DOCUMENT RESUME

ED 087 814 TM 003 438

AUTHOR Scandura, Joseph M.
TITLE The Structure of Memory: Fixed ~~of~~ or Flexible? Structural Learning Series.
INSTITUTION Pennsylvania Univ., Philadelphia.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE 18 Jun 73
NOTE 66p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Artificial Intelligence; Computer Programs; *Individual Differences; *Information Processing; *Learning; *Learning Theories; *Memory; Problem Solving; Recall (Psychological); Simulation

ABSTRACT

Most current information processing theories of cognition and memory share one common feature: the structure (state-space) of memory is fixed and retrieval from memory involves searching through that structure. Learning, where it is treated at all, involves transforming one such structure into another. This form of representation is questioned and the structural learning theory is proposed to take its place. In comparison, the latter theory has a flexible structure and is shown to have greater power and parsimony, particularly regarding individual differences and learning. Supporting data and relationships with research in artificial intelligence and computer simulation of problem solving are also discussed. (Author)

STRUCTURAL LEARNING SERIES

UNIVERSITY of PENNSYLVANIA
Philadelphia

THE STRUCTURE OF MEMORY: FIXED OR FLEXIBLE?

by

Joseph M. Scandura

Report No. 73
June 18, 1973

STRUCTURAL LEARNING SERIES

University of Pennsylvania
Philadelphia, Pennsylvania

This research was supported in part
by
National Science Foundation Grant GW 6796

THE STRUCTURE OF MEMORY: FIXED OR FLEXIBLE?

ABSTRACT

Most current information processing theories of cognition and memory share one common feature: the structure (state-space) of memory is fixed and retrieval from memory involves searching through that structure. Learning, where it is treated at all, involves transforming one such structure into another. This form of representation is questioned and the structural learning theory is proposed to take its place. In comparison, the latter theory has a flexible structure and is shown to have greater power and parsimony, particularly regarding individual differences and learning. Supporting data and relationships with research in artificial intelligence and computer simulation of problem solving are also discussed.

THE STRUCTURE OF MEMORY: FIXED OR FLEXIBLE?

Joseph M. Scandura

University of Pennsylvania

The view that memory is structured goes back to the old gestaltist notion of grouping. It also finds realization in the notion of associative network. In more recent times, memory theorists have borrowed freely from computer science, particularly from the areas of computer simulation and to a lesser extent from the more behaviorally neutral area of artificial intelligence.

In spite of the great variety which exists among current information processing theories, all such theories share one common feature: the structure of memory is fixed and retrieval from memory involves searching through that structure. Learning, where it is treated at all, involves the transformation of an existing structure into a new one.

In the present article, this form of representation is questioned. The first section introduces the notion of a state space (equivalently, problem space, or relational net) and shows how a variety of prominent memory theories are variants on the common theme. In section two, the structural learning theory is reviewed, together with some closely related empirical research. Finally, relationships between the structural learning theory and relational net theories are discussed and an attempt is made to answer the question in the title: is the structure of memory fixed of flexible?

Relational Net Theories of Memory and Cognition

State Spaces

The notion of state space is very general and has been widely used as a basis for representing a variety of theories involving both computer and human information processing. State spaces consist of two kinds of elements, states and operators. In psychological terms, states refer to (encoded) entities of various sorts (e.g., nonsense syllables, words, concepts, even relations). Operators refer to actions which map given states into other states.

State spaces may be represented as shown in Figure 1 by directed graphs in which the nodes refer to states and the arrows to operators.

INSERT FIGURE 1 ABOUT HERE

Examples of state spaces range from associative networks among common nouns (Bower, 1972) to directed graphs representing the possible stages through which a problem solver might go (Newell & Simon, 1972). The typical state space in problem solving, for example, allows for available operators to act on nodes in all possible ways; psychologically a state space may be thought of as the totality of possible paths among the various states.

In particular applications one or more states must be singled out as starting states, and a goal (G) is defined. Goals may be defined in terms of specific states or in terms of properties which specify a class of states (e.g., "is a check-mate position" in chess).

To achieve a given goal in this view, the subject must find a solution path from a starting state to a goal state. (To actually satisfy a goal, of course, the operators in the path must be applied successively to some starting state(s).) One general approach to finding a solution path is to systematically and exhaustively try out all possible routes, either beginning at a starting state or at a goal state. In breadth first methods (for details, see Nilsson, 1971), all operators emanating from a given node are tried first before the outputted states are expanded. In depth first methods, states furthest removed from the starting state are expanded (until some predetermined depth is reached) before new states are expanded.

Heuristic search methods, on the other hand, attempt to expand promising alternatives first and do not necessarily try out all possibilities. Consider, for example, the crypto-arithmetic task

$$\begin{array}{r} \text{DONALD} \\ +\ \text{GERALD} \\ \hline \text{ROBERT} \end{array}$$

in which the task is to assign digits to the letters so that the two resulting addends sum to the third numeral (see Bartlett, 1958 or Newell & Simon, 1972).

INSERT FIGURE 2 ABOUT HERE

An exhaustive search of the space might move (in a depth first manner) until each letter has been assigned a value. These assignments then would be checked to see: (a) if the letters are paired with the digits in a one-two-one manner, and (b) the assignments satisfy the indicated addition requirement. A more heuristic method, suggestive of human behavior, would be to check the one-to-one and addition requirements as each new digit value is assigned. For example, once 4 and 5 are assigned to T and N, respectively, 4 and 5 are no longer valid candidates for L.

Many refinements of state space representations and search methods have been proposed, of course, but the essentials remain as described: the possible states and operators are represented in terms of a relational net (state-space) and search methods are devised for finding paths between given states.

Learning (or storing information) in this view involves transforming given state spaces into new ones. This may take the form of actually constructing a new space or, as we shall see, tagging or in some other way distinguishing certain states and operators in the given space.

Not surprisingly, a wide variety of current models of cognitive behavior, most particularly in problem solving and memory, are increasingly recognized as having a good deal in common (e.g., Reitman, 1970).[1] The major difference seems to be one of terminology. In problem solving, the starting states are referred to as the "given" information and the goal specifies properties to be satisfied by a solution (cf. Polya, 1962). In retrieval from memory, the starting states correspond to (external) recall cues and information which happens to be active in the processer (short-term memory). The goal refers to to-be-recalled items.

Throughout our discussion, problem solving plays a distinctly secondary role and is considered only where this serves to clarify our main argument.

Memory Theories

Because associative models of memory appear to be giving way to the information processing view, it is perhaps surprising that both kinds of models involve state space representations. These models range widely and deal with the free recall of unorganized nouns (e.g., Bower, 1972), the semantic structure of memory (e.g., Rumelhart, Lindsay, & Norman, 1972; Kintsch, 1972; Quillian, 1968; Collins & Quillian, 1972), the structure of paragraphs (e.g., Crothers, 1972), implication (e.g., Frederickson, 1972), and patterned sequences of symbols (e.g., Simon, 1972; Restle, 1970; Glanzer & Clark, 1963; Vitz & Todd, 1969).

Anderson's model FRAN (reported in Bower, 1972) provides perhaps the most clearly defined associative model in this sense. This theory apparently deals successfully with the free recall of unorganized (non-categorized) nouns.

In FRAN, the initial data base (state space) consists of 262 concepts (nouns), each having between 3 and 19 associative connections with the others (determined from Webster's dictionary). The data base in this case may be thought of as representing the associative connections that the population of subjects might conceivably have learned. Particular (sub)lists of nouns are learned in accord with associative principles. On each trial, an attempt is made to tag (i.e., activate) the presented noun, and pathways emanating from this noun are searched for other nouns in the list to be learned. Where such pathways are found, they are marked with a LIST tag.[2] According to Bower, the effect of such markings is to direct the executive (search method) during retrieval toward marked pathways leading from given nouns to others to be recalled. In common with other associative theories, the marking of nodes and pathways (i.e., learning them) is assumed to be a probabilistic process increasing linearly with study time per item.

Anderson and Bower assume that between two and four items are held in short-term memory (STM), together with newly presented nouns and/or retrieval signals. In addition, three or fewer items are assumed to reside in a similar store called ENTRY SET. ENTRY SET consists essentially of those nouns that are connected to the largest number of associates. In effect, a total of about seven items are assumed to be "active" at any one time, an assumption which has become increasingly common in memory theories ever since Miller's (1956) classic paper was written.

Among FRAN's more unique features as an associative model is that items are not retrieved independently but depend on the items initially available and on successively retrieved items. During recall, the information processer is assumed to respond immediately with the four or so nouns held in short-term memory. The short-term memory nouns together with the three on ENTRY SET then serve as starting nodes from which to commence a search through the associative network. The executive (search) process examines the associative connections emanating from these items in a depth first search until a noun is reached from which no pathways emanate. The search continues only along learned pathways. Nouns at the ends of learned pathways are recalled.

Although they generally give greater attention to semantic and categorical features, existing information processing models are also based on state space methods. The model by Rumelhart, Lindsay, and Norman (1972) illustrates this class as well as any. Here again, the data base is a state space (relational net) and retrieval is like running a maze from various starting points to others.

In the Rumelhart *et al*. model, however, unlike the Anderson-Bower model, no formal distinction is made between the *data base* and *processes* which operate on that base. More immediately relevant here, the nodes in the state space consist of concepts (e.g., bird) and actions (e.g., roll) connected by relations. Although Rumelhart *et al*. are not explicit on the point, concepts may be viewed as classes, or equivalently as properties of items which define classes. Such properties are determined by encoding by insertion into classes (for details, see Scandura, 1971). The processes (in the data base) may serve to retrieve information in the data base or to modify the data base through learning. These processes operate under the constraint of a fixed STM capacity.

The model also includes an executive *interpretative* process which encodes information directly into the data base. The executive, together with certain other unspecified primitive routines, are viewed as necessary features of a workable simulation system which are "defined outside of the memory structure itself /Rumelhart _et al._, 1972, p. 210/."

Among the more significant features of the model are: (a) the possibility of defining secondary[3] nodes (e.g., small bat) in terms of primary ones (i.e., small and bat), (b) a taxonomy of rules of formation for (new) relations, concepts, propositions (i.e., concepts which express relationships among concepts), and operators, (c) explicit processes for forming general concepts from a set of examples and for subdividing concepts (e.g., birds that do and do not fly).

Rumelhart, Lindsay, and Norman (1972) feel that three characteristics most distinguish their model from others of the semantic processing variety. First, rather than tagging new items as in the Anderson-Bower model, for example, the interpreter constructs a list of properties (features) of the items. A general feature of the interpreter is that when STM reaches capacity, an attempt is made to reorganize its contents into higher level categories and thereby reduce the memory load. Second, retrieval is viewed as reconstruction of items from remembered characteristics in STM, rather than as searching for connections between items. Although this distinction is important conceptually, it should be emphasized that it is a direct implication of defining the nodes in the state space in terms of properties (classes). Locating to-be-recalled items still involves searching through a state space. Failure to retrieve an item in the Rumelhart _et al._ (1972) view, results when not enough characteristics of the item have been stored. Third, retrieval is thought to be directed according to explicit heuristic criteria, rather than being relatively non-selective, as with the undirected depth first search procedures used in FRAN, for example.

To summarize, a broad range of memory theories conceive of long term memory (LTM) as represented by a state space. Storage, or learning, involves either tagging items in a relational net, or constructing properties of items, which amounts formally to essentially the same thing since both involve transforming one relational net into a new one. During retrieval in most such models, search begins with the items or properties in STM. From there, a directed or undirected search is initiated until the to-be-recalled item is found, or failure results. At a formal level, most information processing accounts of problem solving have the same general form. In this case, the goal is to find a solution path from the given to a problem solution.

The psychological reality which state space theorists impute to their constructs is well summarized by Newell and Simon (1972):

> Human problem solving, we have argued, is to be understood by describing the task environment in which it takes place; the space the problem solver uses to represent the environment, the task, and the knowledge about it that he gradually accumulates; and the program the problem solver assembles for approaching the task /pp. 867-868/.

Limitations

Unfortunately, state space formulations (including Newell & Simon's use of production systems to represent search methods) have a number of important and fundamental limitations. Perhaps the most basic are those pertaining to individual differences in the formation of state spaces, and learning. Again quoting Newell and Simon(1972):

> Our emphasis has been on the problem solver's performance program . . . We brought to bear what evidence we could on the question of how the problem solver, in the face of a new task, generates an appropriate problem space and program and on the commonalities and differences among problem solvers. Our answers to these questions were sketchy, for these areas undoubtedly represent the largest and most important *terra incognita* on the map of the theory of human problem solving today /pp. 867-868/.

Although the importance of individual differences is well recognized, existing state space theories have little more to say about them than the fact that state spaces and processes may vary over individuals.

In dealing with individual differences, the state space theorist is posed with a dilemma. On the one hand, he may employ a separate state space for each subject together with individual processes characteristic of that subject. Such an approach, however, would be antithetical to science. Piaget, for one, has recognized this problem and it is primarily for this reason (e.g., see Furth, 1969) that he chose to deal with the epistemic subject, rather than the individual.

