ED 163 979 IR 006 657

AUTHOR	Bond, Nicholas A.; And Others
TITLE	Studies of Verbal Problem Solving: II. Prediction of Performance from Sentence-Processing Scores. Technical Report No. 87.
INSTITUTION	University of Southern California, Los Angeles. Dept. of Psychology.
SPONS AGENCY	Advanced Research Projects Agency (DOD), Washington, D.C.; Office of Naval Research, Washington, D.C. Psychological Sciences Div.
PUB DATE	Jun 78
CONTRACT	N00014-75-C-0838
NOTE	57p.
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	*Cognitive Processes; Complexity Level; Deductive Methods; Higher Education; *Logic; *Logical Thinking; *Predictive Validity; *Problem Solving; Sentences; Structural Analysis

ABSTRACT

This study explores the extent to which scores on four separate complex reasoning solution processes could predict performance on difficult problems. Definitions are provided for the four processes--intra-sentence processing, inter-sentence processing, ordering, and collecting--and previous work done in the field is outlined. The procedures used in obtaining scores for the four processes and the results of the initial experiment are discussed. The description of a training demonstration, designed to teach the skills needed in complex problem solving, includes the results of this exercise. A discussion of the success of the training procedure, a list of references, and typical sentence-inference sheets are also provided. (RAO)

DEPARTMENT OF PSYCHOLOGY

UNIVERSITY OF SOUTHERN CALIFORNIA

BEHAVIORAL TECHNOLOGY LABORATORIES

Technical Report No. 87

STUDIES OF VERBAL PROBLEM SOLVING:

II. PREDICTION OF PERFORMANCE FROM SENTENCE-PROCESSING SCORES

June 1978

Nicholas A. Bond, Donald McGregor, Kathy Schmidt,
Mary Lattimore, and Joseph W. Rigney

Sponsored by

Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

and

Advanced Research Projects Agency
Under Contract No. N00014-75-C-0838

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Office of Naval Research, the Advanced Research Projects Agency, or the U. S. Government.

Approved for public release: Distribution unlimited.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 87	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) STUDIES OF VERBAL PROBLEM SOLVING: II. PREDICTION OF PERFORMANCE FROM SENTENCE-PROCESSING SCORES		5. TYPE OF REPORT & PERIOD COVERED Technical report 1st April - 30 Sept. 1978 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Nicholas A. Bond, Donald McGregor, Kathy Schmidt Mary Latimore, and Joseph W. Rigney		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-838
9. PERFORMING ORGANIZATION NAME AND ADDRESS Behavioral Technology Laboratories University of Southern California Los Angeles, California 90007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element: 61153N Project: RR042-06 Work Unit: 154-355
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, VA 22217		12. REPORT DATE June 1978 13. NUMBER OF PAGES 47 + vi
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In complex reasoning problems of "the who-done-it" type, four distinct solution processes were identified:

(1) intra-sentence or word-into-symbol processing, where the solver converts the verbal information into strict logical relations;

(2) inter-sentence processing, where the subject has to combine the logic from two or more sentences in order to obtain new inferences;

(3) ordering of problem variables into some rank or numerical ordering scheme;
(4) collecting the logical relations into a reliable format that will reduce memory load and facilitate the "where-to-look-next" Decision.

This study explored the extent to which separate scores on these processes could predict performance on difficult problems.

Scores on the sentence-logic items correlated well(r = .68, N = 34) with number of reasoning problems solved, as did the ordering scores (r = .75). These scores, then, presumably are "closer" to the actual performance than are verbal scores such as McGraw-Hill Reading Rates (r = .40 to .50). Individual timing of inference responses showed that subjects often had long pauses during inter-sentence processing, whereas intra-sentence responding was relatively fast and regular. The inter-sentence portions of the performance appeared to be key discriminators between success and failure.

A small training experiment was carried out with seven new subjects, who were matched on reading scores with the previous group. These subjects were given six hours of intensive, individual practice on the four processes; a standard matrix format was used, and five rules and heuristics were taught which were supposed to facilitate inter-sentence reasoning. The trained people did show improved sentence-logic scores (median about 40% over the comparison group); and if a large reasoning problem contained only strightforward sentences, then the training was very effective. In fact, all seven subjects solved correctly a 4-dimension, 5-variable, negative-disjunction problem within a few minutes. For those problems which hinged upon appreciation of verbal subtleties, though, the training did not help at all.

The investigation supports the idea of rapidly teaching some "logical tricks" in higher-order cognitive operations; but the special training only works if the problem material is clean and unambiguous. One obvious extension of the study is to see if the same increase in performance can be produced in a practical-reasoning domain such as troubleshooting of digital devices; another extension is to look more closely at the verbal subtleties which so effectively prevent solution of some large problem.

SUMMARY

In complex reasoning problems of "the who-done-it" type, four distinct solution processes were identified:

(1) intra-sentence or word-into-symbol processing, where the solver converts the verbal information into strict logical relations;

(2) inter-sentence processing, where the subject has to combine the logic from two or more sentences in order to obtain new inferences;

(3) ordering of problem variables into some rank or numerical ordering scheme;

(4) collecting the logical relations into a reliable format that will reduce the memory load and facilitate the "where-to-look-next" decision.

This study explored the extent to which separate scores on these processes could predict performance on difficult problems.

Scores on the sentence-logic items correlated well (r=.68, N=34) with number of reasoning problems solved, as did the ordering score (r=.75). These scores, then, presumably are "closer" to the actual performance than are verbal scores such as McGraw-Hill Reading Rates (r=.40 to .50). Individual timing of inference responses showed that subjects often had long pauses during inter-sentence processing, whereas intra-sentence responding was relatively fast and regular. The inter-sentence portions of the performance appeared to be key discriminators between success and failure.

A small training experiment was carried out with seven new subjects, who were matched on reading scores with the previous group. These subjects were given six hours of intensive, individual practice on the four processes;

a standard matrix format was used, and five rules and heuristics were taught which were supposed to facilitate inter-sentence reasoning. The trained people did show improved sentence-logic scores (median about 40% over the comparison group); and if a large reasoning problem contained only straightforward sentences, then the training was very effective. In fact, all seven subjects solved correctly a 4-dimension, 5-variable negative-disjunction problem within a few minutes. For those problems which hinged upon appreciation of verbal subtleties, though, the training did not help at all.

The investigation supports the idea of rapidly teaching some "logical tricks" in higher-order cognitive operations; but the special training only works if the problem material is clean and unambiguous. One obvious extension of the study is to see if the same increase in performance can be produced in a practical-reasoning domain such as troubleshooting of digital devices; another extension is to look more closely at the verbal subtleties which so effectively prevent solution of some large problem.

ACKNOWLEDGEMENTS

This research was sponsored by the Office of Naval Research and the Defense Advanced Research Projects Agency under ONR Contract N00014-75-C-0838. The support and encouragement of Dr. Marshall Farr and Dr. Henry Halff, Personnel and Training Research Programs, Office of Naval Research, is gratefully acknowledged.

We also wish to thank Lloyd Lucien for setting up and calibrating the discrimination timing apparatus.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

Here is a logic problem, taken from a publication which specializes in word puzzles:

> Do you know the eight men who formed the president's cabinet in 1895? In case you don't, the following clues should enable you to find their names (Morton was one), their home states, and the posts they held, including Secretary of the Navy.
>
> 1. Four of the men served the full term, 1893 to 1897, in the same posts; the other four--Smith, the Postmaster General, the Secretary of State, and the Attorney General--did not.
> 2. Lamont was not the Secretary of Agriculture, nor was he from Nebraska.
> 3. Herbert was not from Nebraska or New York.
> 4. The Secretary of the Treasury was from Kentucky.
> 5. Harmon was not from West Virginia or Massachusetts.
> 6. The Georgian, who was the Secretary of the Interior, left his post in 1896.
> 7. The men from West Virginia and Kentucky were Carlisle and the Postmaster General, not necessarily respectively.
> 8. The man from Alabama was not the Secretary of War of Agriculture.
> 9. Neither Wilson nor the man from Massachusetts served from 1893 to 1897 in the same posts.
> 10. The Attorney General was from Ohio.
> 11. In 1895, Olney had only recently assumed the post he held, when Harmon entered the cabinet to take Olney's original post.

As you scan the problem, the language seems simple enough; except for a little awkwardness of phrasing in Sentences 7 and 11, the words are clear and the meaning is definite. But this problem is difficult. A typical college student, who is not familiar with this sort of thing, will not finish it in less than half an hour (in one night class of 22 people, not a single person got it in that time). Graduate students and faculty will probably need fifteen to twenty minutes; and without aid, the problem will be forever insoluble for a large fraction of American adults. Some of the difficulties are quite evident: the sheer amount of data presented, the multiple chains of inference required for the answer, the memory load. There is no hope of guessing the right answer, and common sense is probably insufficient.

