DOCUMENT RESUME

ED 030 452 LI 001 487

By-Reilly, Kevin D.

The Library and Human Memory Simulation Studies. Reports on File Organization Studies.

California Univ., Los Angeles Inst. of Library Research.

Spons Agency-National Science Foundation, Washington, D.C.
Pub Date Mar 69
Note-21p., Presented at Seventh Annual Symposium on Biomathematics and Computer Science in the Life Sciences, Houston, Texas, March 26-28, 1969.
EDRS Price MF-$0.25 HC-$1.15
Descriptors-Computer Programs, Computers, Information Processing, *Information Systems, *Libraries, Library Networks, *Memory, *Models, *Simulation

This report describes digital computer simulation efforts in a study of memory systems for two important cases that of the individual the brain; and that of society, the library. A neural system model is presented in which a complex system is produced by connecting simple hypothetical neurons whose states change under application of a transformation rule at intervals determined by random time functions. Analogies of such a system to a library network are drawn. A word-concept model is described in terms of a master file of linked words and phrases with auxiliary files linked to the master. Applications of such a model for student exercises and for library-related and other research projects are discussed. Also discussed are models for library networks for three different subsystems: the computer processing center, the role of the human actors, and the distribution of materials. Such models, though principally developed for library systems analysis, are useful for their formal properties as memory system components. (Author)

THE LIBRARY AND HUMAN MEMORY

SIMULATION STUDIES

by

Kevin D. Reilly

One of a Series of Reports on File Organization Studies

NSF Grant GN 422

Presented at the Seventh Annual Symposium on Biomathematics and Computer Science in the Life Sciences

Houston, Texas

March 26-28, 1969

The Institute of Library Research

University of California

Los Angeles, California

THE LIBRARY AND HUMAN MEMORY

SIMULATION STUDIES

by

Kevin D. Reilly

One of a Series of Reports on File Organization Studies

NSF Grant GN 422

Presented at the Seventh Annual Symposium on Biomathematics

and Computer Science in the Life Sciences

Houston, Texas

March 26-28, 1969

The Institute of Library Research

University of California

Los Angeles, California

TABLE OF CONTENTS

ABSTRACT

This report describes some digital computer simulation efforts in a study of memory systems for two important cases: that of the individual, the brain; and that of society, the library. A neural system model is presented in which a complex system is produced by connecting simple hypothetical neurons whose states change under application of a transformation rule at intervals determined by random time functions. Analogies of such a system to a library network are drawn. A word-concept model is described in terms of a master file of linked words and phrases with auxiliary files linked to the master. Applications of such a model for student exercises and for library-related and other research projects are discussed. Also discussed are models for library networks for three different subsystems: the computer processing center, the role of the human actors, and the distribution of materials. Such models, though principally developed for library systems analysis, are useful for their formal properties as memory system components.

The report provides material of both tutorial and research interest. Of tutorial interest are the scope of the systems analysis, the advanced programming system concepts, (e.g., file organization, data types), and features of programming languages, PL/1, SNOBOL IV, and GPSS/360. Research interest lie in the particular digital simulation models and the comparative analysis of memory systems.

I. Introduction

A great deal of interest centers on the study of memory systems. Taken in its broadest sense that phrase refers to study of organization, functioning, and utilization of a variety of different kinds of systems: information retrieval systems; computer memory; data banks; libraries and library networks; hypothetical neural systems ("brains"). Here, we discuss a few examples of digital computer simulations that we are currently undertaking. The studies cover a broad area overlapping topics in virtually all the areas just mentioned. Our goals for the most part are research; however, we must note a tutorial role.