The alternative is to set up one state space to account for the behavior of all subjects (or, at least, for a given class of subjects), together with a fixed set of processes. In this case, however, the result will necessarily be a theory of averages. Such theories may provide convenient ways of explaining and perhaps predicting average performance of groups of individuals, but they cannot seriously be used to characterize individual processes (Scandura, 1971). Any viable memory theory that purports to deal with individual differences must distinguish between those characteristics which are common to all people and those which make them unique.

Existing theories not only fail to deal with individual differences in a substantive way but they tend to be geared to particular task environments. The model described by Bower (1972) deals with the free recall of unorganized lists while that of Rumelhart _et al_. (1972) was explicitly designed to deal with verbal organization. Both models probably reflect human memory of verbal material to some degree since people can obviously deal with both kinds of situation. Yet neither model by itself allows for this. In FRAN items are treated as wholes, at the same level of abstraction. By stressing properties of items, Rumelhart _et al_. get a somewhat more general state space but at the expense of more processing rationality (e.g., in forming general concepts) than is reasonable or necessary in many situations.

In state space formulations, it is also unclear what are the mechanisms by which state spaces are constructed in the first place. The executive interpretive system of Rumelhart et al. (1972) was designed for this purpose, but if it is so important (as it is), why was it kept separate from the memory theory itself? Equally important, the processes by which state spaces are modified have an ad hoc character that are also treated independently. Clearly, there are relationships between understanding, storing, learning, and searching for information. Exactly, how are these processes related? What are the differences? What do they have in common? With the exception of a fixed processing capacity assumption, state space theories are strangely silent on these matters.

The Structural Learning Theory

Introduction: Competence and the Idealized Theory

With these questions in mind, let us briefly review the structural learning theory (Scandura, 1973) as it pertains to cognition generally, and memory in particular.

The structural learning theory consists of three interrelated partial theories, each of which must be tested empirically in a different way. First, there is a theory of structured knowledge - or, more accurately as we shall see, theories of structured competence. These theories deal with the problem of how to characterize competence: the competence associated with particular behavior constitutes a theory in its own right. The second partial theory brings the behaving subject into the picture. It provides a basis (1) for determining the knowledge had by particular subjects (relative to a given theory of competence) and (2) for telling how that knowledge is selected for use and how new knowledge is acquired. This theory is an idealization in the sense that it applies only where the subject is unencumbered by memory and his finite capacity for processing information. The third theory is still more general and tells what happens when memory and information processing capacity are taken into account. These three theories build upon one another in a natural way, although research on any one can progress independently of the others.

The observer and the subject both play a fundamental role in the theory, corresponding to the above distinction between competence and knowledge. Competence involves rules introduced by an observer to account for behavior he is interested in observing. This behavior, or more exactly, this class of potentially observable input-output pairs, against which actual behavior is to be judged, is predetermined. When the psychologist enters his laboratory, for example, he has a pretty good idea ahead of time what stimuli and what responses he is interested in. Whether or not the subject wiggles in his chair as he elicits the response "MUR" may not only be unanticipated, but typically will also be ignored. Similarly, in testing students to see whether they know the subject matter, the professor can usually determine in advance what are the stimuli and the corresponding acceptable responses.

More important, the rule sets introduced by the observer to represent competence differ in an important way from standard competence theories in linguistics (e.g., Chomsky & Miller, 1963), and indeed, from the formal mathematical (production) systems (e.g., see Nelson, 1968) on which they are based.

A simple grammar, for example, consists of a finite set of rules, and is said to account for an input-output pair if some sequence of rules in the rule set can be found such that the successive application of these rules to the input generates the output. This latter point is particularly important because it is implicitly assumed that the rules must be combined in a very special way. In the structural learning theory rules are allowed to interact in a more general manner.

To see the difference, suppose, for example, that the given class of input-output pairs of interest consists of strings of the form $xB \rightarrow By$ where x is string of a's and y is the binary numeral representing the number of a's (e.g., $aaaaaB \rightarrow B101$, $aaB \rightarrow B10$). A simple grammar which accounts for this class, includes the rules $r_1 = xxBy \rightarrow xB0y$ and $r_2 = xxaBy \rightarrow xB1y$. To account for the pair $aaaaaB \rightarrow B101$, then, we see that

$$aaaaaB \xrightarrow{r_2} aaB1 \xrightarrow{r_1} aB01 \xrightarrow{r_2} B101.$$

Notice that neither of the given rules is sufficient in itself to account for the given pair. It is necessary to assume that the rules may be applied successively as many times as desired.

An equivalent way of accounting for this class is to explicitly include a generalized composition rule * in the characterizing rule set, call it $A = \{ r_1, r_2, * \}$. Accounting for a given input-output pair, in this case, means either that there is a rule in A which generates the output on application to the input or that such a rule may be derived by application of rules in A to other rules in A. More precisely, we say that A accounts for an input-output pair if there is a finite number n such that there is a rule in one of the following sets which generates the output from the input.

$$A = \{ r_1, r_2, * \}$$

$$A^2 = A \cup \{ r_1{*}r_1, r_1{*}r_2, r_2{*}r_1, r_2{*}r_2 \}$$

$$A^3 = A^2 \cup \{ r_1{*}r_1{*}r, r_1{*}r_2{*}r_1, r_1{*}r_2{*}r_2, \ldots \}$$

$$\vdots$$

$$A^n$$

With respect to the above instance $aaaaaB \rightarrow B101$, for example, the rule $r_2{*}r_1{*}r_2$ in A^3 serves this purpose.

It is important to emphasize that these two formulations are mathematically equivalent insofar as computing power is concerned, so in one sense we have nothing new. Mathematical equivalence, however, does not necessarily imply behavioral equivalence, or even as I would propose in this case, behavioral viability. One way to see this is to observe that the composition rule * is just one of any number of different higher order rules that might be included in a rule set. Such rules can greatly increase the power of a rule set. For example, the higher order rule

$$r_a \Rightarrow r_b$$

operates on rules involving a's and converts them into corresponding rules involving b's. Just this one rule doubles the power of the given rule set to include an equivalent set of input-output pairs where the inputs involve b's instead of a's. More important, every time a new rule involving a's is added to the rule set, we automatically get "free," because of this higher order rule, a corresponding rule involving b's.

In contrast with competence, the term "knowledge" refers to a potential for behavior. Knowledge also consists of rules, but these rules are attributed to a behaving subject and are thought of as generating behavior. Previous theories (e.g., see Piaget in Furth, 1969; Newell & Simon, 1972), in which rule like constructs are attributed directly to behaving subjects, have been essentially non-operational. The underlying mechanisms have been difficult if not impossible to test empirically. The Piagetian mechanisms of accomodation and assimilation, for example, are immune in an important sense to behavioral test because the effects of these mechanisms on behavior depend on the knowledge individual subjects have when they enter the learning or testing situation. But, Piagetian theory itself provides no way of finding out what this (individual) knowledge is.

The structural learning theory provides an explicit way of handling this problem. The rules introduced by an observer to account for the behavior of interest are used as an instrument of sorts with which to measure human knowledge. More specifically, the theory tells how, through a finite testing procedure, one can identify which parts of given rules in a competence theory individual subjects know - that is, which rules the subjects can perform in accordance with. The rules in a competence theory in a very real sense serve as rulers of measurement, and provide a basis for the operational definition of human knowledge. It should be noted in this regard that to have behavioral relevance, a rule set must reflect the common culture shared by the population in question.

To briefly review how this is accomplied (for details, see Scandura, 1973), we first note the basic assumption on which the theory rests is that people are goal directed information processers. Further, rules may be viewed as procedures in the sense of computer programs and may be characterized, for example, as flow diagrams or labeled directed graphs (see Scandura, 1973).

INSERT FIGURE 3 ABOUT HERE

Procedures can always be broken down into simple enough steps so that each subject in a given population is able to perform each step perfectly or not at all (cf. Suppes, 1969; Scandura, 1973). In short, each component step of a procedure may be assumed to act in atomic fashion. The behavioral reality of atomic rules has been established, in my opinion, beyond any reasonable doubt (e.g., see Scandura, 1969).

Since each component acts in atomic fashion, each path through a procedure also acts in atomic fashion. That is, each path through a procedure makes it possible to generate responses to a uniquely specified equivalence class of stimulus items, and to no others. Furthermore, there are only a finite number of such paths, since we do not distinguish paths according to the number of repetitions of loops. Collectively, these paths impose a partition on the domain of stimuli to which a procedure applies. This makes it possible to pin-point through a finite testing procedure exactly what it is that each subject knows relative to the initial procedure introduced by the observer. It is sufficient to test the subject on one item selected randomly from each equivalence class. Success on any one item, according to our assumptions, implies success on any other item drawn from the same equivalence class, and similarly for failure.

Knowledge (behavior potential), then, is also represented in terms of rules (procedures), specifically in terms of sub-portions of initial, corresponding competence procedures. It should be emphasized in this regard that the knowledge attributed to different individuals may vary even though only one rule of competence may be involved. The idea is directly comparable to measuring different distances with the same ruler.[4]

None of this is idle speculation. Scandura & Durnin (1973) and Durnin & Scandura (1973) have collected data involving a large number of different tasks, with subjects ranging from pre-school children to Ph.D. candidates. When run under carefully prescribed laboratory conditions, it was possible to predict performance on new items, given performance on initially selected items, with over 96% accuracy. When the testing took place under ordinary classroom conditions, where the subjects were run as a group, the predictions were accurate in about 84% of the cases.

The structural learning theory also provides a precise set of mechanisms by which the rules available to a subject are put to use, and by which new rules are acquired. The basic idea rests on the assumption that human beings are goal directed information processers, and that control shifts among various higher and lower level goals automatically in a predetermined manner, according to the requirements of the situation.

For present purposes, we may think of the mechanism informally, operating as follows: given a task (stimulus and goal) for which the subject does not have a solution rule immediately available, control is assumed to automatically switch to the higher level goal satisfied by rules which do apply. With the higher level goal in force, the subject assumably selects from among available and relevant higher order rules in the same way as he would with any other goal. In effect, if the subject has an applicable rule available, then he will use it. Where no such higher rules are available, the theory assumes that control moves to still higher level goals. Conversely, once a higher level goal has been satisfied, control is assumed to revert to the next lower level.

Assume, for example, that a subject is asked to convert 5 yards into inches, but that he does not know explicitly a rule for accomplishing this (e.g., he does not know that there are 36 inches in a yard). Let us assume, however, that he does know rules for converting yards into feet and feet into inches, together with a higher order rule which operates on pairs of rules such that the output of one serves as the input of the other and generates composite rules.

In this case, control would be assumed to shift to the higher level goal of finding a solution rule. According to the simple performance hypothesis, then, the higher order rule is applied to the yards to feet and feet to inches rules, generating a composite rule from yards to inches. This composite rule satisfies the higher level goal so control reverts to the original goal. Here the simple performance hypothesis is used once again and the composite rule is applied to solve the problem.[5]

Again, none of this is idle speculation. Several experiments (Scandura, 1971, 1973) rather conclusively demonstrate the viability of the analysis, at least under the limited conditions tested. One experiment (Scandura, 1973), for example, involved the composition higher order rule and simple rules for trading objects such as toothpicks for erasers. After training on the requisite simple rules, naive subjects were either trained or not on the higher order rule. Then, they were presented with new pairs of simple rules and tested on problems that required corresponding composite rules for their solution. Correct predictions in this experiment were made in 29 out of 30 individual cases.[6] In a somewhat more complex and demanding experiment (Scandura, 1973), each subject was required to generalize from a specific rule. Correct predictions were made in 50 of 50 cases.

Extension to Memory

In the idealized theory it is assumed essentially that the subject has a single active memory A, consisting of elements (degenerate rules), simple rules, and rules which operate on rules. The contents of this memory, including new elements which may be generated in the course of a computation, are assumed to be readily and uniformly available to the subject. The absence of *a priori* relations among the rules can be represented as in Figure 4A.

INSERT FIGURES 4A & 4B ABOUT HERE

Experiments have shown that this idealization can be approached in practice (e.g., Levine, 1966; Scandura, 1973) but, of course, this will not be the case in all empirical situations. Even familiar information is not always equally easy to recall; witness the "tip of the tongue" phenomenon. In general, at any given point in time some knowledge (rules) will be available (aroused) but other knowledge will not.

With this in mind, Scandura (1973) extended the idealized theory by distinguishing between a long term memory (M), consisting of a cumulative record of all active elements, and that small part of it (A) which is active at any one time (see Fig. 4B). M is a finite set of rules as before; but only rules and (encoded) stimuli in A can generate responses or produce new knowledge. All processing goes on in A.

In developing the theory, Scandura (1973) found it convenient to distinguish between memory theory where the capacity of A is finite but unbounded and where the capacity of A is fixed. Clearly, the memory theory with unlimited processing capacity is more broadly applicable than the idealized (memory free) theory. In particular, the theory applies in situations where certain rules are not immediately available in A, even though the subjects may have previously learned and stored them (in M). In testing the theory, the only essential condition is that the subject not be hampered by his limited capacity for processing information. This can be accomplished, for example, by providing the subject with a pencil and paper, and all the time he needs.

In the theory, stimulation from the environment that enters A automatically becomes part of M. This information remains immediately available to the subject, however, only as long as it remains in A. It can be retrieved at a later time only if it has been stored (via rules) in relation to other information which can serve to cue it. Specifically, storing information involves constructing rules by which to-be-remembered elements can be generated from other elements (that are either given as cues or in A). Retrieving information involves using active rules to generate observables from given cues and elements in A (or in the environment).