Yet the problem can be solved readily enough, once you know how to go about it. And that is an intriguing thing about problems like this--a confusing and difficult task may, it seems, be made a good deal more tractable by certain rules, gimmicks, and heuristics. The solver must somehow structure the problem into something that can be worked on in a fairly routine way. If this structuring, and the subsequent operations, can be made less variable and more efficient, then some control can be attained over the solution process.

In complexity, logic problems[1] lie somewhere between syllogistic reasoning items and the reading of plain-text; the major logical relations are inclusion, exclusion, and ordering. It is assumed that the solver can use the relationships in ordinary life to sort out people and things: thus, 1876 occurs before 1880; Monday is earlier in the week than Thursday; $10.00 is not exactly divisible by 3; mothers are older than their sons but not necessarily older than their husband's nephews; a sportswriter writing a weekly column does not work for the Daily;[2] a physicist will remember his high-school algebra;[3] and so forth. Linear syllogisms, or three-term series, may be nested inside a larger problem--for example, it may be stated that:

Bill earns more than Larry.

Ted earns less than Larry.

Who earns the least?

Logic problems have certain advantages for research purposes. Th are fun to do, at least to some people; and thus they escape the aridi of the syllogistic reasoning item with its deadly "some S is not P" phraseology. If a solver uses a matrix format, as many of our subject: do, then each response entered in the matrix can be observed and timed with relative ease, and the subject can see where he is. Often, a seq of discrete responses can be interpreted in terms of the apparent reaso ing being done by the subject. If the logic problem is being solved a a computer terminal, the supporting software can be arranged to provide a present-status summary, to calculate such quantities as "proportion c necessary information already entered," and to suggest which sentences ought to be combined to yield new inferences (Bond, Gabrielli, and Rigr 1977). Elegant scoring systems derived from information theory can prc duce such indexes as the "discrete entropic cohesion" between problem attempts (Watanabe, 1969; Guiasu, 1977).

[1] "Logic problem" seems to have no standard meaning in academic psych We use the term here to represent multi-sentence membership proble which are amenable to a matrix format. Some puzzle magazines have regular logic-problem section, and there are thousands of people w are regular devotees. Books that contain various brain-teasers of have logic problems in them; but the number of sentences in their problems is usually small, and solution often hinges on a single i

Another reason for studying human performance in logic problems is practical: many technical activities require the conversion of words into strict symbols. Computer programming is the prime example, of course; but there are many others. A maintenance technician who attempts to use a tech manual must convert the words given there into discrete actions appropriate to his problem. And here is a case from the inventory control domain (Gildersleeve, 1970). The following paragraph gives the logic in narrative form:

> When the quantity ordered for a particular item equals or exceeds the minimum discount quantity and the order is from a wholesaler, give the customer a discount and make the shipment. This presumes that there is sufficient quantity on hand to fill the order.
>
> If the quantity ordered is less than the discount quantity, bill at regular rates and make the shipment even if the customer is a wholesaler. Do the same if the sale is retail.
>
> If there is not sufficient quantity on hand, bill as above, ship what can be shipped, and backorder the remainder of the order. It must be emphasized that, in this situation also, even if the discount quantity is ordered, if the customer is a retailer, the discount is not given.

To convert this narrative into a clear action policy, the words must be transformed into a discrete-states decision table, as shown below. Then to carry out the policy, a person has only to observe which of the eight states obtains, and then follow the instructions in that column.

	1	2	3	4	5	6	7	8
Quantity-Ordered Discount-Quantity	Y	N	N	-	Y	N	-	Y
Wholesale	Y	Y	N	N	Y	-	N	N
Quantity-Ordered Quantity-On-Hand	Y	Y	Y	Y	N	N	N	N
Bill At Discount-Rate	X	-	-	-	X	-	-	-
Ship Quantity-Ordered	X	X	X	X	-	-	-	-
Bill At Regular-Rate	-	X	X	X	-	X	X	X
Ship Quantity-On-Hand	-	-	-	-	X	X	X	X
Backorder Quantity-Ordered Less Quantity-On-Hand	-	-	-	-	X	X	X	X

Figure 1. A Discrete-States Decision Table for a Word Problem

Many of the processes in this words-into-symbols process are exhibited also in the logic problem. Indeed, we believe that in an increasingly digitized world, the complex conversion from word to discrete symbol is one of the most important problems in applied psychology.

II. BEHAVIORAL ANALYSIS OF LOGIC PROBLEMS

We first became interested in logic problems because of certain computer programs which were designed to assist the solver who faced a well-structured verbal puzzle. Since we had worked for some years with computerized trouble-shooting aids and diagrams, we hoped that programs which could handle verbal inputs would lead to more effective fault-locating behavior. A human might, for instance, learn to imitate certain features of a word-processing program, and thereby improve performance. Findler's Universal Puzzle Solver (Findler, 1973) accepts logical relations among variables as inputs, and achieves a solution _via_ recursive search through the set of possibilities. At present minicomputer speeds, Findler's program computes a solution in a second or two. Wang's theorem-prover program (Wang, 1961) works by testing "theorems" or hypotheses against the input "axioms" or logical relations. The total number of possible theorems is very large. You keep testing theorems until you find one that is valid; if Proposition P states that the butler is the murderer, on the basis of logical evidence, then a theorem stating that the butler is the murderer will be valid (Raphael, 1975). Incidentally, the basic Wang algorithm is so effective that it proved _all_ the 200 or so theorems of Whitehead and Russell's _Principia Mathematica_ in a few seconds, on a now-outmoded IBM 704 computer. Both the Findler and Wang programs require, of course, a human to translate the verbal problem conditions into a format that the machine can process. In this respect, Findler's routine is much easier to use, as it accepts string inputs for the variables, and prints constrained English sentences as output. Findler's program is also an elegant piece of work from the computer-science standpoint, it has a very compact search routine, and the whole SNOBOL program is less than 200 lines long.

Even so, when we tried the Findler and Wang programs with college students, subjects were rather annoyed at the input constraints, and much preferred a paper-and-pencil environment to the computerized procedure. This was partly due, no doubt, to the slow teletype terminal; it took a while for the subject to enter a logical relation, and if he wanted a summary of progress on the problem, then there was another wait for the machine to clank out a response. Often, the student wanted a "micro-response" from the computer; say, whether a given relation had already been entered or not, or a summary of "what I already know" about a single variable. Our BASIC version of the Findler program did not provide for all queries of this type. Thus, the machine "aid" was often perceived as something of an intruder rather than an assistant (Bond, Gabrielli, and Rigney, 1977).

It soon became evident that a few major activities were discernible in all logic-problem solution attempts, regardless of whether a computerized or manual environment was employed. After a preliminary scan of the problem stem and the sentences to establish the dimensions, a typical solution starts by noting logical identifications and exclusions from the separate sentences. Each one of these may reduce the number of possibilities, and thus will facilitate final classification of all the variables. In the reference problem given on Page 1 of this report, every sentence contains such information. Sometimes the logic is very simple; from sentence 2, we can exclude Lamont from either Agriculture or Nebraska; from sentence 4, we can be sure that the Treasury man and the Kentucky man are the same. But there are more complex and implicit bits of information, too. Sentence 1 yields several relations; we see there that Smith cannot be the Postmaster General, the Secretary of State, or the Attorney General. But we can also infer that Smith and these other three people make up a set of four "short-timers," who did not serve full terms. This part of the solution attempt is intra-sentence processing.

When all the separate sentences have been milked for their logic, the problem usually will not be solved. Information from two or more sentences must be combined, to get new inferences. In our example, combining Sentences 1 and 6 will allow you to infer immediately that Smith was from Georgia, and was the Secretary of Interior. Combining sentences 1, 6, 9, 10, and 11 yields (among others) the firm conclusion that the names of the four short-timers were Smith, Wilson, Olney, and Harmon. This kind of reasoning we call inter-sentence processing; it is often more complex than intra-sentence work, because the immediate memory load is higher, and because it is often hard to know just which sentences should be put together. After watching numerous solution attempts, we believed that there was more variability in the inter-sentence processing than in most other aspects of the performance.

Some problems have an ordering feature, and the solution depends on how skillfully the ordering information is extracted and handled. The "Pie Contest" shown below is rather simple, as our problems go, but it illustrates the ordering aspect. Our early observations indicated that people differed appreciably on this aspect of performance. Some people did not fully utilize the problem datum about the fifth and sixth place pies; this hindered or prevented solution. We often saw subjects keeping little order notes on the margin of the problem sheet. All this suggested the desirability of separately scoring the ordering behavior.

