregation of state and local markets for certain products so that industries can give government purchasers the benefits of innovation and economies of scale.

More recently, the President's message of 1979 reiterated several of these themes, but added the distinctive flavor of creating intergovernmental "partnerships."

As a result of this policy orientation, Federal executive agencies have supported a wide variety of S&T efforts, including programs sponsored by the Department of Housing and Urban Development (HUD) (urban services generally), Department of Transportation (transportation), Law Enforcement Assistance Administration (public safety and criminal justice), Department of Commerce (fire safety and prevention), Department of Health, Education, and Welfare (educational services), and National Science Foundation (NSF) (science utilization by state and local governments). However, no overarching strategy has emerged at the Federal level, and there have been continuing attempts to coordinate Federal efforts, with the establishment of the Intergovernmental Science, Engineering, and Technology Advisory Panel (ISETAP) having been only the most recent attempt (see P.L. 94-282).

In spite of this lack of coordination, the numerous Federal S&T efforts can nevertheless be said to fall into two general categories: *technology-push* policies and programs, and *demand-pull* policies and programs. The following sections are intended to give a brief flavor of these current activities as a basis for discussing emerging priorities over the next five years.

Technology-Push Efforts

State and local governments have implemented or attempted to implement a wide variety of new technologies, many with Federal support. This diversity of technologies is exemplified in Table 1, which shows illustrative technologies in five service areas (criminal justice and fire, transportation and public works, health, education, and planning and other), as divided among three general types of technologies (hardware, computer systems, and data analysis techniques).[3]

From the Federal point of view, support has been provided to promote specific innovation *projects* as well as large-scale innovative programs.

PROJECT SUPPORT

Typical support of innovation projects, found in education and criminal justice, has focused on the identification and dissemination of "exemplary practices." These practices represent innovations that have been tested at a demonstration site, evaluated by some third party, and then communicated in easy-to-understand form to potential users at other sites. In education, this process has been administered under the Joint Dissemination Review Panel, which gives "Good Housekeeping" seals of approval to specific innovations that are then promoted by the National Diffusion Network, helping local sites to implement these innovations. In criminal justice, there is a similar program on Exemplary Projects; and for other municipal services, the International City Management Association has likewise operated an Innovation Transfer Project with support from NSF.

It should be noted that these project-oriented efforts have increased in their sophistication and scope in recent years. Whereas Federal funds were used only to support the initial R&D efforts in earlier years, these contemporary efforts have all begun to deal with the utilization and implementation process, with the knowledge that user capabilities are an essential component of technology transfer. Furthermore, many of these projects have been the subject of third-party evaluations, so that there is now an extensive literature on the factors involved in successful technology transfer efforts and the ways in which barriers to innovation can be reduced.[4]

PROGRAM SUPPORT

At the more programmatic level, the Federal efforts have included such well-known activities as HUD's Operation Breakthrough, Commerce's Experimental Technology Incentives Program (ETIP), and emergent programs in the energy field that have involved the formation of field stations and technology agents. The results of these efforts have been more difficult to assess, and in some cases there have been clear failures. Thus, for instance, HUD's Operation Breakthrough was designed to meet the needs of residents through modularized housing construction. The program was discontinued after sev-

Table 1. Illustrative Technological Innovations Attempted in Various Municipal Services.

Criminal Justice and Fire	Transportation and Public Works	Health	Education	Planning and Other
		Hardware		
Police helicopter patrol Automated status reporting and emergency signaling system for mobile patrol units Rapid water Universal emergency telephone number (911)	Bay Area Rapid Transit System (BART) Dynamic traffic safety displays (lighted signs, improved road illumination, buzzers, etc.) Dial-a-ride system Mercury and sodium street lights Grider sewage pumps Godzilla vehicle system for fully mechanized refuse collection	Two-way closed circuit television for psychiatric consultation Vital function telemetry Technique to extract acid from urine to measure lead in humans New equipment to streamline ordering, procuring, and distribution of medication Methadone maintenance program	Closed circuit television Multimedia educational program Televote system Silent alarm system	Two-way cable television for intergovernmental communication Computer-assisted information system to dispatch multi-agency response vehicles
		Computer Systems		
Remote access infrared data file and search system Criminal justice identification system	Information system for dynamic deployment of police traffic patrols in response to traffic flows	Automated medical history system (AMH) Patient monitoring system Computer-assisted system to analyze electrocardiograms Automated patient registration, identification, and appointment system Blood inventory and information system (BIIS)(computerized clearinghouse of information on regional blood inventory) Hospital information system (medication charts)	Computer-assisted instruction School information system (SIS) to assist decisionmaking in school administration Teleprocessing information system to provide data processing to multiple school districts Computer-assisted testing Information retrieval system to assist decisionmaking in five-county area Computer integrated into science and math curricula	Integrated municipal information system Minicomputer to automate library acquisition system Computer program to assist in housing-policy decisions Computer system to process complaints and violation notices and issue activity sheets to housing inspectors
		Data Analysis		
Computer programs to allocate manpower and devise proportional police rotation schedules Technique to reduce court appearances required of police Police deployment technique Analysis of fire-fighting resources deployment Mathematical models to test effects on service of fire company locations	Computer algorithm to assign bus drivers to routes Cost effectiveness analysis of a water supply system Manpower scheduling system	Model for distribution of mental health funds		Urban dynamics model EMPIRIC (land use planning model) Computer simulation for renewal policymaking Computer simulation model to test new investment and urban change hypotheses Computer simulation of emergency service operations

SOURCE: See Note 3.

eral years, having encountered such difficulties as the following:

- Fragmented markets precluded significant industrial investments;
- Overlapping state and local jurisdictions minimized coordinated efforts;
- Opposition was encountered from groups with vested interests, such as labor unions and trade associations;
- Building codes were not uniform; and
- Uniform testing and performance standards for materials, design configurations, and fabrication methods could not be developed.[5]

One supposition is that Federal agencies have not yet learned how to develop effective technology transfer *programs*, and the development and assessment of such programs is the major topic of a recent congressional document that also exhaustively reviews technology transfer in both the public and private sectors.[6]

Demand-Pull Efforts

Alongside of the Federal efforts to support specific technology projects or programs, there have been major efforts to improve the internal capability of state and local governments to use science and technology. A leading example of this type of effort has been NSF's Intergovernmental Program, which is the successor to two earlier NSF programs, the Intergovernmental Science Program and the R&D Incentives Program.[7]

IMPROVING LOCAL CAPABILITIES TO USE R&D

Demand-pull projects give greater emphasis to organizational *capacity building*—i.e., the ability of local organizations to analyze their own needs and to meet their objectives by using R&D—than to the use of any specific technological products. Thus, for instance, the Pennsylvania Technical Assistance Program (PENNTAP) was formed to provide technical information on a statewide basis to both public and private institutions. Similar activities have been aimed at serving governors (e.g., the Michigan science adviser), state legislatures (e.g., the Model Interstate Scientific and Technical Information Clearinghouse—MISTIC), and mayors (e.g., the Philadelphia science adviser). The effectiveness of these organizational efforts is very difficult to assess.[8] Nevertheless, a continuing Federal policy concern has been that appropriate assessment methods ought to be developed and applied before larger investments are made.

As part of the demand-pull approach, Federal agencies have also attempted to determine the needs of state and local governments more explicitly. Thus, in the management of new R&D projects, a common procedure has been the establishment of a user panel, which is involved in the design and conduct of the research project. This has been an especially dominant activity of the Urban Consortium (Public Technology, Inc.), as well as representatives of state governments. However, recent studies of various user communities—e.g., state legislators, city managers, or county executives—have shown that the identification of user needs is not an easy process. The potential users are much more diverse and fragmented than is often anticipated; the users cannot necessarily articulate their needs very well; and the needs, when articulated, are not clearly susceptible to technological resolution.[9] Nevertheless, the solicitation of user views about their needs has become a prominent element of Federal policy.

ASSESSING LOCAL NEEDS THROUGH THE INTERGOVERNMENTAL SCIENCE, ENGINEERING AND TECHNOLOGY PANEL (ISETAP)

One of the most ambitious efforts to deal with demand-pull factors has recently been undertaken by ISETAP and therefore deserves to be described in greater detail. ISETAP began by approaching state, regional, and local authorities, public innovation groups, and the many public interest organizations that represent state and local interests—e.g., Public Technology, Inc., the National Governors Association, the Conference of Mayors of the United States, and the International City Management Association. These organizations were asked to identify the most pressing state and local problems that Federal R&D programs and policies could address. The result of this initial request was a list of some 800 problems, including statements on almost every aspect of human endeavor: transportation, fire suppression, management of growth, management productivity, health, environment, and energy. These responses were categorized by ISETAP into ten functions, with each function subsequently covered by an ISETAP task force. The categories are listed in priority order below:

1. Community and Economic Development
2. Energy
3. Environment
4. Fire Safety and Disaster Preparedness
5. Health and Human Resources: The Elderly
6. Health and Human Resources
7. Management, Finance, and Personnel
8. Police and Criminal Justice
9. Public Works and Public Utilities
10. Transportation

The rôle of each task force is to articulate state and local needs in the relevant topic area, and to recommend ways in which Federal programs can be more responsive to these needs. In addition to the work of the task forces, the American Association for the Advancement of Science was asked by the National Science Foundation, on behalf of ISETAP, to explore the problem areas under each functional topic. Workshops were held during 1978-1979 by AAAS on each topic, to examine the opportunities for Federal R&D and technology transfer programs and policies.

Concurrently, ISETAP has been interacting with Federal agencies in ongoing R&D programs and policies. Among these agencies are the Administration on Aging (Department of Health, Education, and Welfare), the Department of Commerce, the Department of Housing and Urban Development, the Department of Energy, the Environmental Protection Agency, and the National Science Foundation. In 1978, to strengthen ISETAP's functions, the Office of Management and Budget was designated as being responsible for implementing the recommendations from the Panel. Thus, in theory, ISETAP has had the ability to affect resource allocation decisions about future Federal R&D efforts.