The basic mechanisms of the memory theory with unlimited processing capacity are a direct extension of those for the idealized theory. In retrieval, for example, control may shift among goal levels as before. The relatively small number of rules in A, however, serves to keep within strict bounds the number of rules that must be tested at each stage. Where desired rules cannot be derived or retrieved solely from rules in A, or in the environment, control shifts so as to activate (i.e., derive or retrieve) rules which do make this possible. For example, it is reasonable to assume that some of the rules needed in a derivation, particularly those on which a given rule might operate, may not be active (in A). In this case, the mechanism allows control to shift automatically to what are called domain goals. Domain goals are satisfied by rules in the domains of corresponding available (higher order) rules. Once needed domain information is activated, through derivation or retrieval, control returns to the goal from which the secondary domain mechanism was initiated, and the process continues.

To date, only one series of experiments has been run to test the memory theory with unlimited capacity. This research was concerned with the behavior of individual subjects in particular situations, and involved a demanding new paradigm in which a major task was to insure that the experimental conditions accurately and completely reflected the proposed theoretical requirements. In each experiment the overall results strongly supported the theoretical mechanism under study. There were some minor perturbations in the data of a few individuals in the earlier experiments, however, which led to methodological refinements that were required by the theory but had originally been overlooked.

In Experiment I, the author's assistant made certain modifications in the procedure that were not caught until after 32 subjects had been run. Since they appeared to be minor and could very easily be made by anyone running such an experiment for the first time, it is instructive to consider this experiment in detail.[7]

Experiment I

Method

Materials. The experimental material was similar to that used in an experiment by Scandura and Ackler (reported in Scandura, 1973). These materials consisted of sets of small items such as paper clips and rubber bands which were used in making trades with the experimenter. In addition, there were cards each of which described a rule for trading n stimulus objects for n + m or n - m response objects. On the back of each card was a symbol designated as the "name" of the card. There was also a chart which could be used for locating rule cards by name for making specific trades.

The cards were used to designate two kinds of rules, simple and composite. Simple rules affected trades directly and were represented on 5 x 8 inch cards. The card at the top of Figure 5 designates a simple rule which maps n paper clips into n + 1 blue chips.

INSERT FIGURE 5 HERE

The card at the bottom of Figure 5 designates a composite rule for changing pencils into paper clips, and then paper clips into white chips.

A pair of simple rules is said to be compatible if the output of one matches the input of the other. Compatible rules can be combined to form composite rules. For example, the rules n pencils n + 2 paper clips ($A \to B$) and n paper clips $\to$ n + 1 white chips ($B \to C$) can be combined to form a composite rule ($A \to B \to C$) which maps n pencils into n + 3 white chips. The set of compatible simple rules comprises the domain of a higher order composition rule which maps such pairs into corresponding composite rules.

The chart was a 9 x 9 table in which the entries were names of rules for converting row elements (e.g., pencils) into column elements (e.g., paper clips). The main diagonal and all entries below and to the left of it were blank. Thus, no element could be traded for itself, and no rule had an inverse. For example, there was a rule for trading pencils for paper clips but none for trading paper clips for pencils. The chart could be used to identify rules not immediately available. (This corresponded to retrieval).

Tasks. There were four different kinds of tasks. The first were direct trading tasks in which the subject was given a simple or composite rule card and a set of stimulus objects. He was asked to make the trade indicated on the card.

The higher order rule was used to define a second higher order (H) task in which the subject was presented with a compatible pair of simple A→B and B→C rules and a set of stimulus objects A. The goal was to trade the given A objects for the output (C) of the simple B→C rule (that did not involve the A objects). This could be accomplished by first deriving the necessary composite rule and then applying it. The composite rules (cards) could be derived by rearranging the given pairs of compatible simple rule cards to form composite ones.

In the third, domain (D), task the subject was given a simple B→C rule (e.g., n paper clips → n + 1 white chips) and a set of A items not represented on the card (e.g., pencils). Furthermore, the column on the chart which corresponded to the rule output C (e.g., white chips) was covered so that it was not possible to name the rule which converted pencils directly into white chips. The goal was to find a pair of compatible rules (including the given rule) in the domain of the higher order composition rule. The inputs of the derived simple rule had to match the given A objects (e.g., pencils), and the outputs had to match the inputs (B) of the given card (e.g., paper clips). For example, given pencils and the rule n paper clips $\rightarrow$ n + 1 white chips, the subject had to use the chart to find the rule which traded pencils for paper clips. The (domain) rule for accomplishing this involved locating the row and column of the table corresponding respectively to the domain and range of the desired rule. The entry in each row and column was the name of the indicated rule.

The fourth task was a composite HD task. The stimulus situation was identical to that in the domain task, but the goal, for example, was to trade pencils (A) for white chips (C). This task could be solved by identifying the composition higher order rule as adequate, retrieving the needed $A \rightarrow B$ domain rule from the table, applying the higher order rule to the $A \rightarrow B$ and $B \rightarrow C$ rules to form the needed composite rule, and finally applying the composite rule to solve the problem.

Subjects, design, and procedures. The subjects were 32 elementary school children in second through sixth grades at the Belmont Hills and Lea Elementary Schools in Lower Merion, Pa. and West Philadelphia, respectively. The experiment was conducted with individual subjects in two separate sessions, usually a day apart. At the end of each session the children were offered a balloon.

The first session consisted of pretraining and a Transfer Pretest. The subject was told he was going to play a trading game with the experimenter. He was then taught how to interpret the rule cards and to make trades using the rules represented by these cards. The experimenter pointed out that each rule card had a name printed on the back and that the names of all of the rules were on the chart, but he did not indicate how the chart was used. The subject was shown a rule and told, for example, "This is rule M; it is on our chart. Rule M tells us, no matter how many paper clips I give you, you must give me the same number of blue chips plus one."

The experimenter initiated a number of trades requiring use of the simple rule, providing assistance where necessary until the subject reached a criterion of three consecutive successful trades. The experimenter then gave the subject a set of objects not in the domain of the rule and asked the subject, "Can you use this rule to trade these pipe cleaners [for example] for blue chips?" Regardless of the subject's response, the experimenter emphasized that the rule could be used only to trade paper clips for blue chips.

The experimenter then showed the subject a different rule card and asked him to interpret the rule providing assistance if necessary. Before moving to the next card, the subject was required to use the rule to make three successful trades. This process continued until the subject reached a criterion of three successful trades without assistance, using three different, consecutive rule cards. At this point, it was assumed that when presented with a simple rule card, the subject could apply the corresponding rule.

Next the subject was taught in a similar manner how to interpret and use the composite rules. In the case of the composite rule shown in Figure 5, the subject was told, "Here is a rule for trading pencils for white chips. This rule says that no matter how many pencils I give you, you must give me the same number of paper clips plus two. Then no matter how many paper clips you have, you give me the same number of white chips plus one." As before, the subject was required to perform three consecutive correct trades with each composite rule. Training continued until the subject correctly interpreted three different consecutive composite rule cards. Then counter-examples were given. The subject was shown a composite rule and given a set of stimulus objects not in the domain or was given appropriate stimulus objects and asked to trade for objects not in the range. Throughout the pretraining, the subjects were always told when they were right.

The subject was then given a Transfer Pretest consisting of three tasks: a higher order (H) task, a domain (D) task, and a composite (HD) task. The order of testing was H, D, HD.

On the H task, the subject first was presented with cards representing a pair of compatible rules. The subject had not seen either rule before, but it was assumed, by virtue of his earlier training, that he knew what the cards meant. The subject was told that he could use the simple rule cards and that he might "move" them but he was not shown how to do so. Then the subject was asked to make three trades requiring the use of the corresponding composite rule. (He was never shown this rule directly either before or after testing.) For example, a subject who was presented with the ($A \rightarrow B$, $B \rightarrow C$) rules "n loose leaf reinforcers → n + 3 paper clips," and "n paper clips → n + 1 gummed labels" would be presented in turn with various numbers of reinforcers (e.g., A = 6, 8, and 5) and asked for the appropriate numbers of labels (C). If he made three successful consecutive trades he was rated competent. If he failed on any one presentation he was rated incompetent and the task was not repeated. However, if the subject clearly applied the higher order rule but made an error in counting, the experimenter warned the subject to be very careful, and presented the task again.

On the D task, the subject was presented with a card representing a simple $B \rightarrow C$ rule (e.g., "n paper fasteners → n + 4 rubber bands") and a set of stimulus A objects not in its domain (e.g., pipe cleaners). The C (i.e., rubber bands) column of the table was covered so it was not possible to find the rule converting A objects to C objects (i.e., pipe cleaners to rubber bands) directly. The subject was told, "I want to trade pipe cleaners for rubber bands. We need a pair of rules to let us do that. One of them is going to be this rule. Can you tell me what other rule I need so I can trade the pipe cleaners for the rubber bands?" The experimenter also emphasized that the subject could use the chart (but no training was given on it). If the subject responded correctly, the stimulus object was changed, the rule remained the same, and the task was repeated. If the subject responded correctly three times in a row he was rated competent. Otherwise he was rated incompetent on the D task.

Finally, the subject was tested on the HD task. He was reminded first of what rules he knew or had learned up to that point. For example, if the subject had learned the H and D rules, the experimenter might say, "In this problem you may use all the rules you have learned. You have learned to make trades. You have learned how to rearrange the cards to make a rule. You have learned how to use the chart to find a rule you need. If you need a rule you don't have, you can ask me for it and I will give it to you." These reminders were repeated during the testing if necessary.

Then the subject was given a simple rule and a set of stimulus objects not in its domain, and asked to trade the stimulus objects for the output of the given rule. In order to succeed, the subject had to ask for the necessary card, combine it (perhaps mentally) with the given one, and make the trade. (During testing some subjects misdefined the problem and tried to trade the stimulus A objects for the C output of the given rule by using the given $B \rightarrow C$ card. When this happened, the experimenter drew the subject's attention to the fact that the stimulus objects were not in the domain of the given rule. If the subject asked the experimenter for the wrong rule, he was allowed to choose again, if he wished.) The criterion for competence on the HD task was three correct trades. No reinforcement was given during the Pretest or the Posttests.

The 24 subjects who had failed any of the Pretest tasks, participated in the second session during which training was given on the higher order composition H task and the domain D task. Twenty-two of the subjects(H-D group) were trained first on the H task and then given Transfer Posttest I which was identical to the Pretest. Next, they were trained on the domain task (i.e., on how to use the chart) and given Transfer Posttest II. Later two subjects(D-H group) were trained first on the domain rule. On the Posttests, only the simple rules and stimulus items were changed. All subjects were given both H and D training, even if they were already competent on a corresponding H or D task. No subject was trained on the HD task.

In training on the higher order (H) task, the subject was shown a pair of compatible (A→B, B→C) rules and a set of stimulus A objects. The experimenter demonstrated how the rules could be combined by sliding the simple rules together in the appropriate manner. The subject was then asked to interpret the newly formed A→B→C rule. The A→B and B→C rules were separated again and the subject was given a new set of stimulus A objects and asked to actually trade for C objects. This was repeated until the subject had successfully performed three consecutive trades with the given rule pair. Then new pairs of rules were introduced until the subject made three successful trades with three consecutive, different pairs in a row. Counter examples were then given. Sometimes the subject was given a compatible pair of rules and asked to trade an element not in either domain, or to produce an element not in the range of a rule. Or, the pair presented was not compatible and the subject had to indicate that two rules could only be combined if the output of one matched the input of the other. In this case, the experimenter emphasized that the higher order rule only applied to pairs of rules. In actually solving the problems the subjects were not forced to slide the simple rule cards together if they preferred not to.

In the domain training, the subject was given a simple B→C rule, for example, "n paper clips → n + 2 white chips" and a set of stimulus A objects not in its domain (e.g., pencils). The C (i.e., white chips) column in the table was covered. The subject was told, "I want to trade pencils (A) for white chips (C). We can't do it just with this card so we need a pair of cards. One of them will be the (B→C) card we have here /pointing/. Now we're going to see how we can use the chart to find the other (A→B) card." The subject was then taught how to find the (A→B) rule for trading pencils for paper clips. After a subject retrieved an A→B rule by using the chart, the experimenter held the rule against the given B→C rule that the subject could see that the output of the former matched the input of the latter. Similar tasks involving different simple rules and stimulus objects were presented until the subject was successful at finding the missing rule on three consecutive tasks. The experimenter always emphasized that a pair of rules was necessary to solve the problem, even though one of the rules was given. The question was always phrased, "What pair of rules do I need . . ."

Results and Discussion

Of the 32 subjects given the transfer pretest, eight were successful on the H, D, and HD tasks. Ten of the remaining 24 failed on all three Transfer Pretests; the other 14 succeeded only on the H Pretest.

After training on the H task, all of the first 22 (of the 24) subjects succeeded on the H task on Transfer Posttest I. Except for one of the 22 subjects who also succeeded on both the D and HD tasks, they all failed on the D and HD tasks. After subsequent training on the D task, all 22 subjects not only succeeded on the H and D tasks on Transfer Posttest II but they also succeeded on task HD on Transfer Posttest II.