Parts of our digital simulation modelling for library systems are being designed with the intent of developing tools for usage by students in library systems analysis. Parts of the "word concept" provide illustrations of language data processing and simulation, data structuring, and file organization. Another tutorial aspect concerns the matter of programming ease; librarians, who we hope will soon be making use of models of their own libraries, cannot be expected to become computer experts. The question then that must be answered is whether a serious-minded library school trained individual can perform a reasonably comprehensive study on his own or whether he must be backed up by a staff of researchers and programmers. Our approach often involves use of "highest" level programming languages (e.g., GPSS/360, SNOBOL IV). Links to conventional "higher" level programming languages (e.g., FORTRAN, PL/1) and the System 360 Assembly Language are sometimes necessary, but, insofar as possible, they are to be avoided.

II. Hypothetical Brain Models

Simulation of brain-like systems has been part of the study of Artificial Intelligence for some time now; see, for example, Borko (1962), Greenberger (1962), Green (1963), Wooldridge (1963), Arbib (1964), Fink (1966). It is helpful to mention but a few of the titles of the studies in these references:

Reliable Brains from Unreliable Neurons

Resonant Frequencies in Neural Networks

Formation of Concepts by Means of the Automatic-pattern-interconnection Principle

Computer Simulation of Cognitive Processes

Simulation of Human Thinking

Computer Models of Psychological Processes

Simulation of a Brain

The modelling efforts to be discussed in this section are within this general framework. The first of these is a model for a system of randomly connected simple neurons. Part of the reason for our studying such a model is to aid in the development of formal models for memory systems with possible applications in the area of information retrieval systems. Such models can be utilized for formal analysis of reliability, communication, noise analysis, etc. The second of these is a model for a system of words and phrases interconnected in such a way as to exhibit, upon program analysis, relationships of word concepts. Such a model is of interest for semantic analysis, file structuring, and concept behavior simulation.

A. "Complex" System Model

Walker and Ashby (1966) discuss what they mean by their "certain complex systems" in these terms:

1) structurally intricate,
2) structurally rigid,
3) built of simple parts,
4) functionally homogeneous,
5) not influenced by factors outside the system,
6) clocked (quantized time).

"Structurally intricate" in this context implies random connections between the (large number of) neurons. The "simple parts" refers to the fact that the neurons exist in either of two states (e.g., the "on" or the "off" state). The "functionally homogeneous" refers to that fact that the same transformation rule (defined in terms of the state of the neuron and its (two) afferents) is identical for all neurons. Other aspects of the models are presumably more transparent. If we single out the "structural intricacy" and the "simple parts" as being the most important features of the model, we establish continuity with our further developments.

The model originally proposed by the above authors is easy to program. A number of elementary additions (e.g., setting the neuron connections, as part of the program) are also easy to implement. The generalizations that we are interested in (e.g., outside disturbances; random timing of events; different methods of scoring system behavior, etc.) are not all easy to program. We therefore decided to utilize the user-oriented programming language GPSS/360. Let us briefly discuss some aspects of the programming effort.

A basic element of the dynamics of the system (assuming an already established set of connections) is the programming of the following decision table (the transformation rule for changing the state of the neurons):

Present State of Inputs		Present State of the Element	
L	R	0	1
0	0	e_1	e_5
0	1	e_2	e_6
1	0	e_3	e_7
1	1	e_4	e_8

This can be handled effectively in GPSS through use of a TRANSFER block with a VARIABLE in the "C" field. The VARIABLE produces an unique value, depending on the values of three SAVEVALUES, representing the neuron state and the states of its (two) afferents. Transfer of program control then ensues to a statement which changes the neuron-state SAVEVALUE. A single transaction may be used in a loop with three of its parameters undergoing change and being used in indirect addressing of the neurons and afferents.

Besides the above set of statements, the only other element of the dynamics is that of changing the states of the processes. Indirect addressing similar to the preceding case along with simple assignment statements take care of this need. Each of the two main phases of the program, as described, can be handled as an independent chain of blocks in the GPSS program. This makes interpretation of the program most direct.

The author's use of the system is to study development of cyclic (excitatory) behavior in subsystems, the length of the cycles being one of the points of interest.

