Reprinted from Journal of Structural Learning, 1971 3, 21-53.
(Gordon & Breach Science. Pub. Ltd. 440 Park Ave. South New York, N.Y. 10016)

Deterministic Theorizing in Structural Learning: Three Levels of Empiricism

JOSEPH M. SCANDURA

MERGE and University of Pennsylvania

In spite of the diversity that presently exists in behavioral theorizing, reference to probabilistic notions is all-pervasive. Even support at the .05 level of significance is often enough to elicit whoops of glee from most cognitive theorists. Given this milieu, it is not too surprising that (aside perhaps from computer simulation types and a few competence theorists (e.g., Miller and Chomsky, 1963)), no one seems to have seriously pursued the possibility that deterministic theorizing about complex human learning may actually be easier than stochastic theorizing. And yet, this is precisely what in my own work I have found to be the case.

The purpose of this article is to describe the "rudiments" of a potentially powerful and internally consistent **deterministic partial theory of structural learning**, which could make it possible to explain, and hopefully also to predict, certain critical aspects of the behavior of individual subjects in specific situations. The term "rudiments" is used because at the present time relatively few implications of the theory have been drawn out. The emphasis so far has been on establishing a fit between behavioral reality and the basic constructs and hypotheses of the theory,

As suggested by the title, there are really three different partial theories, each of which must be tested in a different way. First, there is a theory of structured knowledge -- or, more accurately as we shall see below, theories of structured knowledge. These theories deal with the problem of how to characterize knowledge. (The knowledge had by any given individual constitutes a theory in its own right.) Second, there is a theory of idealized behavior which tells how knowledge is selected for use, and how it is learned. This theory applies only where the subject is unencumbered by memory or by his finite capacity to process 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 are not independent of one another, although, as we shall see, research on any one can progress independently of the others and this includes empirical testing.

PRELIMINARY OBSERVATIONS

Before describing these partial theories, some general background may be helpful.

There are three main ideas which my title conveys. The label "structural learning" sets the whole tone for the title, so we consider that first. Structural learning refers to the knowledge a person may have and the behavior (and learning) which this knowledge makes possible. More specifically, structural learning is concerned with complex human learning and behavior which cannot naturally be studied without giving explicit attention to what the subject knows before he enters the learning or behaving situation. Any attempt to study mathematics learning, for example, with reference only to the stimulus situation would be folly to the nth degree. Individual differences in prior knowledge and other intellectual skills in mathematics may be very great indeed, and these differences must be taken explicitly into account in any theory that is to provide a viable account of complex mathematics learning. It should be noted parenthetically that one of the primary requisites for selecting tasks in most traditional studies has been that prior learning be of minimal importance. The reference here, of course, is to experiments on serial and paired-associate learning, classical conditioning, and the like.

Dependence on prior knowledge, then, is important to my conception of structural learning. But this alone is not sufficient. The knowledge involved must also have a reasonably clear structure. In this sense, mathematics, for example, tends to have a clearer structure than, say, the social studies or the humanities. The fact that grammarians, like Harris and Chomsky, have been able to make as much progress as they have in linguistics attests to a good deal of structure in language as well.

The second dominating phrase in my title is "deterministic theorizing". In view of the tradition in psychology against this type of theorizing, it is instructive to consider the paradigm most typically used in testing behavioral theories. First, assumptions are made about how ***individuals*** learn or behave. When stated in their clearest form, as in the stochastic theories of mathematical psychology, the basic assumptions are stated in terms of probabilities. Second, inferences are drawn from these assumptions yielding predictions about group statistics -- that is, about characteristics of the distributions of responses made by the experimental subjects. Third, on the basis of the experimental results obtained, inferences are made about the basic assumptions.

Of course, there is no harm in this as long as it is recognized that the initial assumptions deal with probabilities *(associated with group behavior)* and not with individual processes. But this fact has not always been made as explicit by theorists as might be desirable. What needs to be made clear with such probabilistic theories is that what any given subject does on a given occasion may have little or nothing to do with the particular assumptions made. For example, in stochastic models of paired-associate learning it is usually assumed that each subject has the same probability of learning on each trial. Even the most superficial analysis of relevant data, however, indicates clearly that the probability of success for different subjects may vary greatly. And one cannot attribute this to the fact that the probability of learning is a random variable. This would still not explain the fundamental fact that the probability of success of many subjects tends to be either uniformly high, or low, over different trials.