THE PIE CONTEST

Ann, Bea, Dot, Eve, and Sue won the top six prizes in last year's pie-baking contest at the Centerville County Fair (one of the women was lucky enough to have two entries among the six award-winners). The awards went--not necessarily respectively--to an apple pie, a strawberry pie, and a sweet potato pie. From the following clues, can you determine which pie received each prize, and who baked it?

1. The judges awarded the apple pie first prize and the peach pie second; they decided the cherry pie was too sour to place in the top five.

2. Bea's ranking was lower than Ann's but higher than Dot's.

3. Sue did not receive either first or sixth prize.

4. Ann has never attempted a fruit pie.

5. The sweet-potato pie ranked just below the chocolate pie.

6. The woman who won fifth place also received sixth prize.

On a large problem, the solver will usually need some way of recording his inferences as he goes along. And there will be a "collection" phase where the solver has to go around and "pick up" all the inferences made so far. When this happens, the subject will often mutter something like "...let's see just where I am on this thing." Schwartz (1971) found that a matrix format was best; and puzzle magazines often include a matrix with the problem. We therefore decided to provide a matrix with each problem, and to encourage subjects to use it, even though we knew that some subjects would scribble other kinds of notes as they went along. By this means, we hoped to standardize and control the recording and collecting variable.

Obviously, these four processes do not exhaust the behavioral domain of the logic puzzle. Each "process" can further be separated into finer-grained components; the level chosen by the investigator is often a decision of convenience (Sternberg, et al, 1978). Actual solutions are often multi-level affairs. They have episodes when a rather flickering search over the problem is carried out; perhaps the subject is trying to find something that he has overlooked, or forgotten, or entered wrongly; he can look for lines and columns that are nearly filled up; or he may "change levels" and attempt to get below the superficial membership conditions. Nobody seems to have studied imagery in logic problems, and though our subjects usually do not report strong visual imagery, they may well visualize spatial lists and arrays when they are attempting to order things.

An expert logic-problem subject may not need a matrix aid, and may not do much in the way of recording. Consider again our Cabinet problem on Page 1. After the usual quick scan through it, the expert will often focus on high-information sentences--those that mention several variables. The expert will quickly perceive the importance of the first sentence, which effectively splits the eight cabinet members into two groups of four men. Four men finished the 1893-97 term; four didn't. The short-timers, then, are:

Smith
Postmaster General
Secretary of State
Attorney General

What else do we know about the short-timers? From Sentence 6, the Secretary of the Interior, who was from Georgia, was also one of this group of four, and he can only be Smith. Also, from Sentence 10 the Attorney General is from Ohio. Now we have:

Smith---Interior---Georgia
Postmaster General
Secretary of State
Attorney General---Ohio

From Sentence 9, the Secretary of State is from Massachusetts, and cannot be Smith, Wilson, or Harmon (Sentence 5); so he is Olney. Then Harmon must be the Attorney General, with Wilson as Postmaster General. So our short-timer tableau is already complete:

Smith---Interior---Georgia
Wilson---Postmaster---West Virginia
Olney---Secretary of State---Massachusetts
Harmon---Attorney General---Ohio

The remainder of the problem now breaks easily (Lamont, Herbert, and Carlisle are not from Nebraska, so Morton is). But this rapid "expert" solution, which actually was performed by a staff member, utilized a rather special focus on the first-sentence chain of reasoning. This particular solver liked to find a solution without using a matrix or elaborate notation, and so was especially tuned to cues that encouraged long-but-rapid inference chaining. Much practice, and perhaps hundreds of problem attempts, were required to gain this facility. An experienced solver may even recognize the "style" of the puzzle author; one of our staff members believes that he can identify problems in Dell Crossword magazine which are written by Randall Whipkey. When he sees that Whipkey is the author, he immediately looks for some obscure little verbal cue in the problem, and follows that intensively, in the hope of locating an informative "catch."

But our main interest was not directed so much to spectacular solutions like the one just outlined; we first wanted to see whether separate scores of intra-sentence, inter-sentence, and ordering behavior could predict levels of performance on moderately difficult problems. If so, then further work could proceed to separating these three activities into finer components. Schwartz (1971) had reported that a regular "logic test" was not significantly correlated with performance on hard logic problems; but we hypothesized that his logic test may have been too syllogistic. Logic problems do, of course, follow the rules of syllogistic reasoning; but the structuring of the problem, the word-into-symbol translations, the arrangement of inferences, and the assumptions about the world--these things are not strictly syllogistic processes.

III. PROCEDURES

A. Predictor Variables

1. Sentence Scores. We wanted three separate indicators of word-into-symbol processing: intra-sentence, inter-sentence, and ordering. To produce these scores, the subjects independently derived the logic from a set of eleven sentence sheets. Each sheet had some sentence material, and a specially-labeled response matrix. All the sentences were taken from problems which were not used elsewhere in the study. Six of the sheets had only a single sentence, and the subject's task was to enter into a problem matrix all the dots (definite "yes") and X's (definite "no") that could be inferred from that one sentence. The last five of the sheets had two or three sentences on them, and there the subject had first to enter the logic from the separate sentences, and then to combine information across sentences for new dots and X's. Four of the sentence sheets had ordering implications, so that order-related responses could be separately identified and evaluated. As far as the subject was concerned, no special instructions were given with regard to the ordering variable. Two of the sentence sheets are shown in Appendix A; the second one has an ordering feature.

Scoring of the sentence-sheet matrices was done by counting each dot as one point; each correct X that was separately inferrable was also counted as one point. No additional points were given for row and column X's that followed automatically from the entering of a dot. In the ordering score, only those responses were counted on which at least one dimension (e.g., money, age, days of the week) was clearly ordered, and where the ordering was necessary for problem solution.

The time each subject spent on each sentence sheet was individually recorded by an unobtrusive observer who used a free-running second counter. These times are accurate to within a second or two. Subjects were urged to work along steadily, but not to rush their work. Once a subject had finished with a sheet, he could not go back to it. Exactly 110 inferences could be logically deduced for the whole set of eleven sheets; 36 of these were order-related inferences; and 19 were strictly inter-sentence responses.

2. Reading Scores. As a reference measure of verbal ability, the McGraw-Hill Reading Test was given to all subjects. This test takes about two hours to complete; it yields six separate scores, of which two are reading rates (McGraw-Hill, 1972).

3. Discrimination Scores. Posner showed that closely-timed classification tasks could be separated into additive components (Posner & Keele, 1967). His experimental paradigm distinguishes different levels of processing; for example, the letter pair AA shows physical identity; pair Aa has name identity; and the members of pair AB differ in both physical and name aspects. Typically, the physically identical pair AA is classified as "same" some 50 or more milliseconds faster than is the Aa pair. Hunt and others have explored the possibility that this difference in "code access" time might indicate the basic information processing ability of individuals (Hunt et al, 1973). We decided to include a Posner-type classification as one of the tasks each subject completed.

Our setup used the same letter-discrimination pairs described by Posner & Keele (1967); the major technical difference was that our timing was voice-keyed instead of finger-keyed. After a two second warning, the stimulus pair was presented and the timer was started. The subject looked at two stimulus figures in the display, said "same" or "different," and the vocal response keyed an automatic digital timer. After a series of calibration and practice trials to smooth out the responses, a counterbalanced set of 48 data-taking trials was given for Physically Identical, Name-Identical, and Different pairs. All sessions were individually conducted, in a quiet laboratory room.

B. Criterion Variable

Six problems were chosen from materials we had worked with in the past; all were taken from published puzzle collections, though we made some minor changes to reduce ambiguities. Everybody took the problems in the same order, with problems of highest expected difficulty at the end. Every problem was individually timed for each subject, with 35-minute time limits on the later problems in the series; pretests had shown that half an hour is a practical maximum for a problem, with undergraduate subjects.

C. Subjects

The subjects were 39 undergraduate psychology students at a large Western university; each one received subject-pool credit, and also was paid $2.73 per hour for participating. Five subjects were eliminated: two because they apparently could not solve the simplest logic problem correctly, and three because they had incomplete data.

D. Administration

The McGraw-Hill Reading Test was given in two group sessions; all other materials were administered individually, or in small groups with several observers in the room. Since no subjects had regularly worked logic problems before, about an hour and a half of break-in instruction was given; during this period, several simple problems were worked out, and the matrix format was illustrated and practiced. Each subject had to finish a real problem while the observers walked around the room and observed the dots and X's being entered into the matrix. There were individual differences, of course, in how quickly the instructions were appreciated; but all subjects had to be working efficiently before going on to the scored material.