Our concern is primarily with generalizations of the simple model. A first generalization consists in providing for delays in the system. Delays at the neurons (i.e., as the neuron changes state) and delays of the transmission of the neuron state down its processes to other neurons seem like logical extensions of the system. Such delays are assumed to be random time functions and consequently, the method of scoring the system's performance can be generalized from cycles etc., to total excitations at points where a congregation of neurons are postulated to exist. Synchronization of inputs can be accomplished through use of MATCH blocks; especially of interest would be cases in which afferent impulses were postulated to hold only for a certain period. Other generalizations (e.g., using more afferents and "threshold" logic, as does Culbertson (in Borko, loc. cit.)) might be of interest. Each of these generalizations make the formal system more library-like: delays at the neurons correspond to processing delays; delays in transmission of excitation correspond to message delays; more input channels for some neurons provides a distinction between neurons not unlike the distinctions between relatively busy and non-busy libraries.

B. Word Concept Model

Semantics is a relatively unexplored area of Linguistics, though its importance is being increasingly stressed. An unabashedly computer approach to a "theory and simulation of some basic semantic capabilities" is espoused by Quillian (1967). The basic model, designed principally as a model for long-term memory, consists of system of linked data items: words and word phrases. The links are of different types and portray relationships between the items. Figure 1, taken from the paper by W. R. Quillian, illustrates some of these ideas. PLANT is a principal or "type" node; it has pointers to "token" nodes, represented as being in the same plane. These nodes define, in dictionary style, the principal node. The token nodes also point to principal nodes outside their immediate plane (depicted by dashed lines, leaving the plane).

The diagram illustrates the several different kinds of links between token nodes. These include: class membership (e.g., PLANT 1 is a subclass of

PLANT. 1. Living structure which is not an animal, frequently with leaves, getting its food from air, water, earth.