The Panel's activities, briefly sketched above, have attempted to bring together a very complicated set of relationships involving innovation groups, umbrella organizations for governors, mayors, city managers, budget officers, the police, fire fighters, and other interest groups. Each group has its own set of values, needs, motivations, and policies, and the ISETAP efforts must therefore be primarily considered a political and administrative rather than scientific venture, which is in fact also the major characteristic of most demand-pull efforts.

Summary

This section has briefly portrayed existing Federal programs that use science and technology to improve state and local governments. The existing programs tend not to be centralized or coordinated, with each Federal agency pursuing its own agenda. The efforts do, however, generally fall into one of two categories—technology-push and demand-pull efforts—with the experience of the ISETAP Panel an example of the latter. These two approaches will be continued in the next five years, but with several subthemes that may be discernible, and these are described in the next section.

DEVELOPING FEDERAL R&D PRIORITIES FOR THE NEXT FIVE YEARS

Federal efforts to assist state and local governments will continue to take at least two dominant thrusts reflected by the demand-pull approach: (1) capacity building of state and local agencies and (2) provision of assistance to conduct what are essentially program analysis rather than research functions. As for the technology-push efforts, however, it may be that the appropriate time has been reached for reassessing Federal priorities and for expanding the Federal agenda to deal with state and urban problems more generally, rather than limiting support to the assistance of state and local governments. These governments, after all, only exist to serve citizens and their needs, and Federal policymakers would be remiss if they merely targeted their efforts toward assisting units of government without considering broader problem areas. The following sections review the potential priorities for all three topics: organizational capacity building, program analysis activities, and state and urban R&D priorities.

Capacity Building

The concerns for improving the capabilities of state and local governments to use S&T have a long history, dating back at least as far as 1948, when "little Hoover commissions" tried to identify ways of improving the cost effectiveness of state and local activities.[10]

One priority of the little Hoover commissions was the issue of adequate staffing—one which has still not been fulfilled in the past 30 years. It has been noted that, even now, only one out of every ten state and local employees is prepared to participate in, accept, and understand the implementation of technological innovations.[11] A further problem is how to determine the availability of personnel whose training has given them the advantage of understanding the innovation process. Cost effectiveness may then be determined by the availability and allocation of this type of manpower.

The capacity building approaches remain an elusive strategy for pursuit by Federal agencies. First, the most effective strategies have not yet been identified. There have been numerous attempts, for example, to improve capacity by encouraging personnel transfers (e.g., the IPA program), and to instigate new organizational coalitions among local jurisdictions (e.g., the California Four Cities program), or to create specific networks of practitioners and researchers (e.g., the National Diffusion Network in education). However, these various strategies have not been well articulated, nor have they been compared for their relative effectiveness.

Second, the assessment of capacity building efforts is an ephemeral activity at best. Capacity building, by its very nature, implies payoffs that are diffuse and long term. Most evaluation research techniques, however, are aimed at demonstrating specific payoffs in a short period of time. Whether new methods for assessment can be developed remains to be seen. Organizational behavior in general is notoriously difficult to measure or interpret. In the state and local setting, where job patronage and employment are legitimate goals that parallel service delivery objectives, the tasks are that much more difficult.

Third, the Federal Government may be in a peculiarly weak position to promote capacity building efforts. Again, by its very nature, capacity building implies the use of resources in a manner deemed appropriate by local decisionmakers. Any Federal role would seem to be limited to a spirit akin to the general revenue sharing programs, where Federal funds are disbursed with a minimum of Federal direction or regulation. In contrast, Federal capacity building efforts up to the present time have tended to be dominated by project-type awards, where specific proposals are made for grants or allocations, and where Federal managers play an active monitoring and evaluation function. Such tactics would appear to be inappropriate to the true spirit of capacity building, and are rightly seen by many state and local officials as thinly veiled attempts to promote *Federal* rather than local agendas.

Program Analysis

When state and local officials are candidly confronted about their R&D needs, they will admit that *program analysis* rather than R&D *per se* is what is most helpful in dealing with their policy problems. The distinction between program analysis and three types of R&D is illustrated in Table 2.[12] Essentially, program analysis calls for the specific analysis of a real-life problem, constrained by existing political conditions, costs, and time limitations. The analyst's function is to develop the best set of alternatives, based on empirical evidence, that a decisionmaker can pursue. However, program analysis cannot really be regarded as a research activity in the traditional sense, because research that seeks to identify scientific truths cannot be constrained by arbitrary political, economic, or time limitations.

One distinctive difference between program analysis and R&D is in the nature of the final report or findings from the efforts. The value of program analytic findings peaks with the opportunity to make a decision; after this point has been passed, a fresh program analysis would have to be conducted to suit new political realities (and the old program analysis would only be of historical value). In contrast, the value of R&D findings should increase with the passage of time and with the occurrence of new studies that either corroborate or contradict the initial findings. Findings that withstand these subsequent challenges become accepted as robust, scientific facts, and the original reports are often more dearly valued and cited frequently.

A potential dilemma not squarely faced by most Federal S&T programs is that state and local needs may be dominantly of a program analysis sort, even though Federal policies happen to be geared toward the support of

Table 2. Four Types of Research Activities, Illustrated by Topics on Community and Economic Development

Illustrative R&D Topic	Type of Research			
	Basic Research	Applied Research	Evaluation Research	Program Analysis
Household composition and residential turnover	What is a household?	Develop methodologies for enumerating households; distinguish gentrification from upgrading	Assess impact of CDBG program on residential turnover rates	Provide information to promote or retard turnover in a specific neighborhood
Inventory of appropriate technology	Define various inventions and prototypes	Indicate markets, costs and benefits, etc., of commercial products	Give assessments of specific technologies or programs	Advise citizens on procuring and using specific technologies appropriate to needs
Decentralization of urban services	What is the relationship between citizen and government?	Conceptualize and describe various forms of decentralized governments	Evaluate specific decentralization experiences	Determine whether and how a particular city should decentralize
Role of private lending institutions	Neighborhood investment and disinvestment process, including role of neighborhood "confidence"	Develop models to predict key investment and disinvestment points	Evaluate role of Urban Reinvestment Task Force	Implement Community Reinvestment Act; determine declining patterns in a specific city

SOURCE: See Note 12.

R&D. This dilemma was highlighted at a recent workshop of research investigators and state and local officials, where each group was asked to establish priorities for "research" that would serve state and local needs in the fields of community and economic development. The state and local officials identified their three illustrative priority research topics: The assessment of socioeconomic and housing trends for neighborhoods; the determination of tasks that local government could appropriately assign, share, or contract with neighborhood groups; and the identification of incentive tradeoffs, such as tax abatements or zoning concessions, that would result in private investment in neighborhoods. More importantly, however, the state and local group concluded that:

> In each of these three cases, the type of research that was felt to be most needed was what has been characterized as *program analysis*. While we recognize the importance of basic, applied, and evaluation research, particularly as an input into program analysis, we feel that the immediacy of local government problems suggests that the greatest need at this point . . . is for good program analyses. (Emphasis added.)[13]

The point of distinguishing program analysis from these three R&D activities is that the analytic requirements may necessitate different types of management by Federal agencies. The imposition of a user review committee during the conduct of an R&D project, for instance, would not seem to be relevant to a basic research effort, nor should the quality of basic research be judged by the utilization of specific results. Instead, quality control of a more traditional, academic sort should be imposed. For an R&D project falling into the applied research category, user requirements are potentially relevant, but not in any major way, so that any project advisory group should still mainly consist of research peers. At the same time, one would not impose some of the traditional academic standards on the evaluation or program analysis types of research. Thus, the important outcomes from these types of research are a set of recommendations and the substantiating evidence; no academic or even final report need necessarily be written,[14] with project success judged mainly by the *utilization* rather than the *publication* of results.

Basic and Applied R&D

Neither the capacity building nor program analysis functions should in reality be considered R&D activities. For this reason, it is ironic that NSF, whose main mission is to support S&T, has tended to take the lead role in supporting capacity building and program analysis. These activities have dominated many of the awards made by the Intergovernmental Program as well as some of those by the Applied Science and Research Applications program. Such support would seem to be more suitable to a mission agency such as HUD, whose functions include the provision of support to state and local governments.

Instead of pursuing these types of activities, Federal R&D agencies such as NSF, the National Institute of Mental Health, and the National Institute of Education should consider embarking on a more sustained effort in a third arena—the support of basic and applied R&D to deal with state and local problems. Note that, as previously mentioned, such problems are regarded as being an attribute of residents, communities, and society at large, rather than merely being the captives of units of government. In fact, any new R&D priorities should begin with the assumption that states and cities are basically geographic areas dominated by private activities—including business enterprises, housing markets, and social relationships among neighbors—rather than strictly governmental policies.

URBAN SOCIAL R&D

At the urban level this image runs counter to, and is probably less popular than, the prevailing government service view of the city, in which municipal finances, municipal services, and government regulation are seen as the dominant urban concerns. Yet, considerable evidence exists that a *self-regulated city* was not only the essence of the traditional American city of the nineteenth century, but is also the underlying condition in today's cities.

As but one illustration, take a topic that has been of major importance to residents, policymakers, and researchers alike: public safety. In matters concerning both policing and fire extinguishment, private affairs dominated the early urban scene. Some efforts were purely voluntary. For instance, in policing, self-help movements tended to materialize in periods when a community's revered values, social structure, property, and wealth were threatened. Two phases of American vigilantism were distinguished: a first stage that occurred in the late 1700s and early 1800s when citizens in lawless frontier towns organized to apprehend criminals and marauders, and a second stage of urban neovigilantism that began in the mid-1800s and mainly involved the persecution, oftentimes violent, of racial and ethnic minorities.[15] Similarly, in matters of fire prevention and extinguishment, volunteer fire companies were long the backbone of American fire service. These volunteer companies were gradually displaced in larger cities by a paid force of civil servants, but even today such volunteer companies (or volunteer duty) are the norm in suburbs and smaller cities.