The sentences and the criterion problems took most of the experimental time, with each subject requiring several hours each on these segments of the study. On average, to complete the entire series of measurements, a subject spent a total of about twelve hours, and appeared for five or six experimental sessions. Over a long session, breaks were given about every 90 minutes. All subjects appeared to try hard; most seemed to experience some frustration, and occasional elation, during the criterion problems. Many remarks were heard to the effect that ". . . I felt like I was close to breaking that problem, but I just couldn't get another big dot."[4]

[4] In computer-aided problem presentation, it is possible to score a problem matrix for "percent information already achieved," and to print out this parameter to the solver. Sometimes the solver already has entered enough data in the matrix to get a full solution, but has not "collected" the relations efficiently. This information-collection failure may explain some of the frustration of our subjects; perhaps they had already "solved" the problem, but didn't realize it, because the data had not been put together across rows and columns.

IV. RESULTS

The principal data of the study consist of fifteen scores from each subject; these fifteen scores were:

Criterion Problems Solved
Intra-Sentence Score
Inter-Sentence Score
Ordering Sentence Score
McGraw-Hill Reading (6 scores)
Posner-type Discrimination (5 time scores)

Means, variances, and intercorrelations of these 15 scores are shown in Table 1.

A. Criterion Problems

The number of criterion problems solved correctly ranged from 0 to 5, out of a maximum possible of six. With an average solution score of 1.5, the problems were obviously quite difficult for this group of people. Nobody solved the final problem in the set of six, and only two of the final set of 34 persons got five correct. In general, our predicted order of problem difficulty was accurate, with most of the solutions being achieved on the first two. Since the subjects filled out a matrix for each problem, it was possible to calculate part scores on each problem, and then to sum these over problems. These part scores, however, proved to be very highly correlated with a simple count of number solved; so we simply used number solved as the criterion variable.

B. Sentence Scores

All three sentence scores had rather high variances; there are marked individual differences in translating words into symbols, and all three were highly related to the number of problems solved, with r's ranging from .61 to about .75; two of the scatter plots are shown in Figures 2 and 3.

			Total Sentence Score	Sentence Ordering	Inter-Sentence	McGraw-Hill Reading Rate 1	McGraw-Hill Reading Rate 2	McGraw Hill I	McGraw Hill II	McGraw-Hill III	McGraw-Hill Total	Letters-SD (Diff)	Letters-PI Physically Identical	Letters-NI Name Identical	D-PI	NI-PI
Variable	Mean	S.D.	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Criterion Probs	1.5	1.46	.69	.75	.61	.31	.11	.55	.48	.32	.54	-.42	-.11	-.20	-.53	-.28
2. Sentences	60	18.85		.82	.78	.30	.11	.42	.56	.44	.56	-.22	-.04	-.10	-.31	-.16
3. Ordering	15	8.46			.61	.40	.16	.56	.59	.44	.64	.29	.09	.23	.33	.39
4. Inter-sentence	2.7	1.81				.33	.22	.42	.44	.26	.47	.03	-.15	-.05	.30	.20
5. MGH RR1	48.3	8.92					.76	.43	.49	.26	.46	-.49	-.31	-.44	-.33	-.48
6. MGH RR2	54.0	9.78						.39	.24	.10	.25	-.35	-.10	-.23	-.45	-.37
7. MGH-I	54.7	9.89							.54	.33	.77	-.41	-.15	-.31	-.47	-.48
8. MGH-II	51.4	8.46								.33	.83	-.53	-.39	-.46	-.26	-.31
9. MGH-III	57.5	9.29									.68	-.24	-.15	-.15	-.13	-.04
10. MGH-TTL	55.6	8.64										-.44	-.26	-.34	-.30	-.30
11. Letters (Diff)	729.3	101.12											.83	.86	.31	.15
12. Letters (PI)	647.3	100.82												.93	-.27	.15
13. Letters (NI)	717.1	114.92													-.09	.50
14. D-PI	83.3	57.18														.42
15. NI-PI	70.4	41.49														

Table 1. Means, Standard Deviations & Correlations

N = 34

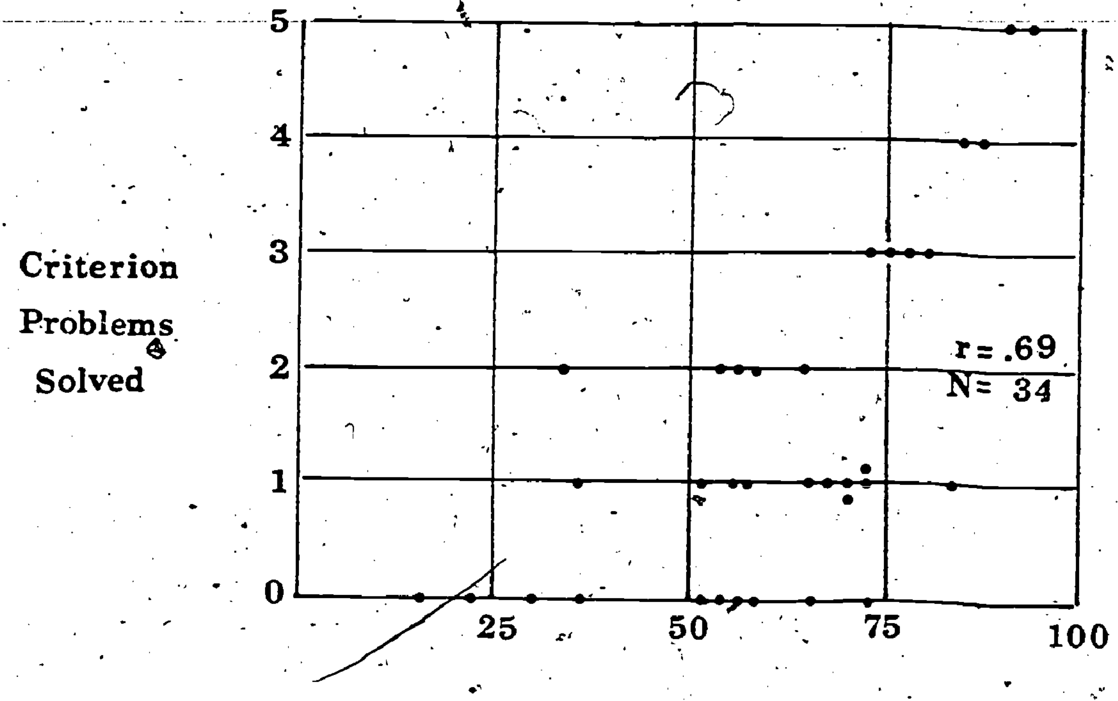

Fig. 2 Total Sentence Score

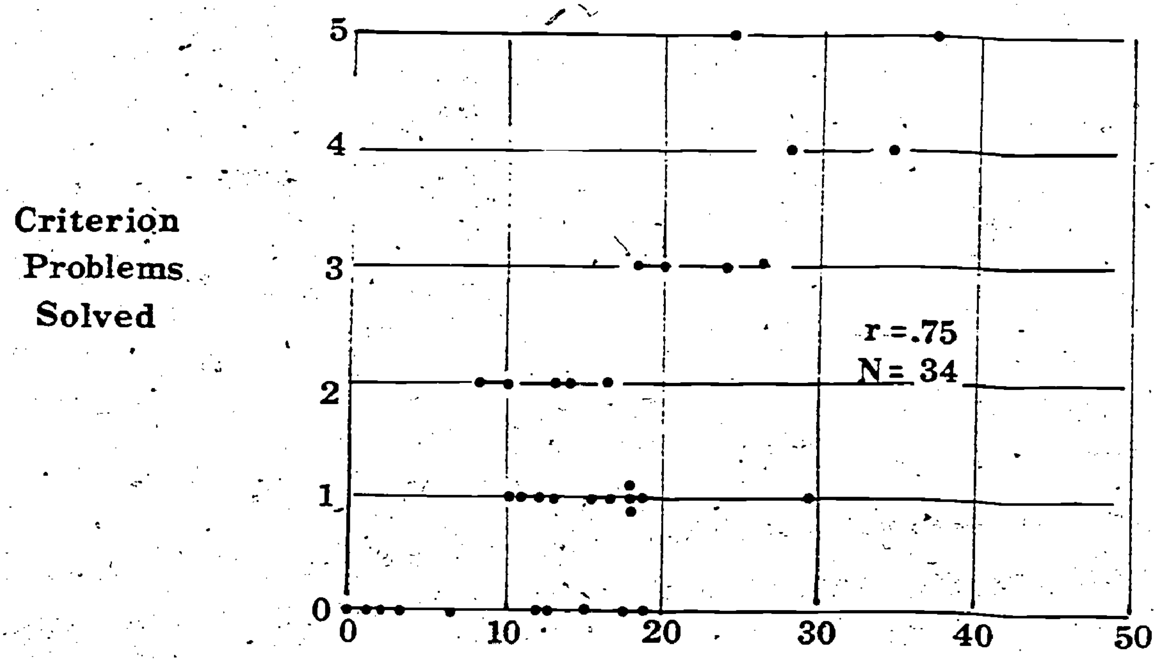

Fig. 3 Linear Ordering Sentence Score

29

The highest correlation obtained with number of problems solved was with the linear ordering sentence score, which had an r of nearly .75. The predictability at the high end of Figure 3 is remarkable; of the eight people who solved three or more criterion problems, all were above the mean on ordering. And only one subject who had an ordering score of 16 or more failed to get at least 3 criterion problems. In fact, the predictability of overall criterion performance from sentence inference is so high that scores from a single sentence were correlated .74 with the criterion problems.