Of the two remaining subjects (in the D-H training group), one failed on all three Transfer Pretests and the other succeeded only on the H pretest. Both subjects succeeded on all three tasks on Transfer Posttest I, after D training, and did so again after the subsequent H training. These results are summarized in Table 1. In the table, "+" indicates that the subject reached criterion and "-" that he did not. Subjects are identified according to age (8, 9, 10, 11, 12) and sex (B, G).

INSERT TABLE 1 ABOUT HERE

With the exception of two of 48 posttests, all of these results are consistent with the theory. Before H and D training, there was no basis for predicting performance on the H and D tasks because there was no way of knowing whether or not the subject had already mastered the requisite rules. The only restriction on the Pretest is that a subject who knows both of the higher order H and D rules should, according to the theory, not only be successful on the H and D tasks, but also on the HD task. The data support this prediction in eight of eight cases.

The same pattern was obtained after the H and D training. In only one of 22 cases did H training lead to success on D Posttest I, and here the subject was also successful on the HD task as would be expected. Although only two subjects were given the D training first, the data suggest an overlap between D training and performance on the H task (as well as on the D task). As before, as well as throughout Transfer Posttest II, success on the H and D tasks was followed by success on the HD task.

All in all, one might be tempted to report strong support for the proposed mechanism. This would be inappropriate, however, because the H training inadvertently was not restricted to the higher order composition rule. The subjects were not only taught how to generate composite rules from pairs of compatible simple rules, but also to use the derived composite rules to solve H tasks. In effect, they were taught both the H rule and the (idealized) mechanism itself, thereby leaving unanswered the question of whether the mechanism itself is innate. This in itself, however, was not a serious problem since the innateness of the idealized mechanism had been tested previously (Scandura, 1973).

The problem came in interpreting performance on the HD task. Our original intent was to determine whether training on a B→C rule, the H rule, and the D rule was sufficient for solving an A→C problem. Assuming that the subject is capable of evaluating the lower and higher level goals involved (See Scandura, 1973, Ch. 9 - especially pp. 287 & 294), the postulated (enriched) mechanism is sufficient for this purpose. Given a set of A objects and the goal of finding the appropriate number of C objects to trade, control would be assumed to go to the higher level goal consisting of rules which apply to A objects and generate C objects. The only available rule for accomplishing this is the higher order H rule. But, the H rule only operates on pairs of compatible simple rules. Hence, control is assumed to go to the domain goal consisting of such pairs. In this case, the only adequate and available rule is the higher order D rule (which applies in situations where a B→C rule and A objects are available). Since the necessary domain elements are available, the rule is applied and a compatible pair of A→B, B→C rules is generated. This pair satisfies the domain goal so control goes to the original higher level goal. This time the H rule is applied and a composite A→B→C is generated. Since the higher level goal is satisfied control goes to the original goal, the composite A→B→C rule is applied and the problem is solved.

Unfortunately, this is not the only reasonable way to account for the HD task results. Because of the nature of the H training, it is just as reasonable, perhaps more so, to assume that the HD task was solved by composing the rules actually taught during the D and H training. To see this, notice that application of the D-rule taught during D training generates the needed A→B, B→C pair. Subsequent application of the combined H-rule and "mechanism" taught during H training, then, not only generates the composite A→B→C rule but also applies it to solve the problem. (It is important to recognize in this regard that generating an A→B→C rule followed by its use is not equivalent to a composition of rules.) Thus, assuming that a subject who has been given the H training also knows the composition H-rule (separate from the mechanism), which seems reasonable with the subjects used, success on the HD task can be explained as follows: After control goes to the higher level goal, the composition H rule is applied to the D rule and the combined H rule and mechanism forming their composition. (Notice that the two rules are compatible since the output of the former serves as input for the latter.) The resulting composite rule satisfies the higher level goal so control reverts to the original goal, the resulting composite rule is applied, and the task solved.

Experiment II

Experiment II was designed to eliminate the second interpretation as a viable alternative.

Method

The materials, tasks, and procedures in Experiment II were identical to those of Experiment I except in training on the higher order rule and in stating the subject's goal in the domain rule training.

Instead of training on the higher order transfer task itself (i.e., training on a rule for solving such tasks), higher order rule training in Experiment II was limited to the higher order H rule for forming composite rules from compatible pairs of simple rules. During training, the subject was first shown two compatible simple A→B, B→C rules and a set of stimulus A objects, corresponding to the inputs of the A→B rule. His goal was to find an A→B→C rule for trading the given A objects for C objects. The experimenter demonstrated how the simple rules could be combined by sliding them together in the appropriate manner. Then, the rules were separated and the subject was given a new set of stimulus A objects and asked to construct a composite rule. The subject did not perform any trades with the rule, as in Experiment I. This was repeated with other pairs of simple rules until the subject was able to form appropriate composite rules when this was possible, or to identify the simple rules as incompatible. In addition, the subject was sometimes given a pair of compatible A'→B', B'→C' rules where A'≠ A and C'≠ C, and, if necessary, instructed why the corresponding A'→B'→C' rule could not be used to trade A objects for C objects.

On the domain task and training, the subject was given a simple B→C rule (e.g., n gummed labels → n-2 rubber bands), and a set of A items (e.g., toothpicks) as before. Also, the C column on the chart which corresponded to the rule output (i.e., rubber bands) was covered so that it was not possible to name the rule which converted A objects directly into C objects.

The way in which the goal was stated, however, was changed. No reference was made to finding a pair of compatible rules for making trades. Rather, the subject's stated task was to find a pair of rules in which the outputs (B) of one (A→B) were the same as the inputs (B) of the other (B→C) and the inputs (A) of the first were identical to the given A objects and the outputs (C) of the second were identical to the outputs of the given B→C rule. One of the two rules, of course, was always the given rule. (To help insure that the task was understood, this was explicitly stated only on the Pretests.) For example, given toothpicks and the rule, n gummed labels → n-2 rubber bands, the subject had to find the A→B rule which converted toothpicks into gummed labels, and indicate that the given rule (n gummed labels → n-2 rubber bands) was the other rule.

During D training, the subject was shown how to locate needed rules by name in the row and column of the table corresponding, respectively, "to the inputs and outputs of the desired rule." (The words "input" and "output" were explained to the subject during the pretraining, and the subjects were taught to identify them on the cards.) Subjects were taught to identify the given rule as one member of the needed pair by giving the name on the back of the rule card. In short, the domain instruction involved using the chart to find compatible pairs of rules but references to the possible use of the rules in making trades were eliminated.

Since the results of Experiment I suggested that D rule training may influence H test performance, second grade subjects (aged 7-8) were used in the D-H group because they would presumably be more sensitive to inadequacies in wording and treatments. The first four D-H subjects were trained on the D task as in Experiment I. The seven other D-H subjects were trained on this task as described above. In addition, there was an H-D group consisting of 10 fourth graders (aged 8-10). These subjects all received the modified H and D training described above.

One of the second graders, a 7 year old boy, was unable to complete the pretraining successfully, and was not included in the experimental comparison. Five of the younger subjects (aged 7 or 8) were unable to complete the experiment in two sessions. In these cases the experiment was spread out over three or more sessions. The length and content of the sessions varied with each subject's attention span and rate of progress. A typical subject might participate in four one-hour sessions with the first consisting of pretraining, the second of further pretraining and the Pretest, the third a review, domain training and Posttest I, and the fourth another review, higher order composition training and Posttest II.

The mean time spent on subjects in the H-D training group was two hours and thirty-five minutes. The shortest time was one hour and fifty-five minutes. The longest, four hours and forty minutes. For the somewhat younger D-H group, the mean time per subject was three hours and forty-five minutes. The shortest time was two hours and twenty-five minutes; the longest, seven hours and fifty-five minutes.

Results and Discussion

The results of the H-D subjects closely paralleled those of Experiment I. After training on the H and D tasks, all 10 subjects on Posttest II not only succeeded on the H and D tasks, but on the HD task as well. Also as expected, H training improved performance only with those 2 subjects who failed on the H Pretest. In no case did H training transfer to success on the D task.

In effect, these results clearly tend to discount the alternative explanation of the Experiment I data lending further support for the proposed (enriched) theoretical mechanism.

The results of the younger D-H subjects, however, were less clear. In 2 or 3 cases (one was a second administration of the experiment to one subject), D training led not only to success on the D task, but also on the H task. Closer scrutiny of our experimental method during D training and testing indicated one possible source of difficulty. In many cases, the experimenter inadvertently showed the subjects how the A→B rules, once retrieved, matched the B→C rules. The very process of showing how the rules matched effectively amounted to instruction on how to form composite rules. It is not therefore surprising that some of the subjects were able to solve the H transfer tasks after D training but not before. It should be noted that this activity by the experimenter was not prescribed by the instructions provided but evolved naturally in the course of attempting to explain rather complex ideas to young and generally untalented children. The fact that this took place was determined by the analysis of the tapes of the experimental sessions by the author and the experimental assistant.

There were also some additional anomalies with these young subjects that were observed for the first time. On four occasions (one subject twice), subjects succeeded on the first H task on the Pretest or Posttest but failed when the number of stimulus inputs to be traded was changed (indicated +?). Why this was so is not clear but it appeared likely that it was due to idiosyncratic features of the particular rules used and/or the subjects themselves. For example, one seven year old boy traded correctly on the first H presentation, but did so without moving the given A→B and B→C cards together. On the second presentation the subject appeared confused, as if he interpreted the repetition of the problem as an indication that he had traded incorrectly on the first presentation. On the second presentation he traded the given A objects for C objects using the B→C card, then traded the C objects for B objects using the A→B card. The stimulus objects were changed several times and on each presentation the subject was admonished to be very careful without positive effect. Also one subject succeeded on both the H and D tasks but failed on the HD task on Posttest I, and another subject (the one who went through the experiment twice) performed similarly again during a second administration of the experiment on the Pretest.

If these latter results are due to other than idiosyncracies, they have important implications for the structural learning theory with unbounded capacity, at least as applied to younger children. It should perhaps be noted in this regard that our D-H children were of the educationally underprivileged variety and had considerable difficulty in learning the material. Attention at times was also a significant problem. These factors suggest that the unlimited capacity assumption may not have been realized in some as yet well understood way. Among the factors that may have been operating are: These subjects may be unable to identify information even when it is readily available in the environment (i.e., they may lack even basic searching skills); another possibility is that the long training required may be indicative of the greater load placed on active memory by the higher order rules which had to be remembered. That is, the younger subjects may have had to remember them in terms of a larger number of chunks than older children, thereby exceeding their processing capacity. Nonetheless, rather than attempt to unravel this complicated set of results in the present series of experiments, I decided to determine first whether it would be possible to further separate D and H training.

Experiment III

Method

The method used in Experiment III was identical to that used in Experiment II, except in the domain rule training. The stimulus situation and statement of the task were unchanged. After a subject had used the chart to retrieve an A→B domain rule, however, the retrieved rule was never held against the given B→C rule. The experimenter did, as before, draw the subject's attention to the fact that the input of the retrieved rule matched the stimulus A objects, and that the output of the retrieved rule matched the input of the given rule. But this was done in a manner that did not reveal how to form the composite A→B→C rule.

The scoring criteria for the H task also were modified slightly. If a subject succeeded on the A→B→C problem but failed on another, he was presented with an entirely new problem and allowed to try again. (This apparently helped to avoid the ambiguous scoring problem observed in Experiment II.)

After four +-- subjects (i.e., subjects who solved the H task and failed the others on the pretest) and four --- subjects had been run, a minor change was made in the procedure. Since four of the subjects on the HD task attempted first to use the given B→C rule to trade A objects for C objects, the subjects were reminded just prior to the HD task that the given rule could only be used to trade B objects for C objects. (All subjects had received such training earlier.)

All fifteen subjects were second and third graders from the Lea Elementary School between the ages of 7 and 9. All of the subjects were trained on the D rule first. Fourteen of the subjects required more than two sessions in order to complete the experiment. The average time per subject was four hours and twenty minutes. The shortest time taken to complete the experiment was two hours and fifteen minutes. The longest time taken was nine hours.

Results and Discussion

The results are summarized in Table 3.

INSERT TABLE 3 ABOUT HERE

After success on both the H and D tasks, all but one of the +-- subjects succeeded uniformly on the HD task. The one exception was a nine year old girl who failed the HD task on Posttest I and succeeded on Posttest II only after receiving special help. On these tasks, she appeared to guess cards randomly and then reject them because they did not "help." On Posttest II when the experimenter finally asked, "What would help?", she indicated that an A→C card would (help) but that the C column on the chart was covered. When asked, "Can it be anything else?" she said she could use an A→B card but did not proceed to look for it. After restating the problem, the subject began to guess cards randomly again. She was then asked, "What were the things [rule cards] you told me would help?" She then looked on the chart for the A→B card and solved the HD problem. Although these results can be interpreted in several ways, it is possible that the original search for an A→C rule was guided by a prelearned selection rule (Scandura, 1973, Ch. 9). Subsequent rejection would tend to have imposed a greater load on working memory perhaps, as the above summary of events suggests, making it difficult for the subject to keep in mind the original goal.