How much better it would be to have a theory which would tell us explicitly what a given subject will do on specific occasions -- a theory which leaves errors in prediction to inadequacies in observation and measurement, and does not make these errors an explicit part of the theory itself. Ideally, such a theory would satisfy the classical conditions for a deterministic theory in the hard sciences -- theories that say, in effect, that given such and such basic hypotheses and these initial conditions, this is what should happen. Given a theory of this sort, probability would enter only where one wanted to make predictions in relatively complex situations where the experimenter practically speaking could not, or did not wish to, find out everything he would need to know and specify in order to make deterministic predictions. In effect, a truly adequate deterministic theory would make it possible to generate any number of stochastic theories by loosening one or another of various conditions which must be satisfied in order for the deterministic theory to apply. (In this regard, see the comments below on levels of empiricism and conditional hypotheses.)

In order to be completely honest, I must mention one further reason why deterministic theorizing appeals to me. I am basically lazy. I have done a good deal of traditional behavioral research, but I dislike with a passion poring over reams of raw data or computer printouts, especially when I know that, no matter what statistics are used to summarize the data, I am losing much, if not most, of what is important. It is perhaps this distaste as much as anything else which has moved me to search for a new and better way to do empirical research on complex human learning. How much nicer to have data which is clearcut, no means or variances to compute, no analyses of variance, or canonical correlations, or factor analyses -- just looking. In this regard, I can't resist the temptation to repeat a little story about an experience I had as a post-doctoral student being initiated into mathematical psychology at Indiana University. The time was the summer of 1962, and the field was bright and promising. As part of my orientation, I was routed about to visit a number of the more prominent names on campus, including one very fine physiologist. Caught up by the emphasis on mathematics given by the psychologists, I asked him what kinds of mathematics he found most useful in his work, and how he used it. His answer was, "We count." After getting over my initial shock, I began to see the logic of his answer, and have been trying to meet his ideal ever since.

Finally, let us consider what is meant by "levels of empiricism“. Recall first that any theory is but a ***partial*** model of reality. It deals adequately with certain phenomena in the sense of providing an adequate explanation for them, but not others. Theories do not apply universally. To make the point in its most trivial sense, we need only note that existing theories of thermodynamics, for example, are not likely to be very useful in explaining paired-associate learning--or vice-versa. As a more realistic example, learning theories such as Hull's provide a far better account of certain simple behavioral phenomena than they do, for example, of the learning of complex mathematical structures. (Partial theories must not be confused with so-called miniature theories of mathematical psychology. Partial theories deal with only certain phenomena of given, broad-based realities. Miniature theories deal intensively with highly restrictive phenomena such as paired-associate learning.)

The general difficulty with most theory construction in psychology, today, is that very little attention has been given to specifying conditions under which theories are **not** presumed to hold. To date, the sole approach to this problem has been an ad hoc empirical one in which experimental evidence is gradually accumulated over relatively long periods of time.

It is my feeling that much can be done along these lines, while theories are actually being constructed. This does not obviate the need for empirical testing, of course. No one believes that we can ever do away with that. But I do think that we can do away with a good deal of it, if theorists would give more explicit attention in their work to identifying these negative conditions.

In constructing a theory, whether it be a mathematical theory or a scientific theory, the theorist has some model, or models, in mind at the time. These models arise basically from particular segments of reality -- but more important here, they usually deal with only certain aspects of that reality. The rest is simply ignored.

This approach may be a viable one in mathematics, where one aims for abstraction. One never knows where mathematical theories may ultimately prove useful (i.e., be applied), and it would undoubtedly be a mistake to tie them in too closely to any particular model, by specifying aspects of these particular models with which the theory does not deal.