Even where formal public services were initiated, however, public safety was still a functionally private affair, in which the relationship between servers and served followed a form of free market exchange.[16] Services were not so much delivered as they were bought, sold, or negotiated. Thus, policemen on the beat had substantial discretion and were largely outside of the control of central administrators. The following aptly describes the supervisory problems at the turn of the century:

> Once the men were dismissed from roll call, their supervisors had no certain way of controlling what they did during their tour of work. The sergeants, who were called roundsmen in Philadelphia and Brooklyn during the early nineteenth century, frequently assigned men 'meets,' prearranged times and places where the supervisors could visually check on them. The only way a roundsman had of discovering what his men were doing was to follow them around and make inquiries among the people who lived and worked on the beats. If he wanted to watch a man at work, he could, and frequently did, accompany him, but this obligated him to neglect his other duties. The men were also isolated from each other, and their only way of attracting attention in moments of distress was by swinging the large rattles which city policemen had been carrying since the sixteenth century.[17]

The new communications technology that has evolved since then has not necessarily overcome these problems. Although call boxes followed telegraph networks, radio cars followed both, and various "pulling" systems have been adopted, no amount of communication has been able to place the police officer under direct, constant supervision. So officers have continued to "coop," to take bribes, to react independently to dangerous situations, and occasionally to be assaulted; and administrators can still do precious little to regulate these encounters. Moreover, the street-level world of the urban police officer provides little clarity or agreement about the nature of the service "problem" or its appropriate solution. What is intolerable vice to one segment of the community may be a pleasurable pastime or means of employment for others. And so police officers have had to deal with numbers runners, prostitutes, and owners of after-hour bars, with the knowledge that citizen demands and preferences are sharply divided. In these cases, police officers have had to mediate conflicting interests and apply ambiguous laws in deciding how to act or, for that matter, whether to act at all.

The importance of private efforts in public safety persists to this day. This is reflected in the continued rise of the private security industry,[18] the increasing frequency of citizen efforts in forming their own patrols,[19] and in the establishment of citizen groups,[20] as well as the general notion that residential "eyes and ears" are an important—and perhaps major—deterrent to crime.[21]

Overall, the view of the self-regulated city is that general improvements in the quality of urban life are determined by the collective activities of private individuals and organizations. The most relevant R&D should, therefore, focus on the basic social control, economic, and demographic trends that are vital to the quality of life and that are likely to influence the city in the future. Thus, further R&D could be conducted on such topics as:

- What makes a good neighbor, and how do residents develop incentives for caring about their homes and public places?
- How do neighborhood housing markets operate and, in particular, what role does *neighborhood confidence*[22] play in determining the viability of a given neighborhood?
- What are the basic economic conditions that favor industrial revitalization in a city?
- What are the urban implications of such demographic shifts as the increase in households without children or households with single-parent families?

For these and similar topics, R&D efforts should identify the dynamics of institutional behavior, interpersonal relationships, and individual decisionmaking (e.g., when to relocate, what services to purchase, and what investments to make) within the overall context of self-regulated systems and subsystems. How these systems work and why they go awry would then serve as the basis for understanding the overall quality of urban life and how it might be improved. Only when these and similar topics are fully under investigation can the appropriate implications for public policy be understood.

Viewed in this broader context, one suspects that these policy implications will in fact call for a much more creative and innovative approach to policymaking. This is because previous policies are likely to be shown to have had negative consequences on the social control, economic, and demographic trends related to improving the quality of urban life. Such consequences include the disintegrative effects on neighborhoods of urban renewal, transportation, and other Federal capital improvement programs, as well as the consequences of Federal programs not aimed toward cities (e.g., Social Security, personal income tax regulations, and defense spending) that have nevertheless had negative impacts on the urban economy.[23] What will be most challenging is whether effective policies can be designed to promote public-private partnerships,[24] and to promote neighborhood revitalization.[25]

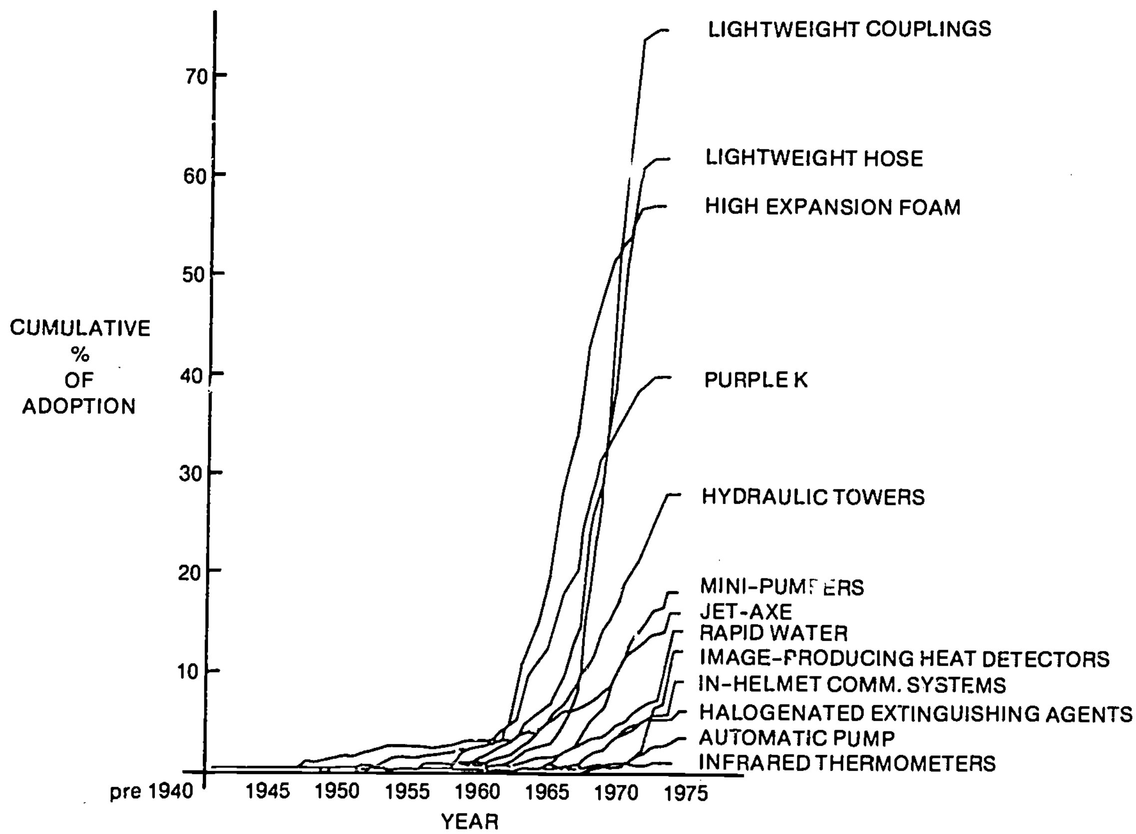
Only in this manner are governments at all levels (Federal, state, and local) likely to develop the creative options needed to suit future public priorities and the milieu produced, for instance, by passage of California's Proposition 13. Thus, it is unlikely that new policies that implicitly require increased government expenditures will be satisfactory. Instead of considering such traditional alternatives as increasing the availability of Community Development Block Grants, for instance, researchers will have to explore more fully the governmental role in regulating (explicitly and implicitly) the affairs of the city. Other options such as the Home Mortgage Disclosure Act of 1975 (Title III of P.L. 94-200), calling for new regulations rather than increased expenditures, might thus have to be developed in order to affect urban markets, whether they are housing, industrial, or service delivery markets.

URBAN TECHNOLOGICAL R&D

In addition to the preceding R&D topics, which assume the importance of the collective activities of private individuals and organizations in urban life, there are also important topics where state and local agency policies are directly involved. These especially involve the use of new technologies.[26] Thus, a second group of R&D topics worthy of continued investigation have to do with the factors associated with successful innovation. One constructive suggestion here is that research should attempt to explain successful experiences rather than continually focus on the barriers to innovation; more is learned about a process if it is studied under these conditions.

On an agency-wide basis, urban fire services have successfully implemented a wide range of new technologies. These are illustrated in Figure 1. which shows the diffusion rates for 13 different innovations (see Feller[9]).

FIGURE 1 Cumulative percentage of adoption of fire-fighting technologies by cities (% based on 551 respondents)



Source: Feller, Irwin, et al., Diffusion of Innovations in Municipal Governments, Institute for Research on Human Resources Pennsylvania State University, University Park, June 1976. Also Eckfield, Richard E., et al., Factors Involved in the Use of University Research by State and Local Governments. Menlo Park, California: SRI International, April 1978.

In general, these curves indicate the degree to which technological innovation is a relatively recent phenomenon, the bulk of the adoptions having occurred only in the past 15 years. For this reason, the research literature regarding the status of urban service innovation has lagged, thus hindering our understanding of why innovation occurs.

One criticism of many of these technologies has been that they tend to be less visible to the public and may not involve dramatic changes in service delivery. However, two illustrative innovations may serve to dispel these notions: the paramedic service (public health) and the use of breath-testing equipment (public safety). In both cases, the innovations have been adopted on a widespread basis and have led to service benefits that are clear to both administrators and the public (see Yin,[27] for more detailed descriptions of the life histories of these innovations).