Altogether, the data indicate that a rather high sentence score is a necessary but insufficient condition for solving the criterion problems. We were surprised at the generally low proportion of inferences that the subjects attained from the single sentences; the median subject managed to get only a little more than 50% of the total number of correct dots and X's that were logically derivable; the top score achieved was 94 out of 110, or about 85% of those possible. Many subjects, then, had trouble extracting the logic from the English sentences, and so their later solution efforts on complete problems were doomed to failure. There were some very low sentence scores; for instance, five people got zero on the inter-sentence measure, which means that they missed all the inferences derived from two or more sentences; so all five would be expected to fail any problem which depended on this kind of process. The generally high correlations between sentences and problems suggested to us that the teaching of "logic extraction" might be the best way to promote excellent performance on logic problems.

Intercorrelations among sentence 1, sentence 2, and so forth were generally positive and moderate; for instance, sentence 6 had a median correlation of .46 with the other ten sentences. There was one exception to this: "Sentence A," the seventh in the series and the first multi-sentence sheet, was correlated negligibly with most other measures. Perhaps the material on this sheet had a technical or semantic defect; more likely, the inter-sentence skills are more difficult and less practiced, and the subjects experienced some cognitive strain when first encountering combinations of sentences. In any case, these moderate correlations between sentence sheets suggest the existence of a reliable dimension for assessing and predicting problem performance.

C. Reading and Discrimination Scores

As can be seen from the correlation matrix in Table 1, McGraw-Hill Reading scores correlated from .11 to .55 with number of problems solved; this was a little higher than we expected, in a college population which was presumably selected on reading. Time scores on the Posner-type classification task also correlated negatively with problems solved and correlated less with sentence scores; for later analyses, the discrimination time scores were subtracted from a constant, so that a high score was a "good" score. In accordance with Posner's results, physically-identical letter pairs yielded the shortest response-time medians (647 ms.). An incidental finding here was that discrimination response times for female subjects tended to be slightly lower than those for males. Two difference scores were calculated, Name Identical - Physically Identical and Different - Physically Identical. The reliability of these difference scores should be appreciably lower than the constiuent discrimination scores.

D. Specific Responses in Sentence and Problem Matrices

With a mean sentence response score of only about half of the possible inferences, and with a difficult criterion set, it was inevitable that some inferences would seldom or never be achieved. This was the case with several of the inter-sentence relations; again, the data show that people have trouble in combining logic across sentences, and in converting this logic into a matrix record. As we reviewed the answer matrices, it often seemed that subjects preferred to stay at a superficial level when combining sentences, and avoided the "deeper" and more complex relationships. Maybe they did this because a superficial "one-pass" reading was often sufficient to get <u>some</u> inferences, and so the subject would be reinforced for such reading. Right now, this is only a conjecture; but the matter should be subject to experimental investigation of the "depth" variable. "Depth of processing" has been shown to affect memorability of learned list material (Craik & Lockhart, 1972); perhaps a subject's "logical depth," or fluency in shifting from one level to another, can influence solution efficiency. Maybe explicit training in analyzing phrases and sentences at different "depths" would alert subjects to the possibility of pursuing more than one level of analysis, and would promote better solutions. We have enough material to score problems for depth-required-to-solve; and we generally know whether a given matrix entry is derivable from superficial or subtle considerations.

Each subject contributed hundreds of timed responses. And there are so many sequences and response rates in the records that we have not yet thoroughly analyzed them. Early in the study, we decided not to pressure the subjects unduly about their rates of responding; but often there were unmistakable slowdowns when inter-sentence processing began.

E. Factor Pattern

One should not take seriously a factor analysis of more than a dozen scores on 34 subjects; but we did put the correlations from Table 1 through a principal-components analysis and Varimax rotation, using the California State University packages. The loadings on the first three factors are shown in Table 2.

Variable	Factor 1	Factor 2	Factor 3
1. Criterion Problems	-.63	-.06	-.10
2. Sentences	-.32	-.05	+.10
3. Ordering	.15	.22	+.13
4. Inter-sentence	-.43	-.13	-.05
5. MGH RR1	.06	-.41	-.70
6. MGH RR2	.18	-.17	-.89
7. MGH - I	.06	-.18	-.28
8. MGH - II	.00	-.47	-.04
9. MGH - III	-.06	-.01	-.03
10. MGH - Total	.05	-.27	-.07
11. Letters (Diff)	-.29	-.80	-.33
12. Letters (PI)	-.11	-.96	-.05
13. Letters (NI)	-.10	-.96	-.11
14. D - PI	-.29	.28	-.64
15. NI - PI	-.01	-.31	-.42

Table 2. Varimax Rotation of
Three Principal-Component Dimensions

N = 34

The factor pattern certainly seems plausible. Problem variates (criterion puzzles and sentence scores) define Factor 1. Factor 2 is clearly a Posner-discriminant dimension with three loadings from .80 to .96; Factor 3 has two high weights (.70 and .89) for the McGraw-Hill reading scores. Two of the Posner difference scores also load on this factor.

According to the "discontinuity hypothesis," there is a rather clear demarcation between fast, overlearned, automatized processes, such as the discrimination skills in the Posner task, and less well-practiced mental operations such as our sentence-logic extraction. The factor structure obtained above is generally in accord with the discontinuity idea; and we have noted before how slow and halting is the inter-sentence exploration and translation. There are appreciable correlations between "good" scores, whether the scores represent "fast" or "slow" operations; but the scores do cluster into rather distinct classes.

We were surprised at the .28 and .46 loadings of Variables 14 and 15 on Factor 3. If Factor 3 is primarily a reading-skills dimension, then why should difference scores from Posner classification load so highly on it? Assuming that the .28 and .46 loadings are not artifacts, we conjecture that time difference between Physically-Identical identification and identification of the other stimuli loads primarily because of rate scoring features in the scores involved. Factor 3 has its only two high loadings on McGraw-Hill "Reading Rates", not on McGraw-Hill total, or McGraw-Hill vocabulary. Perhaps the scanning features that determine Reading Rate have some basic oper tions in common with Posner difference scores. The data from Hunt et al (1973) also suggest that some time-difference scores can be more informative than simple reaction time. Obviously, a next step in this area is to confirm the difference-score phenomenon on a larger data set, and to control experimentally the scanning requirements in reading and in classification tasks. The search for basic information-processing parameters is tantalizing and elusive.

McGraw-Hill Reading Rates are determined on the basis of rapid word recognition; so if a subject were operating under a high-speed instruction, it would not be necessary for the subject to comprehend the McGraw-Hill sentences in their entirety. Instead, the subject could search for a critical key word or two, and give his item response as soon as a key term was observed. There might, then, be natural similarities to the Posner difference scores, because a "fast noticer" on the Reading Rate should be one with a low difference score from the Physical-Identical baseline. This interpretation deserves checking under differing Reading Rate time pressures. We should expect that the loadings of the Posner tasks should be reduced as the McGraw-Hill task shifts from emphasis on speed to comprehension..

F. Practice Effects

Both problems and sentences got harder as the sessions went along; so we cannot assess accurately the impact of the several hours of practice. All subjects, though, were quite comfortable with the matrix format after a few sentence sheets; and they all learned that a nearly-filled-up matrix was a sign that a criterion problem was about to break. A few students came back and asked to finish problems that they had not completed; and some said that they had gotten into a problem so much that, later that day, they found themselves still thinking about it, and perhaps seeing a new inference or two while driving home. In our experience, this kind of interest in an experimental task is rare, and extends quite beyond the Zeigarnik effect.