This is not true in science, however, where the ***ultimate*** aim may be to devise theories that deal with more of the particular reality in question. A theorist may have many more kinds of phenomena in mind in attempting to construct a theory than he can possibly handle at one time. To get around this problem, he may purposefully ignore for a time ***certain*** of these phenomena to facilitate constructing what might be called a ***partial*** theory--a theory which deals with part of the reality but not all of it.

In constructing such a partial theory, it is critically important that the theorist do so in a way that is compatible with the broader reality. Thus, for example, the ultimate aim of competence theorists such as Chomsky (1968) and Miller and Chomsky (1963) is not just to characterize the knowledge had by an idealized human subject -- that in itself might be attempted in any number of different ways. What these theorists want is a theory of knowledge which is likely to be compatible with a more encompassing behavior theory once one is developed (e.g., see Miller and Chomsky, 1963, 463-488). In such cases, it will generally be in the theorist's interest to know just what aspects of reality his present theory does not consider. Stated differently, he must know what boundary conditions must be satisfied in order for his partial theory to apply. Theoretical predictions based on partial theories are necessarily dependent (on such conditions).

In order to test a partial theory, then, the empirical situation must accurately reflect these boundary conditions. Otherwise, the partial theory will simply not be applicable -- by definition. Perhaps the best known example has to do with linguistics, where grammarians, such as Chomsky (1957), assume an idealized knower -- a knower who can use whatever rules are attributed to him without error, and wherever they might be needed. This type of theory seems to be having increasingly important implications for psychology, but it must be remembered that a competence theory of this sort applies only in those situations where the idealized performer assumption is reasonable to make. (There is a close relationship between these ideas and the so-called ecological approach to behavioral science (Wohlwill, 1970), which is becoming increasingly popular of late. In fact, the partial theories described below provide good examples of the kind of theories for which this approach seems to call.)

The first level of theorizing is concerned with the problem of how to account for the behavior of idealized subjects. More particularly, given *a* finite class or corpus of behaviors, the problem is one of how to characterize the knowledge underlying the corpus in a way which accounts as well for the other behaviors of which an idealized knower of that corpus may be capable. Our approach to this problem involves the invention of a finite set of rules of one sort or another which can be used to generate not only the behaviors in the given corpus, although this is an absolute minimum, but also the other behaviors one might wish to attribute to the knower (Scandura, 1970a, for an earlier but closely related version of this goal see Chomsky, 1957). (A rule may be said to account for a class of behaviors if, given *any* stimulus input associated with the class, the corresponding response may be generated by application of the rule (Scandura, 1968, 1970b).)

As one might suspect, there are any number of different ways in which to characterize the same given corpus. The theoretical problem is one of evaluating these various characterizations to determine which best accounts for other behaviors one might wish to attribute to the knower (Chomsky, 1957; Scandura, 1970a, forthcoming). These additional behaviors constitute the predictions.

Consider some of the alternatives. Undoubtedly, the simplest way to account for a given finite corpus is just to list the behaviors involved. Thus, for example, a list of paired-associates might be characterized as a finite set of ***degenerate*** rules (Scandura, 1968) or, equivalently, as a finite set of associations. Clearly, lists of paired associates are not the sort of corpora we usually have in mind in talking about mathematical and other complex behavior, and characterizations, which consist of simple lists of associations, would be essentially sterile in content. If this were all a person could learn, it would be impossible even to learn how to add numbers, addition fact by addition fact. A person could learn at most a finite number of sums, since each addition fact (e.g., 3 + 5 = 8, 25 + 47 = 72, and so on) would have to be learned separately.

A somewhat more realistic characterization of a corpus of behaviors derives from recent attempts in educational circles to define school curricula in terms of a finite number of operational objectives (e.g., Lipson, 1967). Each of the objectives of these curricula amounts to a class of behaviors which can be generated by a rule; the abilities to add, to multiply, to find areas of triangles, and so on, provide obvious examples. It is possible to account for the behaviors represented by such a corpus, then, by simply listing a finite set of rules. In fact, this is essentially what curriculum constructors who have followed this approach have done. The curricula consist essentially of long lists of rules for achieving (operational) objectives, one rule for each objective.