2. Apparatus used for any process in industry.

3. Put (seed, plant, etc.) in earth for growth.

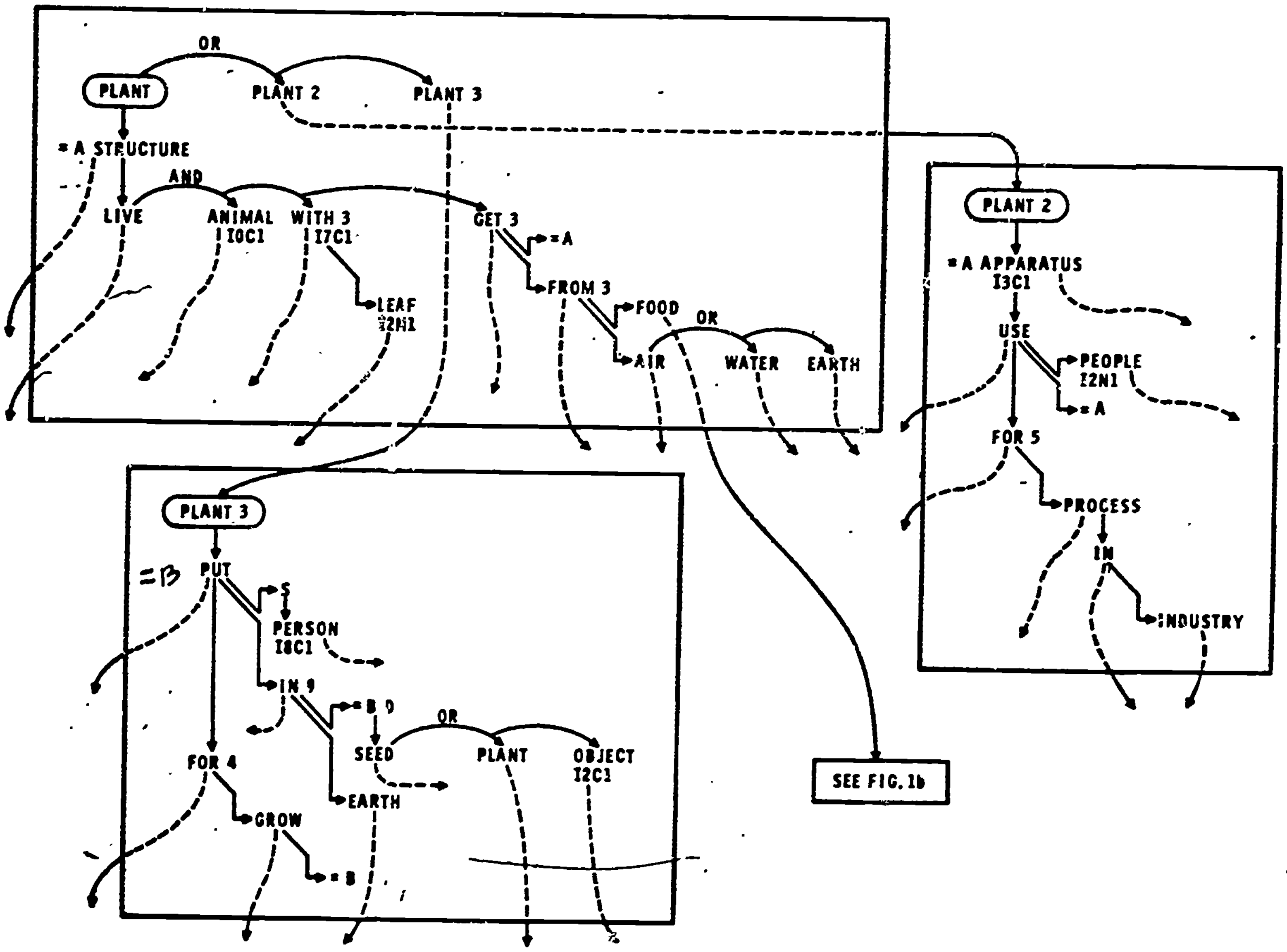

Figure 1: Diagram depicting three planes representing three meanings of plant and illustrating by way of an example some of the features of the system of linked data items constituting the memory system, taken from the paper by Quillian (1967).

"structure"); modification (not just any kind of structure, but a "live" structure); conjunctive set (e.g., a structure that is "not live," "with leaves," and "gets," etc.); disjunctive set (e.g., PLANT 1 or PLANT 2 or PLANT 3); a means of relating any two tokens through a third (e.g., "food" is related to "air or water or earth" via a "from" relation of the "third" type (implying "deriving organization").

The basic use to which Quillian has put his model is to compare and contrast two words. This is done by searching for intersections of nodes. Simple examples are PLANT-FOOD which intersect on the (type) node LIVE; CRY and COMFORT intersect on the node SAD. A "syntax" program joins words about the intersections with some addenda (e.g., the phrase, "among other things") to come up with the final product (e.g., "To cry is, among other things, to make a sad sound; to comfort can be to make someone less sad.").

Our interest in this kind of work has several facets. In the first place, the linked structure concepts of programming have interest in their own right and we have been doing a comparative study of PL/1 and SNOBOL IV as to ease-of-use, generality, core utilization, etc., in handling such structures. In both languages we use recursive procedures. In SNOBOL IV, we utilize the programmer-defined data types, which permit pointer variables as part of the data type. In PL/1, we call directly upon the pointer variables (embedded in PL/1 structures) and the tools for manipulating them. Despite differences in syntax of the languages the data structures and the overall programming logic of the two programs are remarkedly similar. (We are not prepared at this time to indicate the implications of this for programming ease, as the similarity, in part, may be due to the fact that may at times be forcing SNOBOL statements into a PL/1 mold and vice versa.) Our programming effort initially involved data structures which are somewhat more simple than those of Quillian (e.g., binary trees) but more general structures are now being treated. Allowing for features that may be of little use in a direct programming of a Quillian-like system (e.g., deleting of nodes; added fields for sorting etc.) have given the effort a broader manipulations scope, consonant with our need for reasonably broad base for comparison. (Also, some of these capabilities are needed in another (statistical analysis) project.)