V. A TRAINING DEMONSTRATION

Once we knew that criterion problems were so highly dependent on sentence-processing components, it seemed worthwhile to try to teach these skills to a fresh set of subjects. In a few hours, we probably could not materially improve McGraw-Hill Reading or Posner-type classification capabilities, but we thought that the word-into-symbol activities might be subject to some rapid training. In our own working and coding of many logic puzzles, we had developed a little set of rules and heuristics which we had found to be effective; and these techniques seemed to be eminently definable and teachable.

The demonstration was planned for only seven subjects, on the ground that, if you can't demonstrate your technique on a handful of people, then there is little reason to think you can do it with a hundred people. We chose seven more undergraduate psychology students whose mean McGraw-Hill Reading Scores were matched to those of the original set of 34 subjects. This group received exactly the same break-in and criterion-problem procedures as the previous group, but they also went through six hours (one day) of intensive training in sentence-processing skills.

During the training, typical-problem sentences were taken one at a time, and the instructor worked with the trainee until all inferences were correctly extracted. Sometimes it was necessary to give hints and prompts. All trainees learned to work very carefully and rather slowly, since they saw that the criterion was "total correct dots and X's," rather than some speed index. During the last half of the training, three complete problems were worked; if the trainee got hung up, then the instructor gave only enough of a hint to get the process going again; sometimes it was sufficient to point to one of the heuristics on the board.

A. The Training Program

Five rules were explicitly taught for aiding the inter-sentence and data-collection aspects of solution. These rules were:

1. Dot Combination. Suppose a dot (definite "yes") is entered in a problem matrix; to use our cabinet problem again, you discover that Smith is from Georgia, and was Secretary of the Interior. It then follows that "Smith is everything that Georgia is;" and *vice versa*. For example, if Smith and Georgia were on two rows, then every dot and every X in Smith's row now applies to the Georgia row. This rule may appear obvious, but we discovered that many of our subjects did not apply it fully or systematically, perhaps because one of the dotted variables would be in a row, and the other in a column. There is a bit of a knack to applying the rule rapidly in a large matrix; and of course if there are many dots, you have to keep track of the ones you have covered. We taught our subjects to use tick-check marks for this purpose.

2. Choosing Sentences to Combine. Suppose you have already entered all your intra-sentence logic into the matrix. Now to find out whether to combine two sentences, say A and B, look for those sentences that have:

   (a) one or more shared variables;

   (b) the shared variable has to be a "positive" in A, and a "negative" in B; if this is true, then a new inference can be obtained by combining A and B; otherwise, nothing can be learned from combining.

This rule is easily generalized to three or more sentences; in fact, a computer program was written which does this automatically for a given data matrix (students did not use this program as an aid; if they had, solution would often have been instantaneous).

3. High-Information Sentences. Many problems have a long sentence or two in which every dimension, or nearly every variable, is included. These should be especially studied, because they often "split the problem into two parts," or otherwise lead to exclusions that produce many matrix entries. One example is Sentence 1 in the problem on Page 1. Here's another illustration:

> The Collins and the Jones boys won events before Brenda's son, but after Steve and the Allen boy.

If there are five boys in this problem, and all five are mentioned here, there are many inferences that flow from this one sentence. A close look at the ordering data indicates that Brenda's son must have been last, that neither Steve nor the Allen boy can be lower than second, with Collins and Jones in third and fourth places.

4.. Partial but Effective Exclusion. You are working a problem with a five-variable set of people who work for five different newspapers, and you know that one of the five, say Jim, is either on the Daily or the Press; but you don't yet know which one of these it might be. Now you can exclude from Jim any shared exclusion between the Daily and the Press. If both the Daily and the Press, for example, do not have a soccer-column writer, then Jim does not write the soccer-column, regardless of whether he ends up on the Daily or the Press. This logical rule is not often recognized by the ordinary educated person. In this respect it resembles Wason's modus tollens inference task (Johnson-Laird & Wason, 1970; Wason, 1968), which is extraordinarily difficult for educated adults.

5. Partial Ordering. Trainees are encouraged to set up little ordering schemes, where that is possible. As one example:

> John's event was before the horizontal bar event;
> there were three events between them.

Now if the solver knows that there are five events and realizes fully the implications of this sentence, he will see that John's event was first, and the horizontal bar competition was last. (Incidentally, when this sentence is combined with the one about Brenda's son, just above, we see that John must be John Allen, that Steve was the second competitor, and that Brenda's son was in the horizontal bar event).

In the course of the training day, numerous cases arose when these five rules could be applied. A list of them was put on the board, and the student was advised to refer to the list, and perhaps work through it, whenever a problem was at a sticking point. During the last hour or so of training, a moderately difficult problem was worked by the trainee, usually without help.

After their day of special training, the subjects took the same six criterion problems attempted by the earlier group. They also took two additional problems, which Schwartz (1971) had given to 38 Ss at Wayne State University.[5] One of these extra problems was "conjunctive-positive," with three dimensions and five values; the other was "conjunctive-negative," with four dimensions and five values. These problems

[5] Actually, the "negative" problem as published by Schwartz is garbled and insoluble; perhaps there was a clerical error in transcribing it for the journal. Our revised version contained 13 sentences, and is shown in Appendix B.

were added because they were large but they were also "clean"--that is, there were no verbal subleties in the sentences describing the relations. Here is a sentence from the negative-conjunctive problem:

Neither the Japanese nor the Englishman owns a hyena.

The matrix entries here are obvious, and will be done by nearly every trained subject. So even though there are 13 sentences like this, there will be little hesitation in entering the dots and X's.

B. Results

The trained people entered correctly 73% of the possible sentence inferences--a notable gain over the 50% or so correct in the original group. On the old criterion set, they "came closer" to filling up matrices and getting problems, on the average; but their performance total-completion scores were identical to those of the previous group of subjects (average of 1.5 solved, out of six possible solutions). On the two "large but clean" Schwartz problems, however, all seven subjects solved each problem, perfectly and quickly.

We interpret the main result as follows. The Schwartz problems, though large and unwieldy to a neophyte, yield readily to the matrix technique and the five processing tricks that we taught. Hence, we got rapid and perfect performance on those two problems. Our original criterion problems, though, often rest on fairly obscure relations which can be understood by the subjects, when they are pointed out, but cannot always be perceived readily by the subjects. These more subtle factors are called implicit relations by Polich & Schwartz (1974); those authors found that errors in formulating these relations were much more frequent, and much more decisive, than the explicit or surface relations in the problem. It is true that our training helped, even in the subtle problems; the trained people correctly made a larger proportion of the matrix entries; they made fewer errors; they kept working; the list of five rules always gave them something to do, and so they made progress. But they still missed a key inference or two, enough to prevent total solution.

Ordering scores and inter-sentence scores were also notably improved in the trained subjects. There were no zero scores, and rather few erroneous inferences; nearly all mistakes were omissions.

On one of the criterion problems, an observer sat next to each solver and timed every matrix entry that was made. The resulting record then gave a timed solution trace through the problem, and permitted response-rate determination at different stages. There are strong hints in this material that, for the direct declarative logic from the sentences, the response rates are fairly typical across subjects. Figure 4 shows fragments of time records for two subjects who worked on the pie-contest problem. In both cases, the early, regular entries are due to the straightforward recording from the sentences. After this chore has been accomplished by the solver, there are often some delays; presumably, the time is spent in organizing an inter-sentence search strategy. The effectiveness of a special technique or aid can be evaluated by the control that is achieved over such delays; in the present instance, there was still much uncertainty by the trained subjects after the easy matrix entries were skimmed off the top; they took a lot of time to figure out what to do next.