Four of six --- subjects also performed as predicted on the HD task.[8] The other two subjects, however, failed (only) on the HD task on Posttest II. One, a seven year old boy, seemed to have no idea of how to proceed. After a number of apparently random guesses, he became discouraged and unable to concentrate. The other, an eight year old girl, moved through the pretraining relatively quickly but on the HD task refused to investigate any other possibility after finding that she could not trade the given A objects directly for C objects. In this sense, she seemed like the nine year old girl mentioned above.

General Discussion

Overall, these results suggest that the enriched mechanism proposed is a common characteristic of all people. Once a subject knew, or had been taught, the appropriate higher order H and domain D rules, he typically not only could solve the corresponding higher order H (and domain) tasks but he was successful on the even more demanding combined HD task as well. Success on this task cannot easily be explained in terms of the idealized (memory free) theory but does seem compatible with the extended mechanism proposed.

The latter theory also has intrinsic support. The enriched mechanism is a natural extension of that on which the idealized theory is based; the idealized theory has been tested more extensively. What goes under the rubric of memory can be handled by extending the basic mechanism of the idealized theory only slightly to allow for the generation of domain elements. Moreover, these mechanisms provide a highly general framework within which a wide variety of disparate phenomena can be viewed such as simple performance, problem solving, learning, and motivation, not to mention memory (for details, see Scandura, 1973). There is also some evidence to suggest that essentially the same theory can be extended to deal with perceptual and developmental phenomena as well (Scandura, 1973, Ch. 5). The basic mechanisms of the theory appear to be at least as simple as those required for most existing memory theories and, yet, the theory potentially has greater generality.

Although the basic mechanism appears generally compatible with existing memory data, a major limitation is that no serious attempt has been made to date to make direct contact with a large body of memory research.[9] Nonetheless, in accord with known facts, for example, it follows directly that degree of recall should depend on the extent to which the test conditions reinstate the stimulus conditions during storage. According to Scandura (1973), however, the distinction between goals and stimuli provides a basis for making finer experimental distinctions with complex materials than for the most part has been possible to date. According to the theoretical mechanism proposed, it also is immediately obvious why a rule that has been used in the immediate past is more likely to be used than some alternative, even where as in Einstellung experiments, the alternative would otherwise be preferable. Rules in A are applied before rules which must be derived from rules in A. Similar general comments can be made concerning retroactive inhibition and reminiscence, as the theory clearly provides a basis for learning between storage and retrieval. In general, this new learning may either interfere with or facilitate retention.

A major limitation of this research, however, is that it does not take into account what might happen in those cases where a subject is unable to use provided information in the environment effectively or where his capacity to process information is exceeded.

Fixed Processing Capacity

The fixed capacity theory follows Miller (1956) in assuming that A contains 7 $\pm$ 2 "chunks" of information, only here the chunks may be rules as well as elements on which rules operate. The mechanisms of the memory theory with unlimited processing capacity are limited in this regard in that they can serve only to make more rules active. In the fixed capacity theory, mechanisms are also needed to explain how information is deactivated. Although little relevant data are available at this time, it would appear that there are two basic ways in which deactivation might enter the theory: (1) by modifying the basic mechanisms so as to allow for deactivation of goals and rules as well as their activation, and (2) by modifying the rule notion itself so that elements may be "erased" as well as generated. The basic constraint in either case, according to the theory, is the fixed finite capacity of A.

Roughly speaking, Scandura (1973) proposed the hypothesis that goals are also included in A and are deactivated during a learning episode when they become no longer useful. For example, it was assumed that any initial goal would remain in A throughout the course of a derivation because it is always the last as well as the first goal in control. Higher level goals, however, are discharged from A as control reverts to lower levels. Retaining them in A after this serves no critical purpose.

Specifying how goals become active and are deactivated still leaves open what happens when a rule becomes overloaded in the course of a computation. In this case, Scandura (1973) rejected such universal assumptions as: The element in A that is processed first is always dropped first, and proposed instead to add more structure to rules so that they might serve not only to activate (generate) new elements during the course of a computation, but also to deactivate (erase) others. In effect, it was hypothesized that rules might specify not only what is to be done at each stage of a computation but also where each generated element is to be located in A. The placement of a new element in a given location is assumed to erase present contents, much as is the case in abstract automata (Nelson, 1968).

In short, it was proposed in a computation not only that elements can be generated, and thereby added to A, but also that they can be erased in a specified manner. A similar principle of erasure was assumed to apply to the shifting of control among goal levels. Elements, whether they be simple elements, rules, or goals, are assumed to remain in A if and only if they are needed to determine either a future output or an operation that must be performed sometime in the future. For example, let $A = \{S_o, r_n, r_m, o, G\}$, where S_o is a stimulus, r_n and r_m are compatible rules, o is composition and G is the goal. Assume further that $r_n o r_m(S_o)$ satisfies G but not $r_n(S_o)$ or $r_m(S_o)$. In this case, control shifts to G^2 so that now $A = \{S_o, r_n, r_m, o, G, G^2\}$. If A becomes overloaded at this point, something crucial must be erased, but the theory does not specify what. Here, o is applied to $\langle r_n, r_m \rangle$ generating $r_m o r_n$. This time, however, instead of just adding $r_m o r_n$ to A, r_n and r_m may be erased, as they are no longer needed. Similarly, once control reverts to G, G^2 is erased, leaving "more space" for the application of $r_m or_n$.[10]

The theory also includes explicit procedures for determining, in an analytical manner, the memory load imposed by individual processes (rules) as applied to particular instances. Data collected by Voorhies and Scandura (some of which is reported in Scandura, 1973) supports the viability of this method. These data are consistent with the notion that each subject has a fixed finite capacity for processing information. Although processing efficiency depends on the rule used - an extension of Miller's (1956) finding, the basic, physiologically determined processing capacity remains fixed.

Incidentally, the theory treats rehearsal as any other procedure. The data obtained by Voorhies and Scandura strongly suggest that rehearsal in and of itself has no effect on retention. Unless precautions are taken (e.g., Scandura, 1973; Dalrymple-Alford, 1967), however, rehearsal provides opportunities for chunking and thereby may give the appearance of improving retention. It should be noted in this regard, that "chunking" so-called involves processes over and above rehearsal itself, and strictly speaking is not the same (rule) as pure rehearsal.

The Structural Theory of Memory and

State Space Formulations - Contrasted

There are at least four major differences between the two formulations. First, competence in the structural learning theory consists of a finite set of discrete rules. There is no structure to the set of rules itself. The structure, if it can be called that, is imposed by the fixed manner in which the rules are allowed to interact. In state space theories, on the other hand, competence corresponds to a highly structured, fixed network.

Generative grammars provide a convenient way of conceptualizing the relationship between structural competence and state space formulations. A generative grammar, recall, also consists of a set of rules, but these rules may interact only in a very special way. Namely, they must be applied in sequence to successive outputs. State spaces provide a convenient way of representing the possible ways in which the rules (operators) in a generative grammar may be combined. State spaces are not adequate for representing structural competence because rules may be combined and otherwise modified in ways quite different from simple composition.

In effect, it would appear that competence in the structural learning theory is both more general and more constrained. It is more general because of the great variety of higher order rules which are possible. It is more constrained in that the rules are designed to reflect the knowledge had by a given culture or population of subjects, rather than to represent all possibilities.

Unfortunately, the question of how to actually construct a structural competence theory or a state space has barely been touched, even in formal treatments within computer science. In the latter sphere, for example, the selection of a state space has important implications for the search effort required to achieve goals (or retrieve information). Some progress has been made in the problem of description of states and operators (Amarel, 1968) but the processes by which "good" state spaces are devised are very poorly understood. Similarly, in the structural learning theory, we know that the rules must reflect the culture of the population of the subjects in question. Other than general statements to this effect, however, relatively little is known about the specific relationships which must exist between particular populations and rules.

Another facet of problem formulation is handled quite differently in the two formulations - namely, that of forming sub-goals. To date this question seems to have been considered only in state space models, and there only in problem solving (see Newell & Simon, 1972). In state space theories, sub-goals are represented in the state space itself, by means of what are called AND/OR graphs (state spaces) (see Nilsson, 1971). In the structural learning formulation, sub-goals are hypothesized to result from the way in which problems are interpreted (see Scandura, 1973, p. 348). Presenting a subject with a problem statement, for example, is almost universally understood to mean that the subject first is to define the problem - interpret the statement (sub-goal one), and then to solve it (sub-goal two). Defining the problem may involve generating a series of sub-goals, each of which presumably defines a task to be dealt with according to the mechanisms described above. Perhaps surprisingly, interpretation (assigning meanings) in the structural learning theory seems not to require any new mechanisms (see Scandura, 1973, Ch. 7). There is, however, as yet relatively little data relating to this hypothesis.

The second major difference concerns the distinction between competence and knowledge in the structural learning theory. This distinction, in which the knowledge had by individual subjects is defined in terms of competence and the subjects' behavior, has important implications for individual differences. In contrast to the finite, systematic testing procedure provided in the structural learning theory, the only way individual differences can be treated in existing state space theories is by devising separate state spaces and processes for different individuals. Indeed, the distinction between the structure of input information (competence) and the structure of the subject's knowledge has barely been considered (cf. Frederickson, 1972).

Third, turning to learning, we see in the structural learning theory that knowledge acquisition appears to take place according to a simple, very specific mechanism in which control shifts among initial and higher level goals in a predetermined (fixed) manner, a manner assumed to be characteristic of all people. Although we did not attempt to summarize all of his arguments, Scandura (1973) has shown how this one mechanism also deals with motivation, storage and retrieval from memory, and interpretative processes by which meanings are assigned. In state space formulations, on the other hand, learning involves transforming given state spaces, represented for example by tagging states and/or operators or by adding new elements. In contrast to learning, retrieval involves searching through a state space. Motivation has hardly been considered within this framework.

It may be noted parenthetically that, where only part of the relevant knowledge known to a subject is available, a somewhat more general formulation is required. Ignoring processing capacity for the moment, the structural learning theory allows for retrieval (generation) of needed information, including generation of rules (elements) in domains of available rules as well as rules themselves. In state space theories, this corresponds to the commonly assumed situation where only a selected few of the nodes (states) may serve as starting locations.

Fourth, where processing capacity is fixed, as it is both in the "enriched" structural learning theory and in most current information processing theories, specific allowance must be made for erasure of elements from active status, as well as for the generation (activation) of new elements. In state space formulations the processes by which elements are erased and added have a probabilistic and/or arbitrary character. In some theories capacity is assumed to relate primarily to the state spaces themselves. In the Anderson-Bower (1972) theory, for example, admittedly arbitrary characteristics of spaces are used to decide which items (old or new) are to remain active when capacity is exceeded. In others (e.g., Rumelhart, Lindsay & Norman, 1972; Newell & Simon, 1972) capacity relates primarily to processes. Rumelhart et al. (1972), for example, assume a fixed mechanism which recodes active information whenever capacity is reached. In the structural learning formulation, this would correspond to the generation of a new processing procedure (rule). Although probability (or at least nondeterminism) enters the structural learning theory at this level for the first time (e.g., with computations involving given rules), certain general constraints relating to the goal switching mechanism are assumed to govern the erasure of information from active store.

In sum, the structural learning theory appears to have greater generality and parsimony. Critical parts of the theory have also withstood rather demanding empirical test. The situation regarding heuristic power in generating research is inconclusive at present, since both formulations appear pregnant in this sense. It is basically a case of competing paradigms (Kuhn, 1970).

It should be noted, however, that very little work has been done to date in applying the structural learning theory to natural language. The reasons are several, not the least of which is my shared belief (cf. Greeno, 1972) that more progress can be made, at least initially, by attacking less ambiguous kinds of knowledge before moving ahead pell-mell into the man-made morass called "natural language."

Relationships to Heterarchical Systems in Artificial Intelligence

In this section we comment briefly on the relationship between the structural learning mechanism, and the notion of heterarchical control in systems of artificial intelligence (Minsky & Papert, 1972).

For a time artificial intelligence systems were viewed as wholes, as frequently complex programs. As work in the area progressed, the difficulties of building upon earlier work, and even of making changes in existing systems, become increasingly clear because of the close interrelationships among various parts of such systems. To overcome this limitation, heterarchical, or modular planning has been used (e.g., Winograd, 1971, Winston, 1970, Charniak, 1972). Heterarchical systems consist of sets of programs (modules) pertaining to syntax, semantics, line detection, and so on, together with an heterarchical executive which switches control among these "modules" in accordance with a predetermined plan. At the present time, the MIT group is planning ways of enriching the heterarchical control systems they have developed to date to allow for more flexibility (Winograd, personal communication).

It should be apparent that modules in heterarchical systems correspond essentially to rules in the structural learning theory; the executive control structure corresponds to the basic mechanism. There is, however, an important difference between the two. In heterarchical systems, the basic goal is pragmatic. Such systems make it easier to modify and to build upon previous work. No one seriously means to imply that heterarchical control reflects the way people perform, although in developing artificial intelligence systems intuitive judgments are sometimes made with this in mind.

Although any rule system conforming to the structural learning mechanism can be simulated with (in fact is) a heterarchical system and vice versa, this is not the main point. The structural learning mechanism is assumed to be built into people (presumably from birth); it is not learned and need not be taught. While the rules a person knows may increase from time to time, the mechanism is assumed to remain constant.