Clearly, exactly the same idea might be applied in characterizing the knowledge had by individual subjects. A list type of characterization of this sort would have the major advantage of requiring a very simple performance mechanism. Thus, if knowledge is characterized as a list of discrete rules, which operate independently of one another, then a more general theory of performance would need to tell only how such rules are put to use. Since the rules are discrete, no interactive mechanisms need be postulated.

This advantage, however, is also its major disadvantage. Because the characterizing rules are discrete, they cannot account for behaviors, which go beyond the given corpus, except in the most trivial sense. For example, suppose the characterization only included rules for adding, subtracting, multiplying, and dividing. In this case, the subject would be unable to even generate the addition fact corresponding to a given subtraction fact, although one might reasonably expect this type of behavior from a person who was well versed in arithmetic. One might counter that it would be a small thing simply to add a new rule to the original list.

$$c - a = b \rightarrow a + b = c$$

We might even use the distinguishing label "relational rule" since it operates on the elements of a binary relation. Indeed, this is precisely the sort of reply one might expect from curriculum constructors of the operational objectives persuasion. When confronted with the criticism that their objectives do not constitute a mathematically (or otherwise) viable curriculum, they would simply say we can add more objectives.

The trouble with this sort of argument is that it misses the point entirely. Not only would such an approach be *ad* hoc -- which really says nothing by itself except to convey some ill-defined dissatisfaction -- but it would be completely infeasible where one is striving for completeness. To see this, it is sufficient to note that a new rule would have to be introduced for every conceivable interrelationship, and that the number of such interrelationships is indefinitely large. One could easily envision a number of rules so large that no human being could possibly learn all of them. There would not be sufficient time in a single lifetime. The sum total of all mathematical knowledge, which is presently in print, for example, is so vast that no one has, or could, possibly acquire all of it. As vast as this knowledge is, however, a really good mathematician is capable of generating any amount of new mathematics that does not appear in print anywhere. That is, ***he can create***. Much of the new mathematics might be utterly trivial, of course, but the very fact that it exists at all strongly suggests that any characterization such as that described above would almost certainly miss much that is important.

We can get a far more powerful and simple characterization by allowing rules to operate, not just on ordinary stimuli, but on other (lower order) rules as wel1.[1] More specifically, allowing rules to

[1] Higher order rules on rules are common in various branches of mathematics, where they go under the label of functions on functions, but the idea seems not to have generally pervaded either computer science or formal linguistics. In formal linguistics, for example, where the goals closely parallel ours, no one seems to have seriously proposed the use of higher order rules, The closest linguists have come in this regard has been to introduce the notion of grammatical *transformation* between phrase markers (Chomsky, 1957). Rather than higher order rules, transformations correspond more closely to what we have here called *relational* rules (see Scandura, forthcoming).

There are two good reasons why this has probably not been done in the past. First, even grammatical transformations have so far resisted mathematical treatment (Nelson, 1968), and, second, no existing approach to psychology known by the writer provides any real motivation for introducing them. Gagne's (1965) view of problem solving in terms of rules and Miller, Galanter, and Pribram's (1960) TOTE hierarchies come closest.

This is unfortunate, since there is a very simple and intuitively sound reason for allowing rules to operate on (classes of) rules. The main one is just this: There is a very simple and intuitively compelling performance mechanism by which higher and lower order rules may be combined so as to generate completely new kinds of behavior. Furthermore, as shown in the next section, some empirical support for this mechanism has already been obtained.

operate in this way makes it possible to generate new rules and these rules, in turn, may make it possible to generate what might appear to be completely different kinds of behavior. For example, suppose that an idealized knower has mastered the two rules:

(1) $a, b \rightarrow a + b$

(2) $[x, y \rightarrow x \text{ o } y] \Rightarrow [x, y \rightarrow x \text{ o' } y]$

where (1) represents a rule for generating sums of pairs of, say, integers and (2) represents a (higher order) rule which, given a rule of the form (1) for any binary operation, generates a rule for performing the corresponding inverse operation (denoted o'). Such a rule would connect, for example, not only addition of numbers with subtraction, but composition of all sorts with the corresponding inverse operations, whether these operations involved permutations, rotations, rigid motions, or whatever. In this case, application of rule (2) to rule (1) yields rule,

(3) $a, b \rightarrow a - b,$

where "-" is the inverse of "+". Application of rule (3), in turn, makes it possible to generate differences between any given pair of integers a and b where a > b. But, then, isn't this just a simple instance of the sort of thing we have in mind when we think of creative behavior?