A second goal of the efforts in this area concerns research via simulation into semantic capabilities. A hypothetical brain similar to Quillian's, for the denotative aspects of language behavior, is to be merged with an auxiliary file system, perhaps of the Inquirer type (Stone et. al., 1966), for the connotative aspects of language behavior. In such a framework, many aspects of language performance, e.g., response time, role of affective states, etc., can be simulated.

The third goal of the efforts is to develop materials for student exercises for both computer science and library school students. Generally speaking, computer science students are ready for the most complex data structures, semantic analyses, and computer programming system comparisons. Unfortunately, much of the research, particularly in comparative studies, is either out-of-date or quite superficial; hence a tutorial effort geared for advanced students is necessary. We have already commented on the difficulties of educating library school students. Developing approaches for these students that go directly to the heart of the problems is by no means a slight challenge.

III. Information Storage and Retrieval System Simulations

N. Baker and R. Nance (1968) provide a summary of some of the work done in the area of simulation of information storage and retrieval systems. They note that most studies have concentrated on what they call "subsystem problems" (e.g., the composition of the descriptor file, etc.), with emphasis being on how to improve system performance (e.g., "maximizing the relevance ratio," minimizing system response time, etc.). Little attention has been paid to the environment, particularly the human elements, i.e., the users and funders of the system. Our efforts have been directed toward remedying these deficiencies. Our user-behavior model represents one step; a second step is inclusion of a library model that provides the environment in which the computer processing center operates. Together, we have a hierarchy of models at these three levels:

1) Computer processing activities
2) User behavior model
3) Material delivery model

The models are designed to operate as a system. A series of reports is now being written which will describe in more detail what is being done. Here, we shall briefly describe some of the features of the models.

A. The Blunt Model

The model of C. Blunt et. al. (1966) is for a computer-processing center. It can be characterized in the following terms:

1) Operations requiring processing time;
2) Linkages existing between different processing components in the system;
3) Service units available within the system;
4) Availability of the service units in terms of their reliability and schedule;
5) Processing load exerted on the system.

The model is designed to be a "general" model in the sense that mixtures of different components can be incorporated into the model for any particular system. It suffers in generality quantitatively (e.g., a limitation on the number of components) and qualitatively (e.g., inability to specify complex queue disciplines).

We programmed an example (in GPSS/360) similar to that discussed in Blunt's paper and have used it for the basis of some of our work on the analysis of the computer portion of the library network. Reprogramming in GPSS removes many restrictions, e.g., the number of facilities, limitations on statistics gathering. Additional effort of a major type involves developing a priority queue scheduling system to supplant, at least in part, the rigid externally imposed scheduling system and FIFO queue discipline of Blunt, et. al.

B. User Behavior Model

Remedying the deficiency of no user description has been discussed previously (Reilly & Hayes, 1967; Reilly, 1969) and so will be treated very briefly here. Restricting ourselves to a single class of service (e.g., book materials) and to a single user, we may describe the t^{th} use of the library in the following terms:

1)	Probability of Consideration to use the library	CON_t
2)	Estimated Service Time mean	$MEST_t$
	Measure of Dispersion	$SEST_t$
	Current Value of EST	$CEST_t$
3)	Need Time mean (independent of time)	MNT
	Measure of Dispersion (indep. of time)	SNT
	Current Value of NT	CNT_t
4)	Actual Service Time mean (independent of time)	MAST
	Measure of Dispersion (indep. of time)	SAST
	Current Actual Service Time	$CAST_t$
5)	Convenience Factor mean (independent of time)	MCF
	Measure of dispersion (indep. of time)	SCF
	Current Value	CCF_t

The probability, CON_t, determines whether the library is a likely choice for solution of a given information need. If the user thinks another source is more likely, or that the effort to achieve the service from the library is too great etc., he does not consider use of the library. All decisions except temporal ones are umbrellaed by CON_t. Comparison of need time and expected service time is viewed as a separate mechanism and is governed by the variables in 2-3 of the above delineation. Both CON_t and the variables of EST undergo changes with experience of the user. Good service and poor service are defined in terms of a comparison of AST and NT and in terms of the convenience factor. An accumulation of poor service produces a low value of CON_t and (perhaps) a high MEST; this, to a large degree, keeps the user away from making requests until better service is provided. Generally, most interest lies in rules of alteration (for EST variables) that produce an accurate picture of the AST distribution.