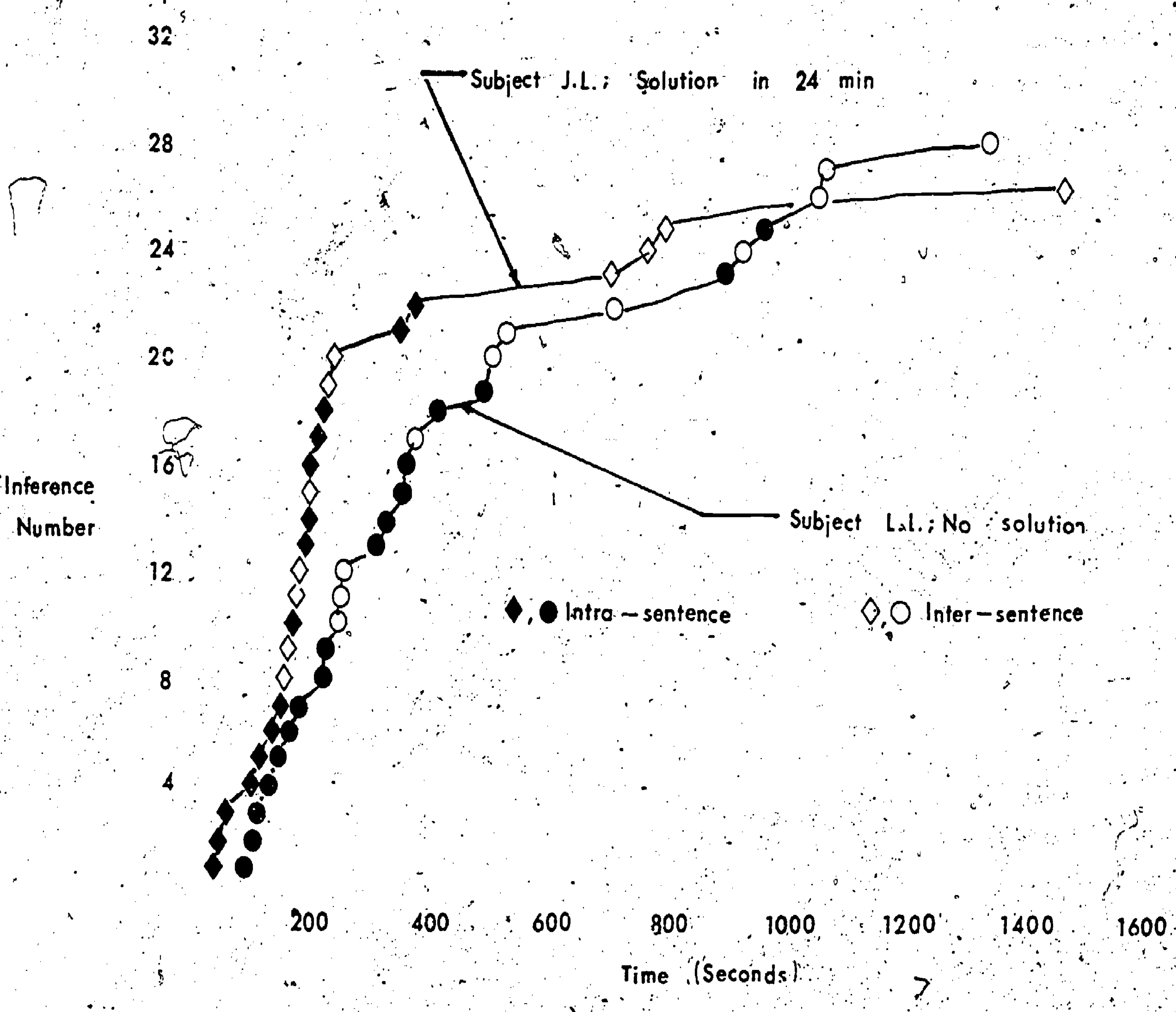

Fig. 4 Time Record for Two Subjects; Pie Problem

Both of the two performances in Figure 4 show the typical "early spurt" of inferences, followed by later pauses of many seconds. A qualitative review of the records, though, suggests that J. L. was not only going faster in the early stages; his solution had a better sense of direction and focus, and his search for "things to combine" seems to be easier to follow, and to make more sense. Subject L. L. was making inferences, all right; but they were logically "jerky," and not well connected with each other. So, L. L. could not find the crucial "dots" he needed, and did not have a sharp search plan when his separate bits of information proved to be insufficient.

It is easy to become fascinated with timed protocol data like this, and to see processes that may not be general. However, we must mention one thing which struck us as we went over the records. It is this. When a subject starts to enter dots and X's in the matrix, he or she does this for awhile, and then slows down. Our most successful people seem to be those who keep going, and keep entering things. We think now that this variable may be part motivational, part habitual, and part cognitive, and is the sort of aspect usually referred to as "concentration," or determination, or mental energy. Perhaps this aspect of performance also would be subject to modification, via training and imitation of problem-solving "models" who enter data at steady rates.

VI. DISCUSSION

Our results were positive in demonstrating the centrality of sentence processing in logic problems, and somewhat encouraging with respect to radically and quickly improving these processing skills in difficult problems. It could be that our training method, if continued over several days or weeks, would have finally produced much better performance on difficult, subtle problems. We are inclined to the view that, if the verbal and semantic subtleties are causing many of the problem hangups, then analysis should be directed to the subtleties themselves. A logical next project, then, would be to assemble the hardest-to-achieve inferences from the present data, to frame plausible conjectures about the reasons for difficulty, and to test the conjectures by systematic variation of the materials. We already know that faulty syllogistic reasoning, though it undoubtedly occurs, is not a major source of failure to our subjects.

The attempt to teach problem solving had an impressive transfer to the two Schwartz problems: perfect performance was attained, when our matrix method and heuristics were applied to a complicated total situation, which was made up of many simple sentences. If this result can be confirmed on a larger sample of people, it should be meaningful to applications in such areas as troubleshooting. Troubleshooting in electronic and mechanical equipment is often a difficult and critically important task; much time is spent on training, on tech manuals, and on aids of one kind or another. Yet good troubleshooters are as hard to find as ever. The aids provided to them are often ineffective;

and the crises continue. When a draft copy of this report was being reproduced, four Xerox technicians were working on a recalcitrant copying machine; and they finally called their office for further help.

In our present framework, the electronic technician who is taught "how to troubleshoot" is like one of our "trained" subjects. The technician realizes the logic of fault isolation pretty well; he knows generally how to interrupt a signal chain with critical tests; but he is operating on a complex "set of sentences" in his head, and in his manuals. These sentences may contain large amounts of information, all right, and the technician extracts some of it; but there are subtleties in his "sentences" which he has not yet appreciated, or has forgotten. As long as these exist, no logical tricks will work; the problem matrix will remain in an incomplete state; and the trouble will not be located. When effective troubleshooting devices are produced, as occasionally happens, they work just because the "sentences" underlying them are clean and clear and the fault-isolation behavior can proceed with the certainty that it is converging on the problem. The quality of the given technician's "sentences" could be assessed by the proportion of correct inferences that could be drawn from his available source material. We could then, in principle at least, estimate the likelihood of a given technician ever finding certain kinds of trouble. Certainly the people who depend on technicians, and who provide their training and reference manuals, should be interested in such an estimate!

More research should be done on the highly-practiced and efficient solver of logic problems, since the behaviors exhibited there are really what the applied psychologist wants. We noted earlier one "expert" solution to the Cabinet problem; that performance had a flow and elegance that seems to be lacking in our rather pedestrian matrix technique. Maybe our matrix skills, which we have shown are quite teachable, should be considered as a prerequisite to a smoother and more rapid expert approach. From our staff experience, there is reason to believe that people who become very fast on the matrix business can go on to expert-type solutions. If this proves to be generally so, then the matrix technique would be supportive to the other cognitive operations, and would not be the major technique used by the solver. Perhaps real fluency in the matrix skills could be taught in, say, a week or less of intensive practice; and then the course could go on to subtle relations.

In an impressive series of studies, Robert Sternberg and his associates have analyzed syllogistic reasoning into a series of about ten operations, and have been able to estimate the time spent on each operation (Sternberg, 1978a; 1978b). Our problem material is often considerably more complex than his three-term syllogisms; but there are several places where the processes seem to be similar. Sternberg's "pivot search," for example, requires the subject to establish a term which will permit the combination of two premises into a single ordered array; this may take several seconds, in his situation. Our solvers must find logical "pivots" too, but the number of admissable alternatives is often much larger than in the three-term case, so discovery may require many seconds, or even some minutes. In Sternberg's models, the "availability" of a solution element is sometimes a function of memory from the immediately preceding operation. This factor might be operative in our logic problems. If sentences are phrased or "bunched" ~~so as~~ to facilitate appreciation of a dimension (age, order, etc.), then increased availability of inclusions and exclusions should show up in the problem matrix. We have tried a few problem variations of this kind on small grab samples of subjects, and occasionally an effect can be obtained. In one version of our "Murderer" problem (Bond, Gabrielli, & Rigney, 1977), we put certain sentences with a shared dimension next to each other, in the hope that appreciation of that dimension would be enhanced. The effect we found was only slight; but there is so much complexity in our problems, and so many ways for the subject to work, that such effects are often masked by the variability. Sternberg gets much of his data from highly practiced subjects on constrained problems; so perhaps a "componential" attack on our logic problems would proceed best with expert solvers.

One of the most powerful features of Sternberg's work is his estimation of the accuracy of various performance models; he gets R^2's on the order of .90 and better for some of his models. Sternberg used the R^2 parameters to decide which of several models is best, and to suggest what proportion of variance is yet to be explained. It is possible to predict performance in our logic problems rather well from sentence-logic extraction activity, as we have seen above. But it is also possible to predict general difficulty of the problems from superficial features, such as number of sentences in the problem, number of variables, and number of sentence combinations required for solution; we get r's in the 60's or higher, for this kind of prediction.