In the first case, paramedic or mobile intensive care units (MICUs) provide emergency care away from a hospital facility. The unit consists of a specially equipped vehicle and a paramedic team that can administer a range of medical care upon instruction from a centrally located physician. The key technological element of this innovation is thus the communication capability of the MICU: the mobile staff can maintain voice contact with a physician, and physiological data about the patient's condition can be sent to the physician via radio telemetry. As such, the MICU goes much beyond either the traditional ambulance service (which provides transport service only) or the traditional first-aid provided by fire or police officers (which is limited to basic first-aid and is not based on immediate physiological measurements or advice from a physician).

The MICU has another distinctive feature—the use of specially trained paramedic personnel. These personnel receive extensive training (i.e., several hundred hours by medical doctors and nurses) and may be alternatively referred to as paramedics, advanced emergency medical technicians (advanced EMTs), or firemedics. These paramedics can perform a variety of medically sophisticated tasks: e.g., cardiac defibrillation, intravenous therapy, electrocardiographic (ECG) telemetry, drug administration, and endotracheal intubation. The paramedic personnel should not be confused with less advanced technicians—usually referred to as EMT-Ambulance or MET-As. The latter are generally ambulance personnel who have received an 81-hour course in basic life support, and they can perform such tasks as cardiopulmonary resuscitation, spinal injury management, splinting, hemorrhage control, and emergency childbirth. Nor should paramedics and MICUs be confused with mobile coronary care units, which were under experimentation several years ago and in which physicians and nurses, rather than paramedics, traveled as part of the unit.

There have been few specific evaluations of the service benefits from MICUs. This has been unfortunate, because the obvious impression at any MICU site is that service payoffs have occurred and can easily be determined. The payoffs can include: lives saved (paramedics describe their life-saving experiences in almost religiously fervent terms), amount and type of treatment provided at the scene, and greater satisfaction by clients and practitioner personnel. The absence of evaluations, however, has led to confusion because the organizational payoffs (as in regional systems) have been unclear and are confused with the service payoffs. However, a sure sign of the success of MICU operations has been their continued diffusion and implementation throughout the country, in urban, rural, and regional settings alike.

In the second case, local law enforcement agencies have recently implemented many improvements for testing and analyzing the amount of alcohol found in a person's breath. These improvements have been a response to both technological developments in breath testing—which can be conducted through processes involving gas chromatography, photometric colorimetry, or infrared photometry—as well as to the continued rise in traffic accidents attributable to persons driving while intoxicated (DWI).

The main breath-testing instruments that have been adopted include the Breathalyzer, the Alco-Limiter, the Gas Chromatograph, and the Alcohol Screening Device. The most common breath-analyzing instrument has been the Breathalyzer, a machine that uses photometric colorimetry and that infers the blood alcohol concentration (BAC) of a person on the basis of a sample of the person's breath. Before such breath-analyzing instruments became available, law enforcement agencies typically had to collect a sample of a person's blood or urine for analysis by a laboratory. In comparison to breath testing, these older procedures were highly complex, costly, and inefficient. Thus, breath testing offers the following advantages over older procedures:

- Less discomfort and inconvenience for the person being tested;
- No need for medical personnel to be present during testing;
- A shorter period of time away from patrol duty by the arresting officer;
- No delay of several days between the administration of the test and the reporting of the laboratory results; and
- No need to coordinate law enforcement procedures with those of a central laboratory. (The

older procedures had to be careful to establish a direct chain of evidence to assure that none of the samples had been exchanged or had accidently become contaminated.)

The preceding illustrations are only two of a larger number of innovations that have been successfully implemented in the past five to ten years by local governments. Further research is needed to identify other examples, to assess the benefits from these innovations more formally, and to examine the factors associated with these successful innovations.

FEDERAL MANAGEMENT OF SCIENCE AND TECHNOLOGY (S&T) PROGRAMS

The preceding section has outlined some of the substantive topics on which further R&D might be productively focused. A general suggestion has been that future NSF efforts might be targeted more toward the support of basic and applied R&D rather than capacity building efforts or program analysis. Beyond the question of the substance of the R&D agenda, however, is the equally important question of how Federal R&D should be managed in the future.[28] New policies need to be developed regarding Federal support for S&T oriented toward serving state and local needs, and two managerial concerns in particular ought to be addressed. The first has to do with the potential conflict between the way that R&D progresses—i.e., through the incremental cumulation of knowledge over a long period of time—and the fact that Federal agencies are often forced to demonstrate their achievements on a project-by-project basis. The second has to do with improving the technical quality of the research that is supported.

Reviewing R&D on an Aggregate Basis

Regarding the first, numerous observations of the conduct of science have shown that R&D progresses in a series of small steps rather than as a result of dramatic breakthroughs.[29] Single research projects must be seen as only producing *findings*, and not scientific *facts*; the latter only emerge after a series of replications and modifications by a host of other research projects or experiments.

Unfortunately, individual research projects have increasingly become the focus of attention in recent years. First, most applied R&D programs supported by the Federal Government require that each research project generate its own "utilization plan." This procedure was installed in NSF's applied programs, for instance, in May 1974. Second, attempts to improve the technology transfer process have often focused on the dissemination of project-based results. Thus, previous hearings on R&D utilization held by the U.S. House Committee on Science and Technology[30] were based on three perspectives, of which one was that new policies could inprove "Efforts within the Administration and the Executive agencies to coordinate R&D funding for dissemination and utilization of results from successful *projects*." (Emphasis added.) Third, there has been an increasing attempt to judge scientific progress on the basis of the design and outcome of single projects. Thus, R&D agencies have been forced to scrutinize individual projects and to avoid such dubious honors as the receipt of the "Golden Fleece" award.[31] Fourth, in terms of R&D utilization, an increasing complaint among users is that when they have attempted to make organizational changes on the basis of recommendations from a single study, they find that subsequently a new study has come out with contradictory recommendations.

The dilemma that has not been adequately confronted is that Federal R&D agencies generally operate on the basis of making individual project awards—which may be defined as a contract, grant, or other agreement whereby an R&D-funding agency supports a specific research project—and administrative accountability is a valid concern in dealing with each award. Such valid concerns include whether the funds were used in a fiscally responsible manner, and whether the outcomes of the project met the objectives stipulated for the project. Current proposals to improve the accountability process, including the implementation of the Federal Grant and Cooperative Agreement Act of 1977, P.L. 95-224, should therefore be strongly supported.

However, new policies and procedures are needed to assess the overall value of scientific progress and impact. These policies should not be based on the results of single research projects, nor should specific R&D utilization efforts be mounted for each individual project. Instead, some type of aggregate function needs to be performed, in which: progress from numerous individual projects is synthesized, a consistent set of scientific facts is accumulated, and these facts are made the basis for actual dissemination and utilization. The shape of such policies or procedures is unclear and should be explored in the future. The questions to be discussed include:

- How should the *aggregate* contribution of R&D efforts on a given topic be assessed, and who should conduct such assessments?
- How can these assessments be related to performance by specific programs?
- In what ways can the aggregation process—whereby results of individual projects are compared, replicated, and extended—be facilitated?

- What are the realistic criteria—e.g., sound fiscal management but not necessarily utilization impact—by which individual projects should be judged?

Improving the Quality of Urban R&D

Regarding the second managerial concern, recent pressure to produce immediate solutions to significant urban problems has in part led to shortcuts in the normal quality-control processes for conducting research. For instance, a major quality-control procedure in scientific research is the publication of results in refereed journals. The review and publication process, together with the time needed for other investigators to assimilate and react to published reports, usually occurs over a multi-year period. Although the process is time-consuming, it is one of the few ways in which scientific ideas may be sorted: some ideas never reach publication, others are published and ignored, and yet others are published and either refuted or extended. Unfortunately, this lengthy process has forced many agencies to develop alternative forms of communicating R&D results. Thus, a common requirement imposed on urban research projects today is the issuance of a final report, which describes an entire study and becomes the basis for policy recommendations. Such final reports may be seen as being available on a more timely basis; however, the reports are not usually subjected to the same rigorous review as journal articles or other academic publications. As a result, less quality control is imposed.

Publication in refereed journals is but one quality-control procedure used in scientific research. Other procedures normally used by R&D-funding agencies include peer review of initial proposals, the maintenance of a balance between solicited and unsolicited research, the use of advisory committees in the course of conducting research projects, and the employment of staff members who are technically competent in the relevant fields. None of these procedures works with great efficiency or effectiveness, and none is a substitute for the more casual indicator of high-quality research—i.e., the participation, whether direct or indirect, of individuals of acknowledged excellence in a research enterprise.

Recent reviews of various applied research programs have noted that such programs have often been deficient in their ability to support high-quality research—i.e., research that is based on the rigorous application of scientific procedures.[32] At the present time, the appropriate courses of action are unclear and might thus be the subject of future inquiry. Even though R&D results should be disseminated in a timely and compact manner, it is unclear, for instance, whether investigators in urban research projects should only be judged by their ability to produce final reports; it may be that requirements for such reports should be complemented by requirements for formal academic presentations, whether in refereed journals, professional meetings, or some other communications medium within the *scientific* community. Overall, the questions to be discussed might include the following:

- How can high-quality investigators be attracted to do applied research?
- What policies and procedures should be implemented by R&D funding agencies to improve the quality of research projects?
- What incentives and requirements should be made in the conduct of research projects to improve their quality?
- How can applied research projects and their results be better integrated within the scientific community—e.g., through the publication of results, discussion in scientific meetings, or other collegial activities?

Inquiry into these and related issues can result in improving the quality of research, which in turn is likely to improve the overall impact of R&D on state and local affairs.