If the extrapolation involved seems too tame to qualify for this distinguished label, consider the following example in which we add another level to the analysis. In this case, we assume in addition to rules (1) and (2) that the idealized knower has also mastered rules,

(4) $[x, y \rightarrow x \text{ o } y] \Rightarrow [\boldsymbol{x}, \boldsymbol{y} \rightarrow \boldsymbol{x} \text{ } \boldsymbol{o} \text{ } \boldsymbol{y}]$

(Note : x, y, o are different from ***x, y, o,*** respectively.)

(5) [(x → y), (y → z)] ⇒ [x → z]

Rule (4) may he thought of as denoting knowledge of generalized homomorphic relationships between pairs of systems such as the system (A) of integers under addition and, say, the system (B) of rational numbers under addition. Rule (5) is extremely general and makes it possible to generate the composite (rule) of any pair of given rules such that the output of one of the rules serves as the input of the other.

Knowing these rules would make all kinds of behaviors possible. For example, the idealized knower would be able to subtract, not only in the first system (A) but in the second system (B) as wel1. To see this, we need only observe that application of rule (5) to rules (4) and (2), yields rule

(6) [x, *y* → x o y] ➔ ***[x, y →x o' y]***,

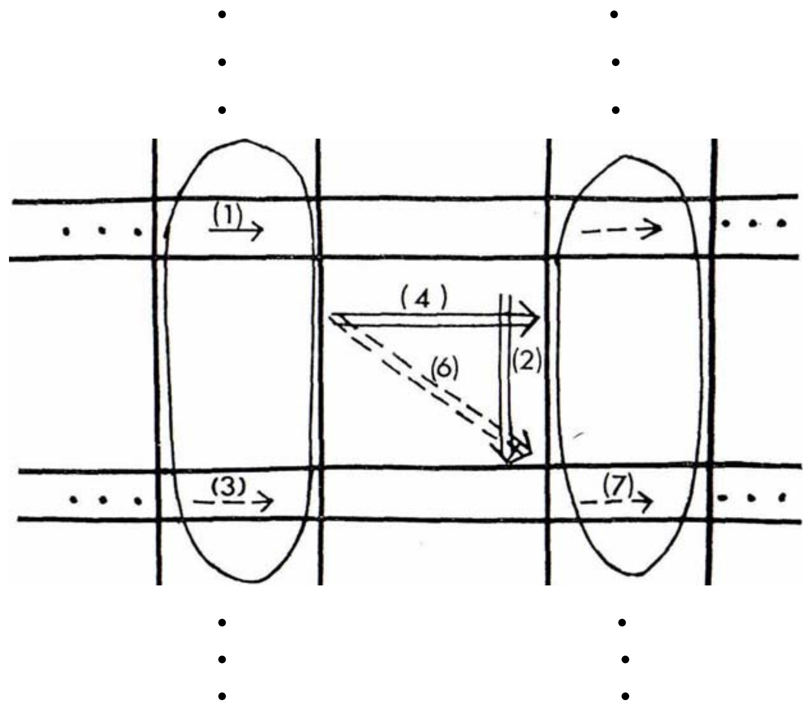
Application of rule (6) to rule (1), then, yields rule

(7) **a, b** → ***a*** **+'** ***b*** or ***a, b*** ➔ ***a*** **-** ***b*** where **+'** = **-** .

Rule (7) is the subtraction rule for system B. The basic relationships are represented schematically in Fig. 1. More details and further examples maybe found in Scandura (1970, forthcoming).