Figure 2 presents a birds-eye view of the kind of output that can be realized from the model. Of course, specific assumptions must be made about the variables in the model and on request costs and returns from satisfied requests. The accuracy of perception of the AST distribution is at issue here under different rules of exploitation of the derived information. Note that MEST values between MAST and MNT appear to be the point of optimal return.

C. Distribution of Materials Within the Network

Libraries, of course, consist basically of materials in "conventional" forms: books, microfilm, reports, newspapers, etc. Consequently, one of the most important aspects of library-based information retrieval networks is the matter of distribution of these resources. Equally important is the distribution of requests over these materials. These features are accounted for by means of a catagorization scheme for both users and material items along subject lines (Reilly, 1968). A cross-fertilization index revealing the extent

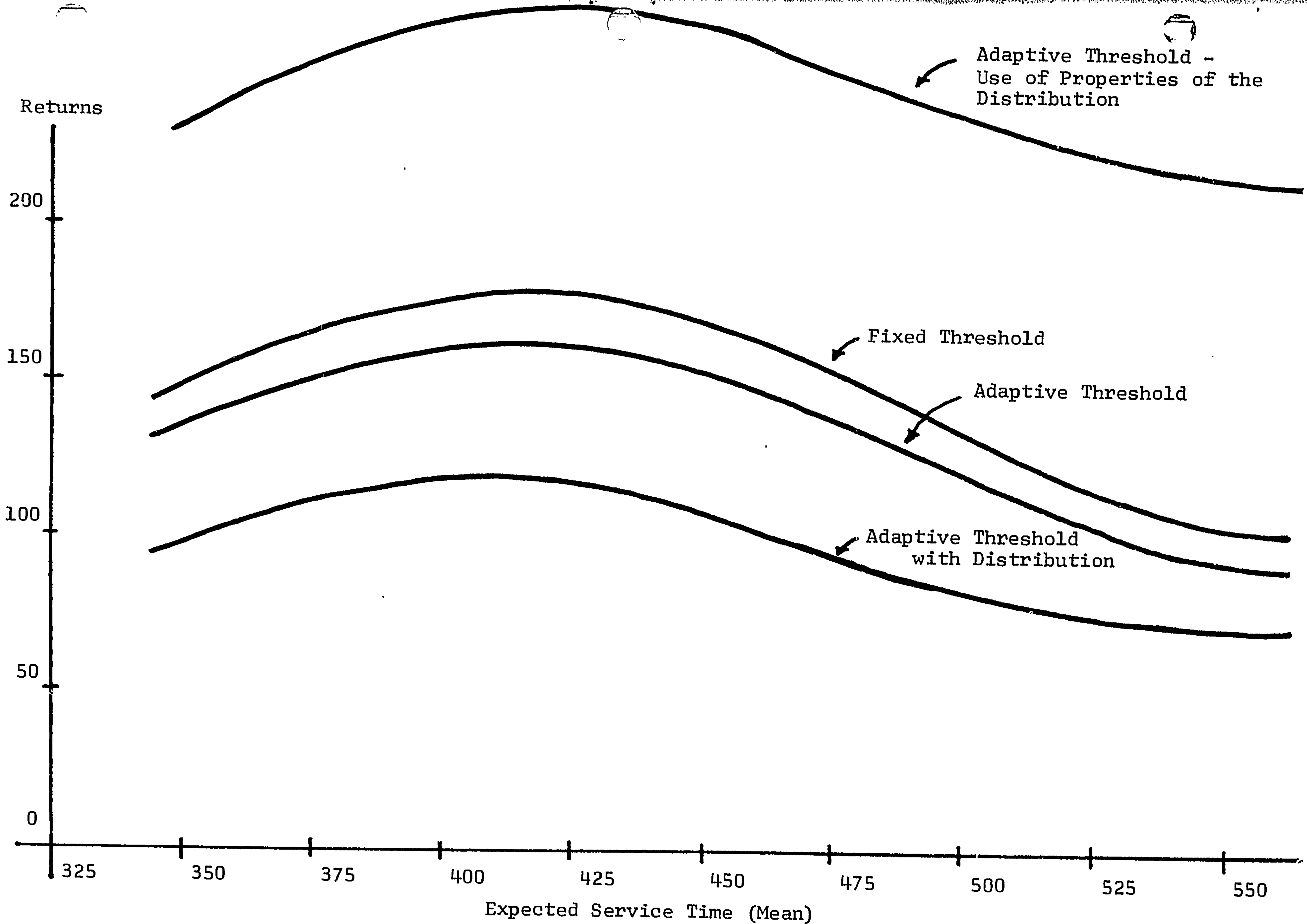

Figure 12: Net returns under different user strategies on assumption of a particular return rate. In those cases where EST is distributed the abscissa is a mean.

to which various areas of study are similar is required, i.e., the extent to which a physiologist seeks biochemical information etc. A representation of the strengths of the member libraries in the various subject areas is required.

A highly computer-oriented network with rapid message delivery at distant locations is destined to uncover the elementary contradiction that locating a material will be a short-time affair, whereas subsequent delivery will be a long-time affair. Thus, though carry-through of a request from its initial phases to the final delivery of the materials may become commonplace but not necessarily smooth. Some form of microreproduction allowing widespread storage of materials or remote delivery via facsimile or similar technology may be the solution to this problem. A below-the-surface problem is that of expansion of the computer file. The very large size of the present machine-readable file is rarely appreciated; yet, we can well expect dramatic increases in the size of this file.