NOTES

1. "Science and technology" is used throughout the text to cover the research programs administered by Federal mission agencies as well as those of the National Science Foundation.
2. Executive Office of the President, *Public Technology: A Tool for Solving National Problems*, Washington, D.C., 1972; and *The Science and Technology Message of the President*, Washington, D.C., 1979.
3. Yin, Robert K., et al., *Tinkering with the System: Technological Innovations in State and Local Services*, Lexington, Massachusetts: Lexington Books, 1977.
4. Public Affairs Counseling, *Factors Involved in the Transfer of Innovations*, report to the U.S. Department of Housing and Urban Development, Washington, D.C., January 1976.
5. Fundingsland, Osmund T., Assistant Director for Science and Technology, Procurement and Systems Acquisition Division, United States General Accounting Office, "Evaluation of Federally Sponsored Social R&D," panel presentation at the American Association for the Advancement of Science Annual Meeting, Denver, Colorado, February 24, 1977.
6. Yin, Robert K., "Contemporary Issues in Domestic Technology Transfer," *Domestic Technology Transfer: Issues and Options*, U.S. House of Representatives, Committee on Science and Technology, November 1978.
7. National Science Foundation, *Networking for Science and Technology in Local Governments*, Washington, D.C., March 1977. Also National Science Foundation, *Panel Report on the NSF Intergovernmental Program*, Washington, D.C., 1978.
8. Ibid., Panel Report 1978.

9. Feller, Irwin, et al., *Diffusion of Innovations in Municipal Governments*, Institute for Research on Human Resources, Pennsylvania State University, University Park, June 1976. Also Eckfield, Richard E., et al., *Factors Involved in the Use of University Research by State and Local Governments*, Menlo Park, California: SRI International, April 1978.

10. Lederle, John W. and Strauss, Dorothy E., Michigan Government Digest No. 4, Bureau of Government, University of Michigan, Ann Arbor, November 4, 1949. Also MacDonald, H. Malcolm, et al., *Outside Readings in American Government*, New York: Thomas Y. Crowell Co., 1952.

11. Brooks, Harvey, "The Problem of Research Priorities," *Daedalus*, Spring 1978.

12. Intergovernmental Research and Development Project, Office of Public Sector Programs, American Association for the Advancement of Science. *Report from a Workshop on Community and Economic Development*, "Neighborhood Development and Stability," Belmont, Maryland, 1978.

13. Ibid.

14. One must distinguish here between the use of a final report for *accountability* purposes and for technology transfer purposes. The suggestion is that successful program analysis efforts are often based on immediate implementation, without waiting for a final report. (The author saw this occur frequently in his experience at the New York City-Rand Institute.) The final report might be prepared several months later, however, in order to satisfy research sponsors or other accountability needs.

15. Brown, Richard M., "The American Vigilante Tradition," National Commission on the Causes and Prevention of Violence. *Violence in America, Volume I*, Washington, D.C., 1969.

16. Yin, Robert K. and Yates, Douglas, *Street-Level Governments: Assessing Decentralized and Urban Services*. Lexington, Massachusetts: Lexington Books, 1975.

17. Rubenstein, Jonathan, *City Police*, New York: Farrar, Straus, and Giroux, 1973.

18. Kakalik, James S. and Wildhorn, Sorrel, *Private Police in the United States: Findings and Recommendations, Vol. I*, Santa Monica, California: The Rand Corporation, R-869-DOJ, December 1971.

19. Yin, Robert K., et al., *Patrolling the Neighborhood Beat: Residents and Residential Security*, Santa Monica, California: The Rand Corporation, R-1912-DOJ, March 1976.

20. Bickman, Leonard, et al., *Citizen Crime Reporting Projects*, U.S. Department of Justice, Washington, D.C., April 1977.

21. Jacobs, Jane. *The Death and Life of Great American Cities*, New York: Random House, 1961

22. Goetze, Rolph *Building Neighborhood Confidence*, Cambridge, Massachusetts: Ballinger, 1976.

23. Vaughan, Roger J., *The Urban Impacts of Federal Policies: Economic Developments*, Santa Monica, California: The Rand Corporation, E-2028-KF/RC, June 1977.

24. "New Partnerships to Conserve America's Communities," The White House, Washington, D.C., March 27, 1978.

25. Yin, Robert K., "What a National Commission on Neighborhoods Could Do," *Journal of the American Real Estate and Urban Economics Association* 5: Fall 1977.

26. Bingham, Richard D., *The Adoption of Innovation by Local Governments*, Lexington, Massachusetts: Lexington Books, 1976. Also see: Crawford, Robert, "The Application of Science and Technology in Local Governments in the United States," *Studies in Comparative Local Government* 7:1-19, 1973; Eveland, J.D., Rogers, Everett M., and Klepper, Constance A., "The Innovation Process in Public Organizations," Department of Journalism, University of Michigan, Ann Arbor, August 1976; Kelly, Patrick and Kranzberg, Melvin, *Technological Innovation: A Critical Review of Current Knowledge*, Vol. 1, Georgia Institute of Technology, Atlanta, February 1975; Lambright, W. Henry and Flynn, Paul J., "Bureaucratic Politics and Technological Change in Local Government," *Journal of Urban Analysis* 4 (1): 93-118, 1977; Zaltman, Gerald, et al., *Innovations and Organizations*, New York: John Wiley and Sons, 1973.

27. Yin, Robert K., *Changing Urban Bureaucracies: How New Practices Become Routinized*, Lexington, Massachusetts: Lexington Books, 1979.

28. Baer, Walter S., et al., *Analysis of Federally Funded Demonstration Projects: Final Report*, Santa Monica, California: The Rand Corporation, R-1926-DOC, April 1976. Also see: Glennan, Thomas K., Jr., et al., *The Role of Demonstrations in Federal R&D Policy*, Santa Monica, California: The Rand Corporation, R-2288-OTA, May 1978; Newton, Robert D., "Contracting under Grants: The Need to Define the Federal Role," *Public Contract Law Journal* 9 (1): 35-44, 1977; Task Force on Criminal Justice R&D, *Criminal Justice R&D Standards and Goals*, Washington, D.C., 1976; Task Force on Science Applications, *Report to the Director of the National Science Foundation*, Washington, D.C., July 1977; Wright, Joseph Ward, "Delineating the Role of the Research Sponsor in the Research Utilization Process: A Case Study of NSF-RANN," Ph.D. dissertation, School of Public Administration, University of Southern California, Los Angeles, June 1978.

29. Rettig, Richard A. and Wirt, John G., "On Escaping the 'Moon Shot' Metaphor," *Policy Studies Journal* 5:168-171, 1976. Also see: Rich, Robert F., "Innovations/Diffusion Research, Public Policy, and Innovations," paper presented at Workshop on the Diffusion of Innovations, Northwestern University, Evanston, Illinois, November 1977; Rosenberg, Nathan, "The Diffusion of Technology: An Economic Historian's View," paper presented at Workshop on the Diffusion of Innovations, Northwestern University, Evanston, Illinois, November 1977; Yin, Robert K., *R&D Utilization by Local Services: Problems and Proposals for Further Research*, Santa Monica, California: The Rand Corporation, R-2020-DOJ, December 1976.

30. Subcommittee on Domestic and International Scientific Planning and Analysis, Committee on Science and Technology, *The Application of Science and Technology to Crime Control*, U.S. House of Representatives, 94th Congress, 1st Session, July 1975.

31. Shaffer, Leigh S., "The Golden Fleece: Anti-Intellectualism and Social Science," *American Psychologist* 32:814-823, 1977.

32. Committee on Research on Law Enforcement and Criminal Justice, *Understanding Crime*, National Academy of Sciences, Washington, D.C., 1977. Also see Committee on the Social Sciences in the National Science Foundation, *Social and Behavioral Science Programs in the National Science Foundation*, National Academy of Sciences, Washington, D.C., 1976.

3 The Impact of Technological Change on the Quality of Urban Life

*by John Paul Eberhard**

SUMMARY

Since World War II, the United States has made enormous investments in the development of science and technology. However, 90 percent of this investment has been directed at programs in the three areas of defense, space, and atomic energy. An opportunity exists to mine the wealth of research and development capability we have created to improve both the hardware and software systems of our metropolitan areas.

This report documents the key technological changes likely to emerge over the next five years and assesses their probable impact on the quality of urban life. Following is a short list of "primary" technological developments and bodies of knowledge identified as the likely driving forces for change over those years:

- Electronics—especially as incorporated in microprocessors and minicomputers.
- Visual communications—especially television as it expands the human experience.
- Space technology—especially as incorporated in satellite communications.
- Nuclear energy—especially as incorporated in electrical generating plants.
- Life sciences—especially as they reduce disease and increase life expectancy.
- Systems theory—especially as used in modeling complex urban interactions.

Each of these primary inventions will in turn be incorporated in other technological developments and bodies of knowledge available to improve the human condition. How well they serve to advance the quality of urban life will not be easy to measure because of the complex interrelationships of human institutions, the economic system, and our collective will. Ways must be found to allow technological competition in urban markets or the new inventions now under development will have great difficulty in displacing the older inventions—all of which depend on high rates of use of fossil fuel energies. A special program of experimental innovation incentives is advocated to set aside the large amount of "paralysis by analysis" presently surrounding policy decisions.

Modifications in housing technology, positive responses to the special needs of the elderly (an increasing sector of the population), the beneficial use of commercial satellite communications, major support for the development of alternative forms of energy (especially solar), and increased emphasis on public health and safety are all advocated as policy issues for Congress and the executive branch. By doing so, the civilization we have come to experience in the latter half of the 20th century

* Director, Architectural Research Associates, Inc., Washington, D.C.

can advance the quality of life for all who live and participate in our urban places. This would raise the level of concern for the problems and opportunities of our cities to one that deals with the welfare of the nation as a whole.