In summary, the essentials of the theory of knowledge as outlined are just these. (1) The knowledge of any *given* individual at any given stage of learning can be characterized in terms of a *finite* set of rules. This implies among other things that there may be as many different theories of knowledge as there are individuals -- or, equivalently, as many theories as there are conceivable curricula to be mastered. **[2006 Update: The distinction made here is more sharply formulated by representing competence {what needs to be learned} in terms of structural & procedural Abstract Syntax Trees {ASTs} and representing knowledge attributed to individuals {via assessment} as programs {data & processes} corresponding to slices through ASTs]** (2) Rules may act on classes of rules as well as on simple stimuli. Allowing rules to act in this way amounts to a simple but conceptually major revision of existent competence theories. (3) For purposes of the theory, it is assumed that the rules may be combined at will and without error as needed. Stated differently, the idealized knower *is* assumed to have mechanisms available for putting the rules attributed to him to use.[2]

[2] There will always be behavior, of course, which cannot be generated by any given finite set of rules. Roughly speaking, when translated into behavioral terms, Godel's (1931) Incompleteness Theorem suggests that no matter how bright an individual, there will always be certain behaviors he will not be capable of performing.



Fig, 1. A schematic representation of the basic relationships described in the text. Solid arrows refer to (pre)learned rules and dotted arrows to derivable rules. Rule (5) by which rules (4) and (2) are combined to give (6) is not represented since this would require a third dimension and would complicate the diagram without adding any clarification.

FOUNDATIONS OF AN IDEALIZED THEORY OF STRUCTURAL LEARNING

The third point above is a critical boundary condition of the theory of knowledge. The theory applies only at the analytical level in the sense that generative grammars account for language behavior. The relevance of the theory to actual human behavior is dependent on our ability to spell out mechanisms which are both adequate to account for how rules may be combined and which are reflected in the actual behavior of human subjects.

It is to this task that we now turn--the task of introducing mechanisms of idealized performance, learning, and motivation into our formulation. The purpose of adding such mechanisms to the theory of knowledge is to obtain an extended theory that deals explicitly with the way in which available knowledge is put to use. This more encompassing theory is still a partial theory, however, one which applies only where subjects are unencumbered by either memory or their intrinsically limited capacity to process information. It should be emphasized, however, that it is a theory, which is assumed to apply no matter what knowledge an idealized subject has available. Thus, even though the knowledge had by different individuals may vary greatly, the same theory of idealized behavior is assumed to hold over all individuals.

The basic assumption on which this theory rests is that people are goal-seeking information processors. In this case, much of what a subject knows becomes irrelevant once a goal situation is specified. Thus, at any given point in time, only a small fraction of the rules available to a knower may be applicable -- namely, those rules which may be used directly or indirectly in satisfying the given goal.

There are three basic kinds of situation with which any viable theory must deal. One type of situation is where the subject knows one or more rules which apply in the given goal situation. The second is where the subject does not explicitly know a rule which applies in the goal situation. The third is actually a refinement of the first, and deals with the question of why, when a subject has more than one rule available, he selects the rule that he does. Why not one of the others? As we shall see, these problems are closely allied with what have traditionally been called performance, learning, and motivation, respectively.

The first case is simplest to deal with. We need only assume that:

(A) Given a goal situation for which a subject has at least one rule available, the subject will apply one of the rules.

Thus, for example, if a subject's goal is to find the sum of two numbers, and he knows how to add, then he will actually use an addition rule.

As trivial an assumption as this may appear, it ***is*** an assumption. It does not follow logically that just because a subject wants to achieve a certain goal and has one or more rules available for achieving it, that he will necessarily use one of them.

Furthermore, the assumption has a number of important implications. One of these is that it provides an adequate basis for determining what might be called a subject's ***behavior potential***, relative to a given class of rule-governed (RG) behaviors. It may be noted in this regard that it is one thing to devise a procedure (rule) which accounts for a given class of RG behaviors and quite another to identify that subclass of behaviors of which a given subject is capable. The first problem is an analytical one and involves inventing a procedure, which accounts for the given class of RG behaviors. No psychological assumptions are involved.