D. Library Models and Human Memory System

D. Norman (1967) lists five operations of human information processing:

1) Sensory transduction
2) Attention and acquisition
3) Short-term memory
4) Long-term memory
5) Retrieval

There are, of course, activities in the library which are analogs to these processes. Some of them appear more clearly after modelling library systems and brain models. (This is part of the rationale for studying both kinds of systems at the same time.) A total systems view for both kinds of systems brings out the comparisons more sharply. Such a view for a library analyst would include the human actors in some integral fashion and for a brain modeler would include stored (e.g., paper) information. A total systems view must not be so total as to postulate that there is no such thing as a brain and that we must view all brains in communication with all stored information. This point of view effectively denies the validity of individual research efforts. This threat, however, is not the current danger; rather most brain modelling seems to reside in a shell in which the ideal appears to be the explanation of the brain's capabilities apart from any environment in which brains naturally tend to find themselves. An example of a case where a broader systems view would have been more advantageous is the following:

> .....To draw strong analogies, however, would be misleading. Libraries are formal structures with rigidly imposed organization; human memory is chaotic and transient.

When looked upon in comparison with its (ideal) assignment to mirror all recorded information, any individual library appears quite "chaotic and transient." Also the formality of structure and the rigid impositions of organization do not seem to truly represent the library community today when vast expenditures of money and effort are going into development of new techniques. We might add that even within computer file systems a great deal of "chaotic and transient" behavior is exhibited (e.g., when an organization has several master files and only a few of them have indexes to them, or when only a part of them can be utilized on-line, etc...).