BACKGROUND

> No matter how completely technics relies upon the objective procedures of the sciences, it does not form an independent system, like the universe: it exists as an element in human culture and it promises well or ill as the social groups that exploit it promise well or ill.[1]

More than 45 years ago, Lewis Mumford pointed out the dilemma for policymaking in those areas of our society where public actions and private needs overlap. We have, in the United States, largely built our huge and successful edifice of science and technology around the three areas of national defense, space exploration, and atomic energy. For most of the years following World War II, 90 percent of our Federal expenditures for science and technology has been devoted to these three areas of research and development. Civilian issues, especially those that involve design, production, and use of the components of our urban areas, have been left to the private sector to develop. For some time now, we have been attempting to mine the scientific knowledge gained and the technologies created by our Federal R&D expenditures to improve the lot of those of us who live in the expanding urban areas of our country. In the next five years we will still be testing whether these earlier R&D expenditures can be exploited to improve the quality of urban life.[2]

The nature of our urban condition is determined by the complex interaction of the "hardware" systems (buildings, machinery, transportation, communications, energy) and the "software" systems we have devised for using this hardware to accomplish our social purposes (education, health care, recreation, life-safety, and employment). The questions of whether we enjoy the kind of life such interactions afford us, whether we receive satisfaction from the support these systems are able to provide, whether we gain personal benefits from the efforts we expend in such a context involve both a personal assessment and a societal one. The effective measures of these interactions in some cumulative (and largely elusive) way determines the "quality of life" we experience. Social indicators to account statistically for such things as the level of schooling achieved, number of infant deaths, and unemployment rates can provide useful indicators at the societal level, but they fall short of reflecting individual taste or style. Living in San Francisco is different qualitatively than living in New York City or Louisville, Kentucky, but for most people those differences, while important, remain unquantifiable.

Science and technology can have a wide range of impacts on either the hardware or software issues of our urban condition, or on their capacity to interact. Some urban issues can and will be improved over the next five years by wise applications of science and technology; other issues may or may not respond to such applications; and still others will likely be aggravated by a continuation of the present trends in spite of, or because of, technological developments. This document intends to discuss current trends in the urban condition, develop some of the issues to which science and technology can apply, and to set out related policy alternatives for the consideration of Congress and other policymaking bodies.

Science can be thought of as the cumulative body of knowledge that is divided into categories such as physics or biology. It is also a mode of practice for a large number of highly trained persons in our society who share a technique of problem solving—the scientific method. This method has limitations in terms of value systems (hence much discussion these days on the ethics of science) and in terms of application to human variability (thus the difficulty experienced in applying the social sciences to design decisions). In an attempt to avoid the pitfalls of prolonged debate on basic research versus applied research, this document will treat science as bodies of specialized knowledge we have available to draw on in addressing our urban condition during the next five years.

Technology is the cumulative set of tools and techniques we have developed to design, manufacture, and test the artifacts we create. The term "technology" is used in other ways, but for the purpose of this document it will be used to mean tools and techniques. A technology thus defined can be as simple as fastening together two boards with a nail and a hammer or as complex as launching and guiding a spacecraft to explore Saturn. Although not all technologies can be applied to the urban condition, in recent years we have made some progress in finding ways to use tools and techniques developed for defense programs or space exploration to design and build better urban hardware systems. In some cases we have even managed to find applications in the urban software systems as well.

In his book *Technics and Civilization*, Mumford reflected on a concept he called "primary inventions." That is, what are the inventions that are the means to other inventions or knowledge that is central to the expansion of knowledge? He suggested that between 1000 A.D. and 1750, the primary inventions or new ideas were mechanical clocks, the telescope, the scientific method, cheap paper, print, the printing press, and the magnetic compass.[3] Without the benefit of history, it is less easy to suggest a similar list for the primary inventions and new paradigms for knowledge that will be dominant in the next five years, but the following seem to fit his definition:

- Electronic technology—especially as incorporated in the computer.
- Visual communication—the special form of electronic communication that has broadcast images to expand the personal experience of people everywhere.
- Space technology—which makes it possible to get beyond the earth's field of gravity with devices such as communications satellites.
- Nuclear energy—especially as applied to electrical power generation, with all its inherent dangers.
- Life sciences—the body of knowledge that is dramatically changing our understanding of the human condition.
- Systems theory—an intellectual technology of major importance that allows us to model the relationships of complex interactions.

To attempt to confine the scope of this document within manageable limits, these six primary inventions will be the principal focus. In the analysis that follows and the issues discussed later, it would have been possible to include hundreds of ideas, problems, and opportunities. Time and context simply will not make that possible. However, it is the author's judgment that these six primary inventions will be the base for key scientific and technological areas of impact on the urban condition during the next five years. Consequently, they are the ones most important for policy consideration in urban areas, as related to science and technology.

ANALYSIS

Urban areas are neither easy to define—where does New York stop as an urban area as contrasted to a political jurisdiction?—nor to encompass by human thought. Trying to observe an urban area is like observing a living organism in which life cycle processes have been speeded up by time-lapse photography. It is virtually impossible for us to recognize all the changes that are going on simultaneously. We know that there are interrelationships that cause one set of changes to affect others, just as a decrease in the oxygen supply of humans will cause our lungs and heart to respond differently. We can often understand how these interrelationships work by collecting data on the performance of the separate elements, but even as we do the analysis we also know from experience that the set of relationships is changing. For example, to suggest that a modification in the traffic patterns of a city would improve the economic aspects of downtown merchants is probably logical; however, not enough of the variables are likely to remain constant to let us measure the results of our policy decision.[4]

Still another difficulty in bounding the urban condition is the fact that every metropolitan area in the United States presently consists of many urban nodes, even though there is usually a center city that is commonly referred to as "downtown." Recognition of these urban nodes makes it less clear where to draw the line between urban and suburban areas. It is not uncommon to find clusters of working places (office buildings and light industry) surrounded by shopping malls, hotels, theaters, churches, and even medical clinics located just on the edge of the historical city limits of the central city. Such urban nodes have all the physical facilities usually associated with urban as contrasted to suburban areas. Therefore, in looking at how science and technology influence such factors as population distribution, it is less useful to speak of "urban nodes," "residential neighborhoods," or "central cities" and "suburbs" than to speak of cumulative urban areas as "metropolitan" and to distinguish them from "rural" areas simply on the basis of population densities.

Trends

An analysis of some key trends can give us clues to the metropolitan condition that is likely to emerge in the next five years.

AGE OF POPULATION

There is a decided shift coming in the distribution of ages within our population. In the next five years, the relative percentage of the population between 5 and 20 years old will continue to decline; the population between 40 and 60 will continue to decline in absolute terms (the result of nonbirths during the Depression years); and, most dramatically, the population over 75 will increase rapidly. The increasing portion of the population who are beyond 75 is one of the benefits of modern medical technology, but it brings with it the responsibility to shape our physical environment to make a better fit between the sometimes limited physical abilities of the aged and the kinds of buildings and metropolitan areas where they will live. Sufficient pressure is mounting in our society to recognize this need, so that in the next five years we should see more attention paid to the support requirements of the manmade environment for the aged, just as we have begun during the past five years in providing better access for the physically handicapped.

INCOME

In spite of the inflation that nibbles at the edges of progress, there does appear to be a more equitable distribution of income, which causes metropolitan areas to have a larger number of middle class residents. Fewer people are involved with menial work as technological improvements continue to provide mechanical and electrical equipment to do routine jobs, and there are more and more people with a high school education or better. The market for white collar workers is both a cause and effect of this increase in education. The service sector of the economy is expected to continue to pace the creation of new jobs, and new jobs in turn often mean new buildings and new communities.

HOUSING[5]

It becomes more likely with each passing year that the housing industry is on the brink of a technological revolution. A rapid change in housing production would be characterized by a shift to more capital-intensive processes and away from the labor-intensive processes. For more than 25 years, financial and real estate institutions have been able to produce innovations within their sector of the industry that enable a large number of people to buy houses, even at inflated prices. Until (or unless) there is some genuine market pressure on the technology of designing and building houses, little is likely to change.

The cost of land and the selling prices of houses are inflating at record levels. (In 1978 the average house price increased by 15.7 percent.) This increase has had little relationship to the cost of construction but seems driven primarily by the general real estate market, which has become an attractive investment for anyone able to put together sufficient financial resources to make the downpayment and keep up the monthly costs. In this environment, there is little incentive to develop technological changes. It seems reasonable to predict that in the next five years the cost of houses will not be affected by such changes, since the real estate market will continue to determine the price of houses.

OTHER BUILDINGS

Buildings other than houses are likely to be changed in the future as the result of two different, but related, influences. One is the rapidly increasing cost of fossil fuel energy, and the other is the interest in renovating and reusing older buildings. (Energy will be discussed further under "Issues.") Most techniques of construction and most building products were developed for new buildings. As the interest in renovating and reconstruction continues to expand, techniques are beginning to emerge that have been specifically developed for making changes in existing buildings.

TRANSPORTATION[6]

Just as energy issues will impact the design of buildings, so will the growing shortage of oil impact transportation technologies. How serious the oil shortage will become in the next five years is subject to widely divergent points of view. It seems clear, however, that the automobile will not be as dominant a transportation system in the future. Five years is not a very long time for new developments to take place—the average time from concept to marketplace distribution of new products is about seven years—and genuine public concern with the impending oil shortages is only a few months old. However, if gasoline shortages continue to plague automobile owners, they may well turn to alternative transportation modes at a rapid rate, including the substitution of communications for transportation.

COMMUNICATIONS AND INFORMATION PROCESSING[7]

The communication technologies of metropolitan areas are likely to experience many substantial and dramatic changes in the next five years. In 1982, the first major domestic communications satellite (owned and operated by Satellite Business Systems) will be launched. This will provide substantially new means of electronic communications and information processing, as discussed further below.