This is a strong claim, something which no responsible person would make concerning executive systems currently used in heterarchical systems. Among other things, it is very unlikely that an existing control system would be useful in systems other than the one for which it was designed. It is my contention that benefits might accrue in artificial intelligence and, of course, in simulation if structural learning like control structures were used.

Conclusions

By way of summary, let us return to the questions with which we began. In a theory of memory, what parts should be fixed? What parts should be flexible?

It would appear from the structural learning analysis, that while certain portions of cognitive theory seem to be fixed, much more appears to be flexible. Furthermore, the question of what is fixed and what is flexible enters in a number of different ways. Competence theories, for example, are fixed, at least for given populations and particular content. Knowledge, however, is flexible. It varies over individuals, although there are specific methods for determining knowledge from individual behavior and a fixed competence theory.

According to our analysis, the mechanism by which knowledge is selected, put to use, and acquired, also appears to be fixed. Finally, it would appear that each person has a fixed finite capacity for processing information, a capacity rooted not in the processes used but in the physiological character of man.

REFERENCES

Amarel, S. On representations of problems of reasoning about actions. In D. Michie (Ed.), Machine intelligence 3. New York: American Elsevier, 1968.

Bartlett, F. C. Thinking: An experimental and social study. New York: Basic Books, 1958.

Bower, G. H. A selective review of organizational factors in memory. In E. Tulving, & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Charniak, E. Toward a model of children's story comprehension. Report No. AI-TR-266, Artificial Intelligence Laboratory, M. I. T., 1972.

Chomsky, N., & Miller, G. A. Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology. Vol. II. New York: Wiley, 1963.

Collins, A. M. & Quillian, M. R. How to make a language user. In E. Tulving, & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Crothers, E. J. Memory structure and the recall of discourse. In E. Tulving, & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Dalrymple-Alford, E. C. Perception and immediate memory. British Journal of Psychology, 1967, 21, 63-67.

Durnin, J. H., & Scandura, J. M. An algorithmic approach to assessing behavior potential: Comparison with item forms and hierarchical technologies. Journal of Educational Psychology, 1973, in press.

Frederickson, C. H. Effects of task-induced cognitive operations on comprehension and memory processes. In J. B. Carroll, & R. O. Freedle (Eds.), Language comprehension and the acquisition of knowledge. Washington: Winston, 1972.

Furth, H. G. Piaget and knowledge: Theoretical foundations. Englewood Cliffs, N.J.: Prentice Hall, 1969.

Gagné, R. The acquisition of knowledge. Psychological Review, 1962, 59, 355-365.

Glanzer, M., & Clark, W. H. The verbal loop hypothesis: Binary numbers. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 301-309.

Greeno, J. G. On the acquisition of a simple cognitive structure. In E. Tulving, & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Kuhn, T. The structure of scientific revolutions. Chicago: University of Chicago Press, 1970.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Levine, M. Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 1966, 71, 331-338.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Minsky, M., & Papert, S. Artificial intelligence: Progress report. Artificial Intelligence Laboratory, M. I. T., 1972.

Nelson, R. J. Introduction to automata. New York: Wiley, 1968.

Newell, A., & Simon, H. A. Human problem solving. Engelwood Cliffs, N.J.: Prentice Hall, 1972.

Nilsson, N. J. Problem-solving methods in artificial intelligence. New York: McGraw-Hill, 1971.

Polya, G. Mathematical discovery. New York: Wiley, 1962.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), Semantic information processing. Cambridge, Mass.: M. I. T. Press, 1968.

Reitman, W. R. What does it take to remember? In D. A. Norman (Ed.), Models of human memory. New York: Academic Press, 1970.

Restle, F. Theory of serial pattern learning: Structural trees. Psychological Review, 1970, 77, 481-495.

Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. A process model for long-term memory. In E. Tulving, & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Scandura, J. M. Rule-governed behavior in mathematical structures. University of Pennsylvania Structural Learning Series, Report 48, February, 1969.

Scandura, J. M. Deterministic theorizing in structural learning: Three levels of empiricism. Journal of Structural Learning, 1971, 3(1) 21-53.

Scandura, J. M. Mathematics: Concrete behavioral foundations. New York: Harper & Row, 1971.

Scandura, J. M. Competence, knowledge and their relation to human behavior. University of Pennsylvania Structural Learning Series, Report 69, November, 1972.

Scandura, J. M. Structural learning: I Theory and research. New York: Gordon & Breach, 1973.

Scandura, J. M., & Durnin, J. H. Assessing behavior potential: Adequacy of basic theoretical assumptions. Journal of Structural Learning, 1973, in press.

Simon, H. A. Complexity and the representation of patterned sequences of symbols. Psychological Review, 1972, 79, 369-382.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. Acta Psychologica, 1969, 30, 276-315.

Suppes, P. Stimulus-response theory of finite automata. Journal of Mathematical Psychology, 1969, 6, 327-355.

Vitz, P. C., & Todd, R. C. A coded element model of the perceptual processing of sequential stimuli. Psychological Review, 1969, 76, 433-449.

Winograd, T. Procedures as a representation for data in a computer program for understanding natural language. Report No. AI-TR-17, Artificial Intelligence Laboratory, M. I. T., 1971.

Winston, P. Learning structural descriptions from examples. Report No. AI-TR-231, Artificial Intelligence Laboratory, M I. T., 1970.

FOOTNOTES

1. State spaces (tree diagrams) are also widely used in analyzing the hierarchical structure of subject matters (e.g., Gagne, 1962). Such structures correspond to levels of refinement of rules in the structural learning theory (for details, see Scandura, 1973, Chapter 5).

2. The model also allows for learning during recall test trials but this complication need not concern us here.

3. The authors also define higher order nodes which, analogous to Gagne's (1965) use of the term, are concatenations of other nodes, not to be confused with higher order rules in the structural learning theory (Scandura, 1973). The latter operate on classes of rules and may, for example, generate composite (concatenations of) rules.

4. Even though knowledge is always defined in terms of the rules in a predetermined competence theory, it must not be thought that such knowledge is arbitrary. If two or more rules of competence each provide a consistent basis for assessing behavior potential (i.e., if performance on the respective equivalence classes is homogeneous), then the respective (sub)rules used to characterize knowledge are necessarily equivalent. Furthermore, any viable competence theory in this view must be capable of withstanding behavioral test (Scandura, 1972). Competence and knowledge are analogous to the chicken and the egg insofar as priority is concerned.

5. In actuality, this mechanism is oversimplified. For details concerning an enriched mechanism which deals with rule selection (where two or more rules apply), and which allows for false starts (i.e., backtracking), see Scandura (1973, Chapter 9).

6. There was reason to believe in the one deviant case that the conditions of the experiment had not been adequately fulfilled. The subject was run through the same experiment a week later, using different rules, this time with positive results.

7. In order not to mislead the reader, it should be mentioned that this "first" attempt came after a considerable amount of preliminary pilot work.

8. The results of a seventh --- subject were difficult to interpret because the D test on Posttest I indicated that the D training had not been effective. After Posttest I the subject was retrained on the D rule and Posttest I was repeated. Then he was given the H training and Posttest II. In both cases, he succeeded on all three tasks.

9. According to Scandura (1973), the main reason that this has not been done is because the theory seems to call for different kinds of data. The theory does not seem to provide any compelling insights into free recall, for example. Its major advantages seem to lie in the analysis of memory of more meaningful knowledge which can be readily and unambiguously represented in terms of rules.

10. Scandura (1973) made no attempt to deal with the question of processing time, deferring here to ongoing research in the area (e.g., Sternberg, 1969). Scandura's position was that processing time may ultimately be traced to certain physiologically based behavior constants of individual subjects, in the same sense that the processing capacity of individual subjects is fixed.

TABLE 1

Experiment I: Summary of Results*

		H-D Training																						D-H Training		No Training (Prior Knowledge)							
		9	11	9	9	9	9	9	10	10	9	9	9	10	10	10	10	9	10	10	12	10	12	8	9	9	11	11	11	10	11	11	11
Interpreting simple & composite rule cards		G	B	B	B	B	B	G	G	B	B	B	B	G	G	B	G	G	B	B	G	G	G	G	B	G	G	B	G	B	B	G	G
		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Transfer Pretest	H	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+
	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
	HD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
Training		H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	D	D								
Transfer Posttest I	H	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+								
	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+								
	HD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+								
Training		D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	H	H								
Transfer Posttest II	H	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+								
	D	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+								
	HD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+								

*Encircled data indicates a result which was contrary to prediction.

Table 2

Experiment II: Summary of Results

	H-D Training										D-H Training										
	9 G	10 G	8 G	10 G	9 B	9 B	10 G	9 B	10 G	10 G	7 G	7 G	8 G	7 G	8 B	8[b] G	7 B	7[b] B	7[c] B	8 B	7 B
Interpreting Simple and Composite Rules	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+ +	+	-
Transfer Pretest																					
H	-	-	+	+	+	+	+	+	+	+	-	-	-	-	-	+?	-	+?	- +?	+	
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- +	-	
HD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- -	-	
Training	H	H	H	H	H	H	H	H	H	H	D	D	D	D	D	D	D	D	D D	D	
Transfer Posttest I																					
H	+	+	+	+	+	+	+	+	+	+	+[a]	+[a]	-	-	-	+?	-	+	- +[a]	+	
D	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+ +	+	
HD	-	-	-	-	-	-	-	-	-	-	+[a]	+[a]	-	-	-	-[a]	-	+	- +[a]	+	
Training	D	D	D	D	D	D	D	D	D	D	H	H	H	H	H	H	H	H	H H	H	
Transfer Posttest II																					
H	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+ +	+	
D	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+ +	+	
HD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-[a] +	+	

a. Results which ran contrary to predictions.

b. Subjects succeeded on first presentation of H task (indicated +?) but failed thereafter.

c. After failure on HD task in Posttest II, this subject was run again using different rule cards.

TABLE 3

Experiment III: Summary of Results

D-H Training

		7	8	7	9	8	9	9[c]	8	7	7	7	8	8	8	9[b]	
		G	B	B	B	G	G	G	B	B	G	B	G	B	G	B	
Interpreting simple & composite rule cards		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Transfer Pretest	H	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	
	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	HD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Training		D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
Transfer Post-test I	H	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+[a]
	D	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-[a]	+
	HD	+	+	+	+	+	+	-[a]	+	-	-	-	-	-	-	-	+[a]
Training		H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	
Transfer Post-test II	H	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	D	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	HD	+	+	+	+	+	+	+	+	+	+	-[a]	-[a]	+	+	+	

a. Results which were contrary to prediction.

b. Since the results of Transfer Posttest I indicated that the D training had been ineffective, the D training and Transfer Posttest I were repeated (using different cards) before going to H training and Posttest II.

c. This subject only succeeded after special help which is described in text.

FIGURE CAPTIONS

Figure 1. The directed graph represents a state space in which the nodes represent states and the arrows represent operators or relations. S denotes the starting state and the G denote goals.

Figure 2. This figure shows a portion of the state space for DONALD + GERALD = ROBERT illustrating a simple heuristic search. Once a digit has been assigned to one letter (e.g., 4 to T), it cannot be assigned to other letters thereby reducing the search.

Figure 3. The directed graphs labelled 1, 2, 3, and 4 represent the four paths through the indicated procedure for generating the "next" numeral in Base Three Arithmetic. The sample S-R pairs belong to the four equivalence classes defined by the paths.

Figure 4A and 4B. Schematic representations of memory in the idealized theory of structural learning (4A) and in the "enriched" theory of memory (4B).

Figure 5. Samples of simple and composite rule cards.