It might be of some value to point out a few of the more commonly recognized library activities in Dr. Norman's list. We find an analogy to sensory transduction in a process such as that of building of a bibliographic (machine-readable) file. First, information in book or microfilm etc. form is read (scanned) by an indexer, with his thoughts transformed to paper. A keypunch operator translates this sensed message into punched card form. The information must be transformed again into tape format, whereupon it can be transmitted efficiently (by mail) to individual libraries. An analog of attention and acquisition occurs when the new records arrive at the individual library. New master file records can be added onto the end of the master file. If inverted indexes are used (say on subject headings, authors, and titles), a relatively complex set of operations ensue in which sorting and merging operations take place. It is in this operation that we see also a kind of short-term memory. After a period of time, the index records become merged into the index file and are "lost." A quite complex set of operations are needed to bring those records back into existence. Manifestations of short-term memory also occur during the retrieval process, when temporary records are constructed only to vanish when more permanent traces of them are put down. Another form of short-term memory might correspond to the latest master-file records that are (perhaps) kept upon direct access devices, only later to be added to the permanent (long-term memory) tape files. Long-term memory also manifests itself in terms of paper records, the card catalog. In order to keep the total file within size bounds, purging rules are discussed. These correspond to a loss of long-term memory. Retrieval of records and parts of records is dependent on a program, often designed at a relatively simple level. However, extensive forms of file exploitation are under investigation in which heuristics are built in in such a manner that the program simulates some of the gyrations of the human retrieval process. A hierarchy of auxiliary files provides "levels of access" to this library-brain. With time limits placed upon requests, thus restricting the access process and the heuristics, the output will result in more or less complete and insightful responses.

Our final note is on the question of the formation of a general memory systems model. In a very limited sense a programming system of the type of GPSS, the user-oriented programming language with defined data types and a limited number of action-oriented statements, along with a set of prescriptions on how to use it to simulate a given class of systems constitutes a theory on that class of systems. This is, of course, not a totally satisfactory theory, the search for which might well prove to be as elusive as that for a general theory of (all) systems.

REFERENCES

Arbib, M. A. Brains, Machines, and Mathematics. New York: McGraw-Hill, 1964

Baker, N. R. and R. E. Nance, "The Use of Simulation in Studying Information Storage and Retrieval Systems." American Documentation. October 1968.

Blunt, C., Duquet, R. T., and P. T. Luckie, A General Model for Simulating Information Storage and Retrieval Systems. HRB-Singer, Inc. Science Park, State College, Pennsylvania, April 1966.

Borko, H. (Ed.) Computer Applications in the Behavioral Sciences. Englewood Cliffs, New Jersey: Prentice-Hall, 1962.

Fink, D. Computers and the Human Mind. Garden City, New York: Anchor Books, Doubleday & Company, Inc., 1966.

Green, B. F. Digital Computers in Research: An Introduction for Behavioral and Social Scientists. New York: McGraw-Hill, 1963.

Greenberger, Martin (Ed.) Computers and the World of the Future. Cambridge, Massachusetts: MIT Press, 1962.

Norman, D. "The Library and Human Memory." Part 13 of the Final Report on Mechanized Information Services in the University Library. December 15, 1967, 22 pp.

Quillian, M. Ross. "Word Concepts: A Theory and Simulation of Some Basic Semantic Capabilities." Behavioral Science. Vol. 12, 1967, 410-430.

Reilly, K. D. and R. M. Hayes. The Effect of Response Time Upon Utilization of an Information Retrieval System. Presented at the Operations Research Society of America Meeting, New York, May 31-June 2, 1967.

Reilly, K. D. "Outline for a Simulation Study of the California State Library Network." Part 5 of the Final Report on Specification of a Mechanized Center for Information Services for a Public Library Reference Center. Los Angeles, California: Institute of Library Research, July 1, 1968, 15 pp.

Reilly, K. D. Digital Computer Simulation Study of Library-Based Information Retrieval Systems. Presented at the Third Annual Sumposium on Computer Science and Statistics, Los Angeles, January 30-31, 1969.

Stone, P. J., Dunphy, D. C., Smith, M. S., and D. M. Ogilvie. The General Inquirer - A Computer Approach to Content Analysis. Cambridge, Mass.: MIT Press, 1966.

Walker, C. and W. R. Ashby. "On the Temporal Characteristics of Behavior in Certain Complex Systems." Kybernetik, 3. Band, 2. Heft, May 1966, pp. 100-108.

Wooldridge, D. E. The Machinery of the Brain. New York: McGraw-Hill, 1963.