"Chip" technology (miniature electronic circuits now being produced at dramatically reduced prices and greatly increased capacity) is also expected to explode in many directions. The inclusion of microprocessors in everything from the home cooking center to traffic control devices will make it possible to invest many pieces of equipment with a form of intelligence not previously possible. Minicomputers are expected to find their way into every nook of the business world, as well as into the home via games and home accounting systems. Each day finds another new product announcement from some sector of the electronics industry. Parallel developments in communications technology make it possible to imagine large urban complexes within the next five years with much more sophisticated management and control capabilities than we now have.

If these changes in the hardware components of metropolitan areas are combined with the population and income trends indicated earlier, it seems clear that the software systems will also be experiencing associated changes. These changes are likely to be seen in education, health care, recreation, work, and local government. Other areas will likely be affected as well, but these five seem to be where the most visible changes will occur.

EDUCATION

Education at the elementary and secondary level in metropolitan areas is one area of public service that has the potential to benefit from the list of new technologies mentioned earlier. However, during the next five years there may be countervailing conditions that will prohibit these benefits from being realized. The increasing numbers of people who are well beyond the child-rearing age means that their votes are much less likely to support increases in local school budgets. The general decline in school-age populations means that there are surplus teachers and surplus school buildings, especially in older, established neighborhoods in the cities and suburbs. The general public attitude about holding the line on local taxes (the Proposition 13 syndrome) is not likely to diminish in the next five years. All of this suggests that primary and secondary educators are not likely to be able to fund many technical advances in their areas within this time period. However, in higher education we may experience a different set of conditions. As more and more mature persons seek out university programs whether for required knowledge or general learning, it seems likely that higher education will turn to mass communications as a technological aid. Scientific knowledge and technological ability are the highly marketable commodities of higher education institutions.

HEALTH CARE AND SAFETY

Health care seems likely to continue to make rapid improvements, as does the related area of personal safety. For example, communications satellites connected to computer-based networks will provide advances to health care data systems. National health programs are on the way that provide protection for all citizens against the financial catastrophes that all too often have accompanied the treatment of major illnesses. One of the reasons for the high cost of such medical care is the high level of medical technology generated during the past decade. With some reasonable assurance that even higher costs can be managed through a national health insurance program, it is likely that the next five years will see a very large number of additional technologies emerge for health maintenance as well as health restoration. This would seem to be particularly likely for the increasing portion of the population over 75 that needs many new prosthetic devices but presently lacks the collective funds to represent a market for development of such devices by the private sector.

Some improvements in personal safety will relate to preventing automobile accidents (or, as Dr. William Haddon, president of the Insurance Institute for Highway Safety, prefers to call it, "the avoidance of unwanted transfers of energy between moving vehicles and humans.").[8] The improvements in vehicle design that will be required by Federal legislation already enacted will likely produce substantial reductions in injury and death, and the avoidance of automobile wrecks will have a major impact on total accident statistics. Dr. Haddon says that in 1978 the single largest cause of death for people between 6 months and 37 years old was accidents of various kinds.[9]

Dr. Philip Handler, president of the National Academy of Sciences, has suggested that "there are a host of social institutions that did not exist when most of us were young, designed to protect individuals as well as organizations from catastrophes: social security, unemployment compensation, benefits for the disabled, Federal Deposit Insurance Corporation, Federal disaster assistance, and so on. It is taken for granted that there should be regulations in support of the health, safety and well-being of individuals (the FDA, OSHA, EEOC, laws against fraud, diverse laws concerning licensing, etc.)"[10] The results in recent years have been reduced risks to life, as compared to the past.

RECREATION

The decreasing working hours and increased incomes available to most urban dwellers have engendered a large collection of new leisure time activities. Dr. Handler points out that the "number of books published annually in the United States increased from 11,000 in 1950 to 40,000 in 1974. The number of symphony orchestras increased five-fold in the two decades following 1955. Statistics concerning sporting, picnicking and fishing equipment, outboard motors, boats, foreign travel, theater going, etc., are staggering as compared with any prewar experience."[11] There could be a wide range of possibilities for technological changes to continue moving into the leisure markets, but again there are some counteracting trends. The cost and availability of gasoline is already changing the pattern of vacation planning, and the recreational vehicle market is apparently nearing bankruptcy. Electronic technologies are more likely to find their major markets in home entertainment units, thus shifting theater and concertizing to studio activities to be recorded for playback in home centers, rather than live performances. All of this suggests that leisure time activities will doubtless benefit from the changing technologies, but traditional recreational businesses may suffer.

WORK

More than 20 years ago, some observers began to suggest that our technological and scientific improvements to industry and commerce were going to make productive work as we had known it largely disappear. Predictions were made that only 10 percent of the population would be required to work to produce all the goods we needed to maintain our standard of living. Leaving aside the construction industry, this seems to be the case in most sectors of manufacturing. What has happened is that there has been a subsequent dramatic increase in the number of white collar service workers. Science and technology related to the primary inventions being discussed in this document would seem likely to maintain the pace of job creation in the service sector of the economy—if we can satisfactorily manage our fossil fuel energy situation during the next five years.

LOCAL GOVERNMENT

Local governments make up another sector of our urban metropolitan area that might benefit from the new technologies. Some of the smaller, more affluent towns and cities (especially those in the "sunbelt" area) will probably be able to gain these benefits and thus demonstrate their potential. But the Proposition 13 public attitude is likely to be persistent enough during the next five years to prevent local governments from making any substantive investments in new technologies. Federal support via Federal revenue sharing could help stimulate developments by local governments, if the natural tendency to use revenue-sharing funds to rescue the local budget can be avoided. Many local government assistance institutions are in place to help make technology advances happen.

ISSUES

A subject as broad and diverse as the future of urban areas can be counted on to provide hundreds of issues of concern to citizens, policymakers, and observers of scientific and technological development. This document will limit itself to issues that emerge from the prospect of applying the six primary areas of inventions and the trends discussed earlier.

Technology Breakthroughs

Probably the clearest technological breakthrough now visible is the one being precipitated by "chip" technology—a manufactured electronic component about an eighth of an inch square that contains all of the complex circuity needed to perform an enormous variety of mathematical, logic, and data manipulation functions. The chip is being incorporated in relatively small *microprocessors* and fairly inexpensive *minicomputers*. These are new products well on their way through the development cycle that will likely be as pervasive in our lives as television and the telephone.

A second and related technology well on its way to being a commercial reality is the communications satellite. Satellite Business Systems (SBS), a joint venture of IBM, COMSAT, and Aetna Insurance Company, expects to be selling $500 million in communications, broadband capability to U.S. Government and industry as early as 1982. They expect to use this capacity for computer data transmission, financial planning and recordkeeping, environmental controls, security systems, educational programs, and entertainment (including electronic puzzles, a network for playing chess or bridge, electronic lotteries, etc.). SBS expects that smaller towns and suburban communities will be able to compete with larger central cities by having satellite links that will make possible the following:

- Continuing education programs for professionals that would not be available locally—e.g., special programs for physicians about new forms of treatment.
- Teleconsultation between professionals, including graphic materials needed for such consultations—e.g., a local firm of engineers consults an expert on the earthquake design provisions of a new bridge.
- Expansion of information services in many ways, with the possibility of illustrations as well as words being communicated—e.g., information for the city engineer on a new hydraulic pump being developed in a laboratory 2,000 miles away.
- Sharing of expensive regional resources such as large computer data systems or high-technology medical diagnostic equipment.
- Group interactions via teleconferencing that avoid time-consuming and expensive travel—e.g., local elementary school teachers can participate in a national meeting 3,000 miles away.[12]

The combination of microcomputers and microprocessors linked together into a national network via communication satellites suggests many applications in managing a metropolitan area. Emergency plans for dealing with natural disasters; critical data on potential epidemics linked to the Center for Disease Control in Atlanta, Ga.; security systems that make it possible for smaller, more mobile police forces to cover a larger area; and engineering and architectural networks tied to such national data resources as the National Climate Center in Asheville, N.C., will be possible as a result of these technological changes.

When satellite technology is combined with local networks of cable television, the prospect for educational programs in the home becomes a likely development. The prospect for service industry employees to upgrade their knowledge and skills will be greatly enhanced by such technological developments.

Energy Concerns[13]

Nuclear energy as a source of generating electrical energy seems certain to increase in scope and scale of application, in spite of the controversy that surrounds its inherent dangers. As the shortages of petroleum and natural gas grow more acute over the next five years, and subsequently as the price of fossil fuels continues to climb, two basic directions in energy policy seem likely: increases in our supply of electrical energy by applications of nuclear power and increases in our attention to nontraditional alternatives (including conservation).

The energy crisis has already begun to change people's thinking about alternative energy sources and the possibility of conserving the use of fossil fuels. If the availability or cost of such fuels accelerates even more rapidly, there will be more pervasive changes in how metropolitan areas are constructed in the future and how they are modified now to reduce our dependence on energy from fossil fuels. The future possibilities in this area may seem to portend serious difficulties, and it is indeed possible that the consequences could be detrimental to our economy. But there are signs that the results could actually be beneficial. For example:

- Active solar energy systems are technically feasible and effective, but are not effectively marketed at the moment. Additional research funds from the U.S. Department of Energy could increase knowledge and the application of current technologies.

- Passive solar energy costs the consumer almost nothing, since it is based on building design principles, but it is dependent on widespread education of design professionals before it becomes common practice.

- Forms of biomass, "gasohol," waste wood conversion, and methane all show promise and are prime candidates for increased public and private investment.

- Control devices that can be used in large buildings or at the urban scale to obtain a better fit between the needs of users, climate conditions, and systems responses are still largely in the development stage.