Figure 1

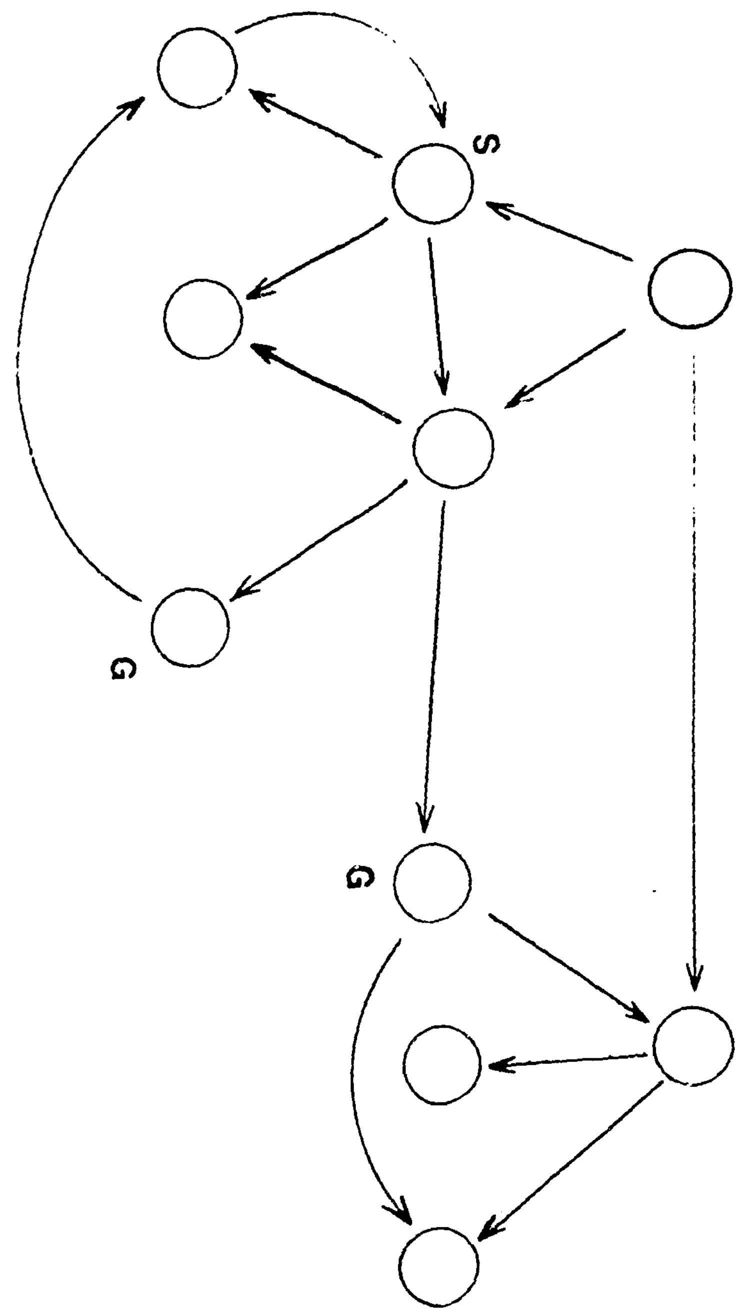


Figure 2

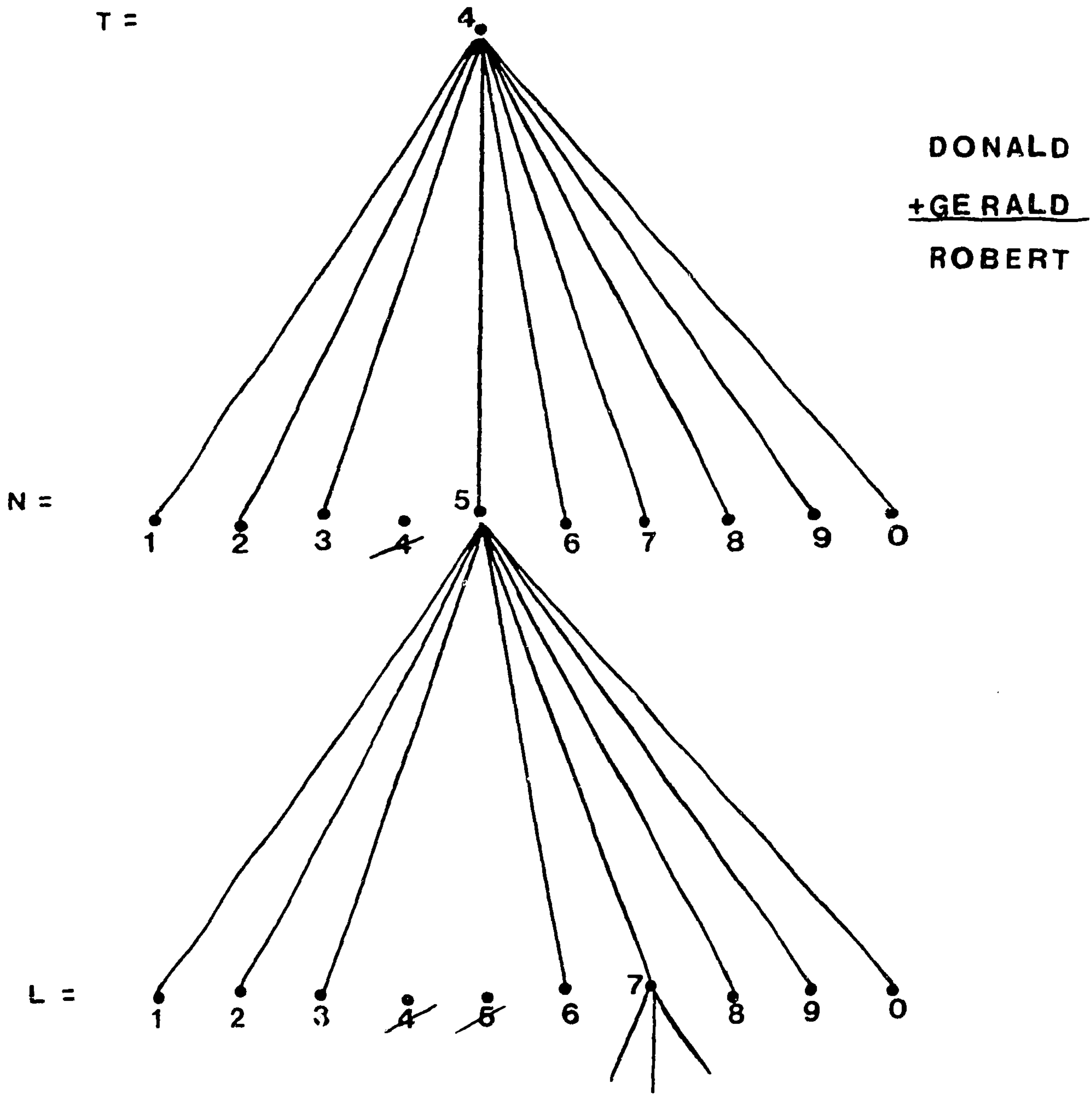
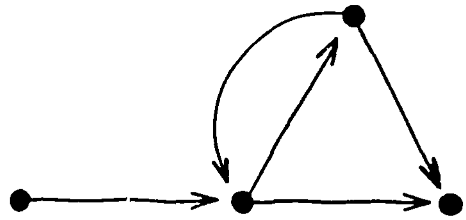


Figure 3

Sample

Stimuli--->	Responses
101--->	102
12--->	20
2--->	10
222--->	1000

Total Graph



Paths (Subgraphs)

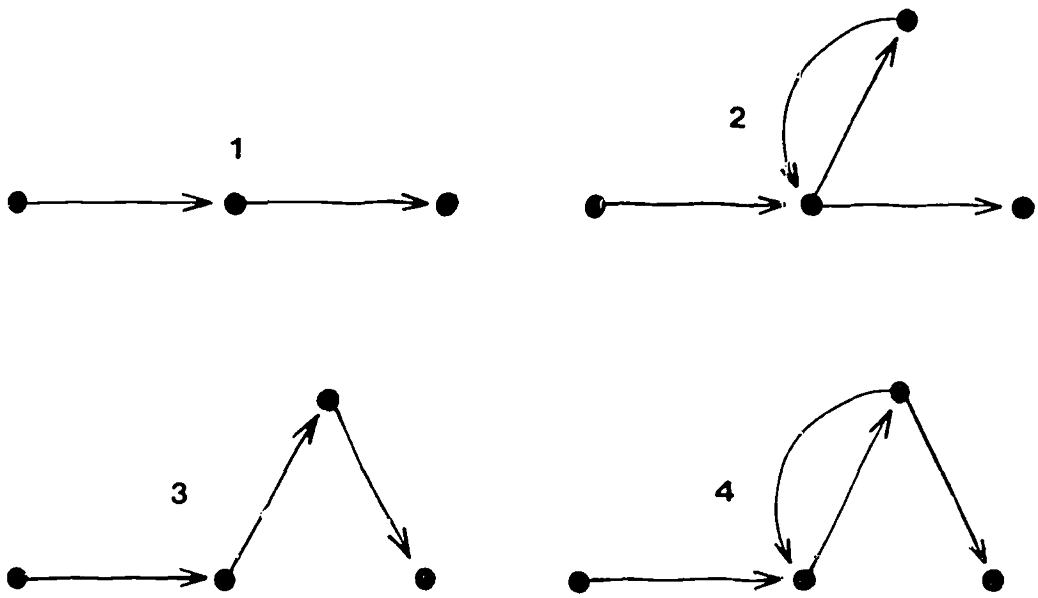
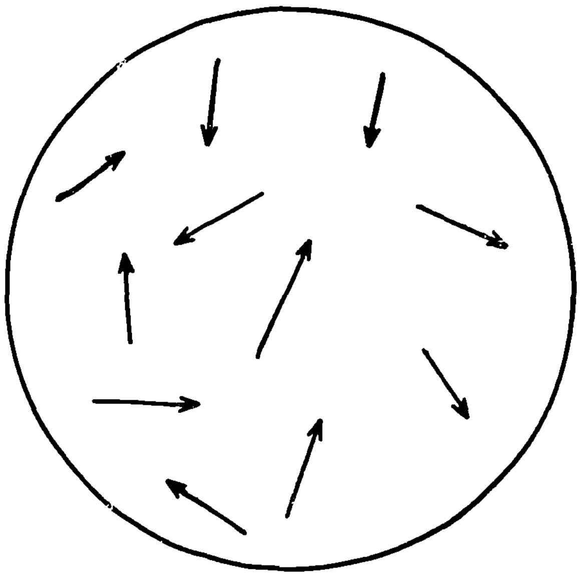


Figure 4A



A

Figure 4B

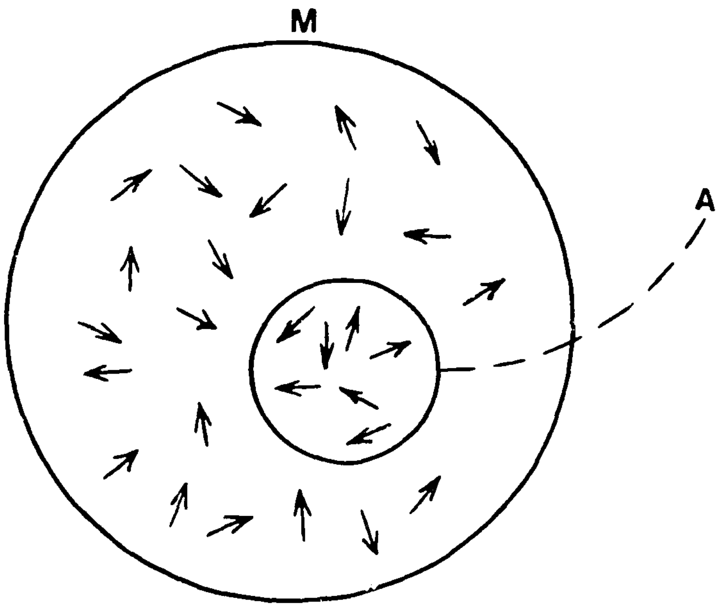
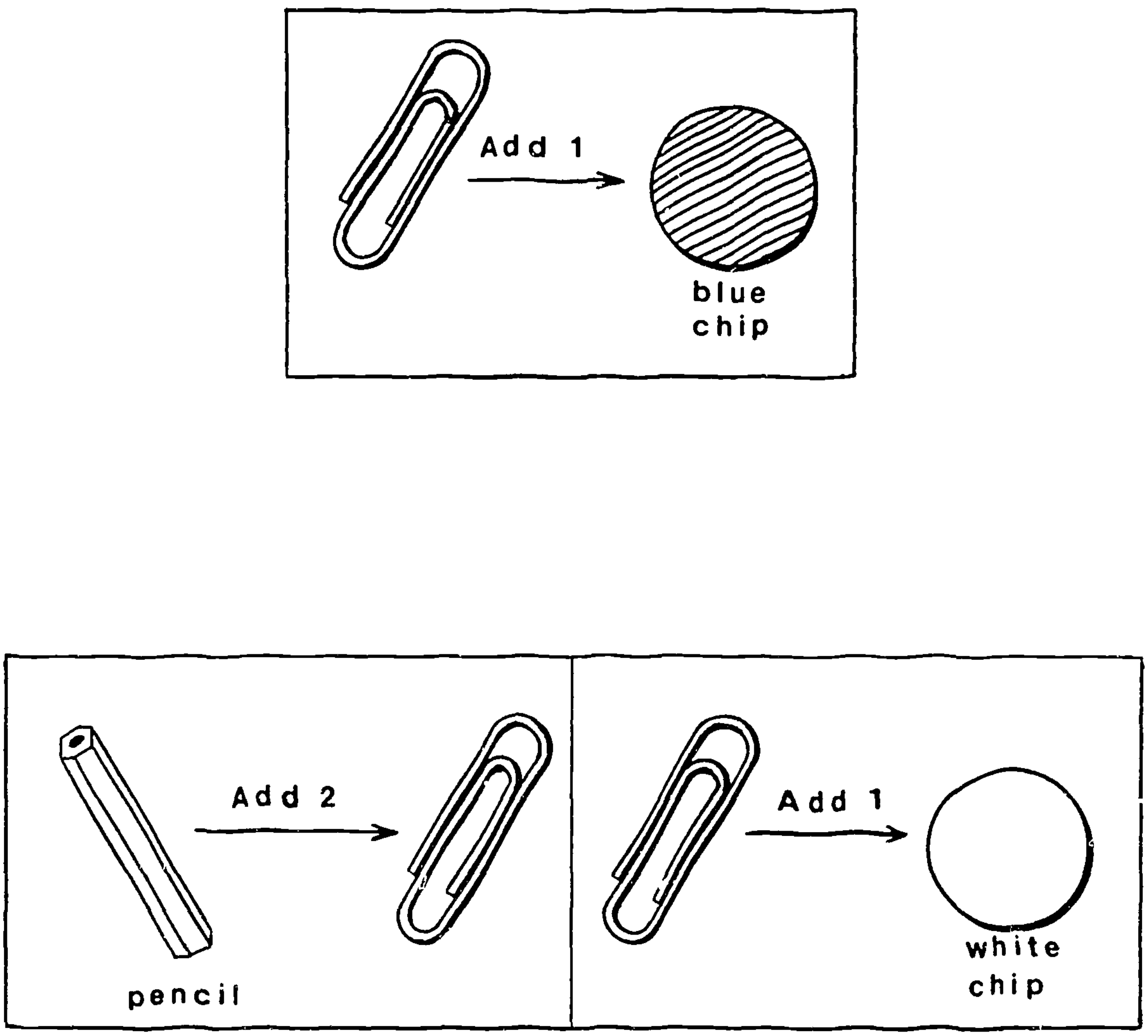


Figure 5



MERG
(Mathematics Education Research Group)

University of Pennsylvania

Reports

(Place of Publication is shown in parentheses.)