We believe that a measure of logical "depth" is needed for the inferences in our problems, and we are now exploring a simple three-level depth score. If this scoring scheme works, then probability of solution in a given problem might be a function of a person's average "realized depth" in several similar problems. Given a person's ability in perceiving relations at various logical depths, we could model his/her performance on any given problem, via the depth parameters of the problem, and a digital simulation program.

Data from our problems have some significance for individual differences in cognition. High-verbal ability people tend to be faster at making "name" matches than are low-verbal ability subjects (Hunt, Frost, & Lunneborg, 1973); and high-verbals are also much faster than people with low-verbal ability in making taxonomic category matches and homophone identity matches (Goldberg, Schwartz, & Stewart, 1977). Our results suggest that appreciation of the strict logical relations implied in a sentence is also related to verbal ability. Perhaps the "logic encoding rate" will turn out to be a useful information-processing parameter for individuals.

Hayes et al (1977) employed logic problems in a simulation study of human reasoning. Their work is in the Carnegie-Mellon "production system" tradition. First, they asked human subjects to make relevancy judgments about the material in each of the problem sentences. As expected, the subjects were able to ignore extraneous information, and to focus on the key elements of the problem. The investigators then hypothesized that three problem-structuring processes were operating: (1) a SETS heuristic which identifies groups of items; (2) a TIME heuristic which tags items containing time-related phrases such as "yesterday;"

and (3) a QUESTION heuristic which places great relevancy-value on items which are in a query mode, and which tend to define a solution to the problem. A SNOBOL computer program was written to imitate these processes, and the program output of "relevancy" or "meaning" resembled the human judgments rather closely.

Our data certainly confirm the critical importance of activities like SETS, and the powerful orienting effects of QUESTION sentences. Indeed, we believe that skill in rapidly defining sets, and in separating large sets into subsets, is a distinctive mark of the good problem solver. For an inexperienced subject, the set-defining activities may be observed from the very beginning of the solution process. There is also a close resemblance between the Hayes TIME heuristic and our "ordering" variable. One of the first things an expert solver looks for is information about ordered arrays in the sentences. And if order data are there, the expert generally hopes to find that one or two problem sentences are especially rich in the order "subset" logic; often, too, the "big" ordering sentence leads rather directly to "smaller" exclusion statements; so the sentence-combining decisions are easier.

Our seven trained subjects, we believe, would solve the Hayes "All-sports" problems very quickly. This is because all the sets and relations are obvious (even though some are irrelevant), and because our matrix skills would afford rapid exclusions of the (few) name-sport pairings.

There appears to be no intrinsic barrier to coding problem sentences for their semantic contents, with a relatively full listing of meaning categories for each term in a sentence. If this can actually be done, then production-system models and aiding programs could capture the essential behaviors in logical problem solving. The results of the present study indicate that the actual processing of problem data can be routinized, once suitable inputs are recorded into a matrix, or into some other kind of memory. We know how to teach, and to assist _via_ computer, the logical-inference part of the problem. But there are still many technical issues in getting good data from the sentences into the logical processor; and these must be mastered before an aided problem-solving system can be confident of solution in really different problems.

REFERENCES

Bond, N.A., Gabrielli, W.F., & Rigney, J.W. Studies of Verbal Problem Solving. I: Two Performance-Aiding Programs. Los Angeles, Calif.: Behavioral Technology Laboratories, University of Southern California. Technical Report No. 83. August, 1977.

Craik, F.I.M., & Lockhart, R. Levels of processing: A framework for memory research. Journal of Verbal Learning & Verbal Behavior, 1972, II, 671-684.

Findler, N.V., & Willis, B.M. A "Universal" Word Puzzle Solver. International Journal of Man-Machine Studies, 1973, 5, 53-74.

Gardner, M. Mathematical puzzles & diversions. New York: Simon & Schuster, 1961.

Gildersleeve, F. Decision Tables. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Goldberg, R.A., Schwartz, S., & Stewart, M. Individual differences in cognitive processes. Journal of Educational Psychology, 1977, 69, 9-14.

Guiasu, S. Information theory with applications. New York: McGraw-Hill, 1977.

Hayes, J.R., Waterman, D.A., & Robinson, C.S. Identifying the relevant aspects of a problem text. Cognitive Science, 1977, 1, 297-313.

Hunt, E., Frost, N., & Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), Advances in learning and motivation. Vol. 7. New York: Academic Press, 1973.

Johnson-Laird, P.N., & Wason, P.C. A theoretical analysis of insight into a deductive problem. Cognitive Psychology, 1970, 1, 134-138.

Polich, J.M., & Schwartz, S.H. The effect of problem size on representation in deductive problem solving. Memory and Cognition, 1974, 2, 683-686.

Posner, M.I., & Keele, S.W. Decay of visual information from a single letter. Science, 1967, 158, 137-139.

Raphael, B. The thinking computer. San Francisco: W.H. Freeman, 1976.

Schwartz, S.H. Modes of representation and problem solving: Well evolved is half solved. Journal of Experimental Psychology, 1971, 91, 347-350.

Sternberg, R.J., Guyote, M.J., & Turner, M.E. Deductive reasoning. New Haven, Conn.: Yale University Psychology Dept., Technical Report No. 3, January 1978.

Sternberg, R.J., & Turner, M.E. Components of syllogistic reasoning. New Haven, Conn.: Yale University Psychology Dept., Technical Report No. 6, April, 1978.

Wason, P.C. Reasoning about a rule. Quarterly Journal of Experimental Psychology, 1968, 20, 273-281.

Watanabe, S. Knowing and guessing. A quantitative study of inference and information. New York: John Wiley, 1969.

Wylie, C.R. 101 puzzles in thought and logic. New York: Dover Press, 1957.

APPENDIX A

Typical Sentence-Inference Sheets

Five women received special honors at the annual awards luncheon of the ladies hospital auxiliary.

1. Ann and Mrs.Trask, each work one day a week; Bea, Ms. Quinn, and the woman who received the ribbon all work two days; each day of the week is worked by at least one of the five.

2. The woman who received the ribbon works with Ms. Ross on Tuesdays and with the armband winner on Mondays.

3. The woman who received the certificate sees Eva at work every Friday and Mrs. Sussman every Tuesday.

	ANN	BEA	CLAIRE	DONNA	EVA	armband	badge	certi	pin	ribbon	M	Tu	W	Th	F
POTTER															
QUINN															
ROSS															
SUSSMAN															
TRASK															
M.															
Tu.															
W.															
Th.															
F.															
armband															
badge															
certif.															
pin															
ribbon															

APPENDIX B

Modified Schwartz Problem

The five men in order of their salaries, from least to greatest are: the Army's employee, Mark (whose las name is not Reiner), the shoe salesman, David, and Mr. Dixon (who does not sell jewelry.)

		SALARY					GOODS				
		$9,500.	9,700.	9,750.	10,000.	10,500.	Furniture	jewelry	shoes	sports	women's apparel
LAST NAME	DIXON										
	GRAY										
	HOLMES										
	HOLT										
	REINER										
GOODS	furniture										
	jewelry										
	shoes										
	sports										
	women's apparel										

Modified Schwartz Problem

1. The hyena's owner doesn't live in the white, yellow, red, or green house.
2. Neither the Japanese, the Indian, nor the Englishman lives in the green house.
3. Neither the American nor Canadian owns a zebra.
4. The tea drinker doesn't live in the blue house, and doesn't own a turtle, hyena, ox, or horse.
5. Neither the Japanese nor the Englishman owns a hyena.
6. The beer drinker isn't English, doesn't live in the red house, and doesn't own an ox.
7. The zebra's owner doesn't live in the yellow or red house, and doesn't drink milk.
8. The coffee drinker doesn't own a zebra, or an ox, live in the yellow, blue, or red house.
9. The Japanese doesn't live in the red, blue, gree, or yellow house.
10. The American doesn't live in the red, blue, or green house.
11. One of the men drinks whiskey.
12. The American does not drink whiskey or milk and does not own a house.
13. The milk drinker does not live in a blue house.

	White	Yellow	Green	Blue	Red	Zebra	Hyena	Ox	Turtle	Horse	Tea	Beer	Whiskey	Milk	Coffee
Japanese															
Indian															
Englishman															
American															
Canadian															
Tea															
Beer															
Whiskey															
Milk															
Coffee															
Zebra															
Hyena															
Ox															
Turtle															
Horse															