Each of these new energy ideas—and many yet to be invented—can create new industries, new jobs, and new inputs to the economy. More importantly, such technological developments help us avoid costly balance-of-payment problems with the OPEC nations, and they help reduce the eventual dependence we will have on nuclear generating plants.[14]

The Life Sciences and Technology

The union of life sciences and systems theory may seem to be a strange confluence of ideas, but in many of the basic ingredients of metropolitan life it is just such a confluence that seems to offer the impetus for change over the next five years. It has already been suggested that the life sciences are, and will continue to have, a dramatic impact on the composition of our population. With birth control measures reducing new additions and disease control increasing the life expectancy of those now living, the average age of the population continues to increase. The shift of the work force (those between the ages of 25 and 60) to service sector jobs leaves a partial vacuum in the manual labor and menial job categories. However, these jobs seem to be attractive to new immigrants (both legal and illegal) and thus make it possible to continue to operate some of our more labor intensive industries, such as construction, maintenance, and certain forms of agriculture.

Under these conditions, if we are to make effective links between science, technology, and metropolitan conditions, some breaks with traditional methods seem necessary. Systems analysis based on performance requirements of the software systems (education, health care, recreation, etc.) could be a powerful tool in setting directions for these links. Unfortunately, the history of urban development over the last 30 or 40 years is based on creating markets for traditional inventions rather than creating the conditions for innovation (see policy proposals, below).

Housing Issues

Urban decay seems to breed more urban decay, like a giant cancer. If we are to restore decayed inner city neighborhoods to some semblance of a "decent place to live and a fit environment" (the goal of major U.S. housing legislation), we are most likely going to have to use the powers of government control of land use. It would, however, be possible to leave ownership of land in private hands if private greed can be contained within the public interest. Government spending (not loan guarantees) will be required to restore damaged areas while still allowing low income people (especially the elderly) to live within the inner city. If all the costs and benefits could be weighed without prejudice, it would probably be possible for citizens to support such action. Systems analysis is a tool that could be used for this purpose.

The task of organizing technological change in the housing field seems to be related to the difficulty of organizing a market that would be large enough (and capable of being sustained over a long period of time) to attract the investment capital needed to make a substantial difference. The technological revolution in agriculture that achieved dramatic increases in productivity, by substituting capital investment in farm machinery and fertilizers for human labor, has been an elusive dream of some people in the housing industry for decades. There is little reason to expect this situation to change in the next five years. Unfortunately, even if some technological miracle made it possible to build houses in which the walls (or the floors or roof or any other single component) cost 80 percent less than at present, the reduction in price to the home buyer would be scarcely visible. Some other component of cost—including profits to the host of lawyers and sales agents who get their share of the selling price—will expand to absorb the difference between costs and market prices. Consequently, housing increases in metropolitan areas will continue to fall into three major categories: the new high-rise housing "fortresses" of the well-to-do; the restored houses of the affluent new professionals; and the generally shabby housing of the low-income residents.

POLICY PROPOSALS AND PROGRAM SUGGESTIONS

Government programs, the result of both explicit and implicit policies, act in several ways to harness the knowledge gained from scientific research, or the tools and techniques that emerge from technological development:

- Some programs are intended to rescue urban areas that have fallen into decay or otherwise gotten in trouble. For example, the U.S. Small Business Administration will make loans to businesses in a city devastated by a flood or other major disasters.

- Some programs are meant to stimulate some sector of the urban society to improve economic conditions. Housing programs, urban rehabilitation efforts, and activities of the U.S. Economic Development Administration are examples.

- Some programs are designed to stimulate innovation and the creation of new ideas where there is insufficient economic incentive for the private sector to do so. Recent grants to stimulate the use of solar energy are in this category.

- Some programs are designed to bring the beneficial results of science and/or technology to a sector of society that lacks the visibility to stimulate the private industry market to make such benefits available. Making buildings accessible to the handicapped or providing special transportation for the elderly are examples.

Each of these ways can be appropriately used, as long as the policymakers are clear on their goals. Experimental housing programs (such as the Operation Breakthrough project in HUD in the late 1960s) with rhetoric that sounds as if they are meant to stimulate new ideas should not be subverted to become an indirect form of housing subsidy. Nor should the reverse be attempted.

Suggestions for policy initiatives are given with the recognition that programs are more likely to accomplish policy objectives when Congress is clear in its intention and specific in its program plan. Whenever possible, it also seems sensible to build on institutional capabilities and local program plans, rather than to start an entirely new Federal effort. This seems especially valid with respect to the large number of urban innovation efforts already begun but not often funded adequately.

Following are four program suggestions that illustrate each of the four ways to use policy directions based on effectively applying science and technology in metropolitan areas:

- To help smaller metropolitan areas adequately prepare for potential major disasters, mitigate the consequences of such events when they do occur, and provide for disaster relief after the event, an emergency network could be established via the proposed communications satellite. The National Hurricane Center in Miami, Fla., and other climate-related services have emergency networks already established. The former Office of Emergency Preparedness, now a part of the Federal Emergency Management Agency, probably has plans along this line in development, but Congress will want to encourage a comprehensive coverage (as contrasted to a specialized service tied primarily to climate-related events) ranging from potential aviation disasters to earthquakes. The new U.S. Fire Administration, for example, should be part of such a network to help in the event of a catastrophic fire in a smaller metropolitan area lacking the local resources to manage the consequences.

- While housing programs have historically been used to stimulate the economy, policy directions from Congress have normally resulted in loan guarantees being advanced by the U.S. Department of Housing and Urban Development. Even the urban renewal programs of the early 1960s were managed by Federal guarantees of local banking funds used for investments in land purchases by local government authorities. No technological innovation has resulted from such programs, only the stimulation of existing practices.

- If Federal funds were combined directly with local government initiatives of land reform in order to aggregate very large areas in metropolitan centers, free them from private speculation, remove historic streets and alleys, and begin a new "city within the city," then major technological innovations could be encouraged. A few years ago, Detroit undertook, in a seven-block-square area of its downtown, a program that has resulted in the Renaissance Center. The transportation, housing, public services, and energy management innovations made possible by such large-scale, three-dimensional development can be significant departures from existing hardware concepts limited by traditional land use patterns and development schemes.

- To change the economic incentives for technological innovation in housing, the Federal Government could launch a major housing program with the following attributes: If the land were publicly acquired within metropolitan centers by legislation, a housing program could then follow. The housing would be "packaged" in large enough volumes that there would be sufficient incentive for private companies to make the capital investments required to change substantially the technological base for construction. The housing would not be sold through the private real estate market but would be leased through local authorities that also had the responsibility to organize and provide the urban support services for those who leased the housing. As a condition of obtaining contracts to develop the housing, all participating companies would be required to provide a well-developed training program for the unskilled who live in the area under development, thus helping the local labor force make the transition to the new technology.

- The U.S. Administration on the Aging could be given broad financial support and authority to use the services and programs of other Federal agencies, to support state and local governments, and to provide incentives to the nonprofit sector to mount a series of programs for making urban areas and urban housing more responsive to the special needs of the aging. The alternatives of having the aged assemble in specialized communities or urban "ghettos of the old" would not appear to add to the quality of life for any of us.

Each of these suggestions has a large number of policy choices that could become alternatives. There is not sufficient scope within this paper to discuss all of the existing programs at Federal and local government levels that might relate to one or more of these suggestions. Because it is possible to evaluate large numbers of alternatives, because there are persons whose jobs call for coordination of policy initiatives, and because the efforts to measure economic consequences and environmental impacts are well staffed, we are too often stymied by a form of "paralysis by analysis." While it is true that such analysis can tend to prevent the launching of ill-conceived or redundant programs and thus do serve a public purpose, we need some relief valves that allow a few innovative programs to be tried in a climate that does not demand complete information as a condition for getting started. An "experimental innovation incentives" program might be tried again through the National Science Foundation or U.S. Department of Commerce. If such a program had our metropolitan areas as targets, we might then find ways to stimulate truly innovative concepts from the available wealth of scientific knowledge and technological capability.

REFERENCES

1. Mumford, Lewis. *Technics and Civilization*. New York: Harcourt, Brace & World, Inc., 1934, p. 6.

2. "Technology for the City." *International Science and Technology*, No. 57, September, 1966.

3. Mumford, op. cit.

4. U.S. Department of Transportation. Urban Mass Transportation Administration. "People Movement for Downtown Improvement." January 1977.

5. Center for Community Change. "Opportunities for Abuse: Private Profits, Public Losses and the Mortgage Banking Industry." October 1977. See also Construction Sciences Research Foundation. See also Fredrickson, Donald S.

6. Congress of the United States, Office of Technology Assessment. "Technology Assessment of Changes in the Future Use and Characteristics of the Automobile Transportation System." 1979.

7. Poppel, Harvey L. "The Information Revolution: Winners and Losers." Reprinted from *Harvard Business Review*, January-February, 1978. See also *Business and Society Review*, Winter 1978–79.

8. Insurance Institute for Highway Safety. "The Highway Loss Reduction Status Report." Vol. 14, No. 7, April 30, 1979.

9. Fredrickson, Donald S., "Low-Level Ionizing Radiation." Statement before the Subcommittee on Energy Research and Production and the Subcommittee on Natural Resources and Environment Committee on Science and Technology, U.S. House of Representatives, June 13, 1979. Haddon William, Jr., M.D. and Baker, Susan P., M.P.H. "Injury Control" Insurance Institute for Highway Safety. March 1978.

10. Handler, Philip. *Science, Technology, and Social Achievement*. Unpublished paper delivered at the Edison Centennial Symposium, San Francisco, Calif., April 1979. p. 7.

11. Hollomon, J. Herbert and members of the Center for Policy Alternatives, Massachusetts Institute of Technology. "Government and the Innovation Process." *Technology Review*, May 1979.

12. Brown, James H. "Advanced Communications and Rural Communities." Satellite Business Systems, 1979.

13. "Comprehensive Community Planning for Energy Management and Conservation." Executive summary prepared for U.S. Department of Energy. December 1977. See also Fredrickson, Donald S.

14. Zuckerman, Harriet and Miller, Roberta Balstad. "Social Indicators and Science Indicators."