1. Scandura, J. M. An exceptionally rare athlete. 1959. (Mentor Magazine, 1959, 30-31.)
2. Scandura, J. M. Abstract card tasks for use in problem solving research. 1964. (Journal of Experimental Education, 1964, 33, 145-148.)
3. Scandura, J. M. An analysis of exposition and discovery modes of problem solving instruction. 1964. (Journal of Experimental Education, 1964, 33, 149-154; also in P. C. Burns & A. M. Johnson (Ed.) Research on Elementary School Curriculum and Organization. Boston: Allyn and Bacon, 1969; and in Dissertation Abstracts, 1963, 23, No. 8.)
4. Scandura, J. M. The other role of schools of education. 1964. (Western New York Study Council Journal, 1964, 15, 11-13.)
5. Scandura, J. M. Fractions--names and numbers. 1964. (The Arithmetic Teacher, 1964, 11, 468.)
6. Scandura, J. M. Probability theory--basic concepts. 1965. (New York State Mathematics Teachers Journal, 1965, 112-118.)
7. Scandura, J. M. & Nelson, J. L. The emerging research role of the subject matter educator. 1965. (Journal of Research in Science Teaching, 1965, 3, 51-53.)
8. Scandura, J. M. Educational research and the mathematics educator. 1965. (The Mathematics Teacher, 1965, 58, 131-138; also in The Mathematics Education, 1967, 1, 1-14.)
9. Scandura, J. M., Johnson, D. A., & Thomason, G. M. Federal funds for the improvement of mathematics education. 1965. (The Mathematics Teacher, 1965, 58, 551-554; also in The Arithmetic Teacher, 1965, 12, 496-499.)
10. Scandura, J. M. Familiarization and cue selection. 1965. (Psychological Reports, 1965, 17, 27-38.)
11. Scandura, J. M. Familiarization and one-trial learning: A positive interaction. 1965. (Psychological Reports, 1965, 17, 27-38.)
12. Scandura, J. M. Problem solving and prior learning. 1965. (Journal of Experimental Education, 1966, 34, 7-11.)
13. Scandura, J. M. Algorithm learning and problem solving. 1966. (Journal of Experimental Education, 1966, 34, 1-6.)
14. Scandura, J. M. Prior learning, presentation order and prerequisite practice in problem solving. 1966. (Journal of Experimental Education, 1966, 34, 12-18.)
15. Scandura, J. M. & Behr, M. Prerequisite practice and criterion form in mathematics learning. 1966. (Journal of Experimental Education, 1966, 35, 54-55.)
16. Scandura, J. M. Problem solving and prior learning. 1966. (The Reporter, Central New York School Study Council, School of Education, Syracuse University, 1966, 17(3), 1-5.)
17. Scandura, J. M. Teaching--technology or theory. 1966. (American Educational Research Journal, 1966, 3, 139-146; also in N. I. E. Journal, 1, 11-17; and in Scientia Paedagogica Experimentalis, 1967, 4, 231-243.)
18. Scandura, J. M. Precision in research on mathematics learning: The emerging field of psycho-mathematics. 1966. (Journal of Research in Science Teaching, 1966, 4, 253-274.)
19. Greeno, J. G. & Scandura, J. M. All or none transfer based on verbally mediated concepts. 1966. (Journal of Mathematical Psychology, 1966, 3, 388-411.)
20. Scandura, J. M. & Roughead, W. G. Conceptual organizers in short-term memory. 1966. (Journal of Verbal Learning and Verbal Behavior, 1967, 6, 679-682.)
21. Scandura, J. M. A new doctoral program in mathematics education research. 1966. (The Mathematics Teacher, 1966, 8, 729; also in American Mathematical Monthly, 1967, 74, 72-74.)
22. Scandura, J. M. Concrete examples of commutative non-associative systems. 1966. (The Mathematics Teacher, 1966, 59, 735-736.)
23. Scandura, J. M. Training leaders in mathematics education: A contrast with science education. 1966. (Journal of Research in Science Teaching, 1966, 4, 227-229.)
24. Scandura, J. M. The nature of research in mathematics education. 1967. (The High School Journal, 1967, 50, 227-237.)
25. Scandura, J. M. An alternative approach to research in mathematics learning--a reaction to Suppes's paper. October, 1967. (Journal of Research and Development in Education, 1967, 1, 33-43.)
26. Scandura, J. M. Critique of "Research in the improvement of verbal problem solving." December, 1967.
27. Scandura, J. M. Training in mathematics education: A proposal for action. 1967. (The Mathematics Teacher, 1967, 60, 714 & 719.)
28. Scandura, J. M. & Wells, J. N. Advance organizers in learning abstract mathematics. 1967. (American Educational Research Journal, 1967, 4, 295-301.)
29. Scandura, J. M., Woodward, E., & Lee, F. Rule generality and consistency in mathematics learning. 1967. (American Educational Research Journal, 1967, 4, 303-319.)
30. Scandura, J. M. The basic unit in meaningful learning--association or principle? 1967. (The School Review, 1967, 75, 329-341; also in Proceedings, 75th Annual Convention APA, 275-276, 1967.)
31. Scandura, J. M. Learning verbal and symbolic statements of mathematical rules. 1967. (Journal of Educational Psychology, 1967, 58, 356-364.)
32. Scandura, J. M. Learning principles in paired-associate lists. 1967. (Psychological Reports, 1967.)
33. Scandura, J. M. Concept dominance in short-term memory. 1967. (Journal of Verbal Learning and Verbal Behavior, 1967, 6, 461-469.)
34. Scandura, J. M. & Satlow, J. An analysis of existing curricular materials in mathematics. 1968.
35. Scandura, J. M. (Ed.) Report of a conference on elementary mathematics education and the Pennsylvania State Department of Public Instruction. 1968.
36. Scandura, J. M. Con Ebel (Ebel's "Strawman"). 1968. (Phi Delta Kappan, 1968, 49, 286.)
37. Scandura, J. M. and Anderson, J. Research in mathematics and science education. 1968. (School Science and Mathematics, 1968, 58, 185-192; also in Educational Research and the Science Educator, 1968, 52, 353-358.)
38. Scandura, J. M. Research in psycho-mathematics. October, 1968. (The Mathematics Teacher, 1968, 61, 581-591; also in The Mathematics Education, 1968, 3; and in R. B. Ashlock & W. L. Herman (Ed.) Current Research in Elementary School Mathematics. New York: Macmillan, 1970.)
39. Roughead, W. G. and Scandura, J. M. "What is learned" in mathematical discovery. 1968. (Journal of Educational Psychology, 1968, 59, 283-289.)
40. Scandura, J. M. Expository teaching of hierarchical subject matters. 1968. (Journal of Structural Learning, 1969, 2(1), 17-25; also in Psychology in the Schools, 1969, 6, 307-310.)
41. Scandura, J. M. The formulation of research on subject matter learning. July, 1968. (Psychology in the Schools, 1968, 5, 330-341.)

Scandura, J. M. and Durnin, J. H. Extra-scope transfer in learning mathematical strategies. 1968. (Journal of Educational Psychology, 1968, 59, 350-354.)

43. Scandura, J. M. New directions for theory and research on rule learning: I. A set-function language. 1968. (Acta Psychologica, 1968, 28, 301-321.)
44. Scandura, J. M. New directions for theory and research on rule learning: II. Empirical research. 1968. (Acta Psychologica, 1969, 29, 101-133.)
45. Scandura, J. M. New directions for theory and research on rule learning: III. Analyses and theoretical direction. 1968. (Acta Psychologica, 1969, 29, 205-227.)
46. Scandura, J. M. Federal control and basic research in education. 1968. (School & Society, 1969, April, 227-228.)
47. Scandura, J. M., Barksdale, J., Durnin, J., & McGee, R. An unexpected relationship between failure and subsequent mathematics learning. 1968. (Educational Studies in Mathematics, 1969, 1, 247-251; also in Psychology in the Schools, 1969, 4, 379-381.)
48. Scandura, J. M. Rule-governed behavior in mathematical structures. February 2, 1969.
49. Scandura, J. M. A research basis for mathematics education. November 10, 1969. (High School Journal, 1970, 53, 264-280; Educational Studies in Mathematics, 1971, 3, 229-243; and Journal of Structural Learning, 1970, 2(3), 1-18.)
50. Scandura, J. M. The role of rules in behavior; Toward an operational definition of what (rule) is learned. November 25, 1969. (Psychological Review, 1970, 77, 516-533.)
51. Scandura, J. M. & McGee, R. An exploratory investigation of basic mathematical abilities of kindergarten children. December 1, 1969. (Educational Studies in Mathematics, 1972, 4, 331-345.)
52. Scandura, J. M. A theory of mathematical knowledge: Can rules account for creative behavior? January 28, 1970. (Journal for Research in Mathematics Education, 1971, 2, 183-196.)
53. Ehrenpreis, W. & McGee, R. Research implications for the elementary school: From the laboratory to the classroom. March 1, 1970.
54. Scandura, J. M. & Voorhies, D. The role of irrelevant attributes and irrelevant operations in rule learning. May 20, 1970. (Journal of Educational Psychology, 1971, 62, 352-356.)
55. Scandura, J. M. Deterministic theorizing in structural learning: Three levels of empiricism. May 20, 1970. (Journal of Structural Learning, 1971, 3(1), 21-53.)
56. Scandura, J. M. What is a rule? May 21, 1969. (Journal of Educational Psychology, 1972, 63, 179-185.)
57. Scandura, J. M. Problem solving ability in elementary mathematics--its nature and measurement. February 8, 1968.
58. Scandura, J. M. Theoretical note: S-R theory or automata? (A reaction to Suppes' reaction). In Structural Learning I.
59. Scandura, J. M. A plan for the development of a conceptually based mathematics curriculum for disadvantaged children. October 15, 1970. (Instructional Science, 1972, 247-263.)
60. Scandura, J. M. An algorithmic base for curriculum construction in mathematics (A theory of mathematical knowledge). October 29, 1970. (In Lamon, W. E. (Ed.) Learning and the Nature of Mathematics. Chicago: Science Research Associates, Inc., 1971 .)
61. Scandura, J. M. On the value of status studies in education. (Journal for Research in Mathematics Education, 1971, 2. 115-117.)
62. Scandura, J. M., and Durnin, J. H. Assessing behavior potential: Adequacy of basic theoretical assumptions. November 30, 1971. (Journal of Structural Learning, in press.)
63. Durnin, J. H., and Scandura, J. M. An algorithmic approach to assessing behavior potential: Comparison with item forms and hierarchical technologies. January 1, 1972.
64. Ehrenpreis, W., and Scandura, J. M. An algorithmic approach to curriculum construction in mathematics: A field test. March 1, 1972.
65. Lowerre, G., and Scandura, J. M. Conceptually based development of individualized materials for critical thinking based on logical inference. May 1, 1972.
66. Scandura, J. M. Mathematical problem solving. June 1, 1972. (American Mathematical Monthly, in press.)
67. Scandura, J. M. The role of higher order rules in problem solving. September, 1972.
68. Scandura, J. M. Structured knowledge: Static or dynamic? (Contemporary Psychology, 1972, in press.)
69. Scandura, J. M. Competence, knowledge and their relation to human behavior. (Talk given at Psychonomic Society Meeting in St. Louis, Missouri, November 2, 1972.)
70. Scandura, J.M., Durnin, J. H., Wulfeck, W.H. II. Higher order rule characterization of heuristics for compass and straight edge constructions in geometry. December, 1972.

Books

Scandura, J. M. Mathematics: Concrete Behavioral Foundations. New York: Harper & Row, 1971.

Scandura, J. M., Durnin, J. H., Ehrenpreis, W., & Luger, G. An Algorithmic Approach to Mathematics: Concrete Behavioral Foundations. Harper & Row, 1971.

Scandura, J. M. Structural Learning I: The Theory and Empirical Research. New York: Gordon & Breach Science Publishers (440 Park Avenue South, New York, New York 10016), 1973.

Scandura, J. M. (Ed.) with contributions by Arbib, M., Corcoran, J., Domotor, Z., Greeno, J., Lovell, K., Newell, A., Rosenbloom, P., Scandura, J., Shaw, R., Simon, H., Suppes, P., Wittrock, M., & Witz, K. Structural Learning II: Views and Approaches. New York: Gordon & Breach Science Publishers, 1973, in press.