Determining a subject's behavior potential, however, necessarily depends on what can be assumed about the mechanisms that govern human behavior. The basic idea goes like this: Given any familiar class of RG behaviors, like the class of addition tasks, we can usually identify those rules (algorithms) which the subjects in question are likely to use in solving the problems. We do not automatically know which aspects of these algorithms any given subject is capable of, however. To find out, we must test the subject. But on which instances is he to be tested -- how are they determined? The standard approach, of course, is just to select a random sample of test instances and then make probabilistic predictions about future performance on other instances in the class.

This approach is rejected in favor of systematic selection of test instances and deterministic prediction on individual items. To see how this can be accomplished, we first note that every algorithm for solving a given class of (RG) tasks can be represented by a directed graph (see Scandura, forthcoming). For example, the task of generating the next numeral in Base Three Arithmetic can be represented as follows.

Sample Stimuli Responses

0 → 1
1 → 2
2 → 10
10 → 11

Total Graph

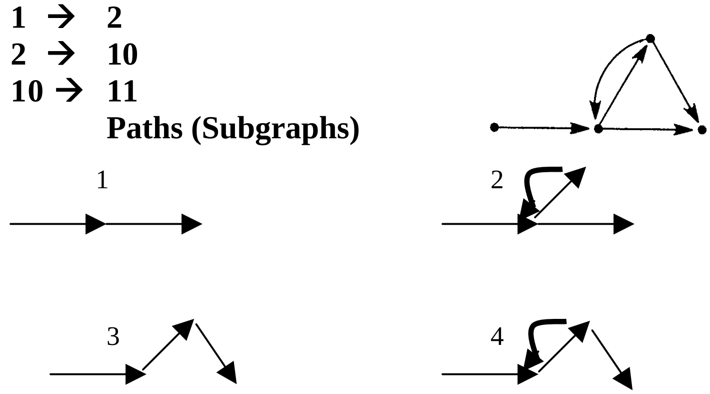


Fig. 2. Sample stimuli and responses for the task of generating the next numeral in Base Three Arithmetic, together with the (total) graph of a procedure for generating the behavior, and four graphs representing the four behaviorally distinguishable paths through this procedure.

In Figure 2, the arcs correspond to rules, which are assumed to act in ***atomic*** fashion. That is, success on any one instance of such a rule is tantamount to success on any other, and similarly for failure. We have obtained sufficient empirical evidence over the past seven years to demonstrate the existence of such rules -- in a wide variety of situations (e.g., Scandura, 1966d, 1969a). The points correspond to branching rules, that is, decisions, which must be made in carrying out the algorithm on particular test instances.

The ***subgraphs*** at the bottom of Fig. 2 correspond to the four possible paths through this procedure, which may be used in solving particular problems. Since the constituent rules are all atomic, it follows that each of these paths also acts in atomic fashion. Hence, to determine the behavior potential of a given subject with respect to this algorithm, we need only select one test instance for each path. In this case, the base-three stimulus (response) numerals, 101 (102), 2 (10), 112 (120), and 222 (1000), correspond respectively to the four possible paths. Accordingly, the behavior potential of a given subject on this class of tasks can be uniquely specified by his performance on just these four test instances -- as long as the atomic assumption is valid. (Hence, the assessment is conditional.) Any other set of four stimulus representatives of these paths, of course, would do equally well. Although its role was hidden in describing this method of assessing behavior potential, the methods' validity depends directly on the simple performance mechanism. According to this mechanism, if a subject has a particular path available for solving a given task, then he will use it and use it consistently on all instances to which it applies. That is, of course, assuming that the subject's goal remains the same.

None of this is idle theoretical speculation. Over the past several months one of my students, John Durnin, has collected a good deal of evidence which provides support which goes far beyond the bounds of what is normally considered sufficient evidence. In a total of 204 predictions utilizing a variety of tasks and subjects of greatly differing abilities and grade levels (from the preschool through graduate school), we have had a ***grand total of seven errors in***

prediction. A sample of this data is given in Table 1 for a procedure involving eight paths.

Let us next consider what happens when a subject has *not* explicitly - learned a rule for achieving a given goal. In this case, the subject has a problem in the classical sense -- a problem situation, a goal, and a barrier between them.

The major theoretical problem is to explain what happens when a subject is confronted with such a situation. If the problem can be formulated in a way that lends itself to prediction, so much the better. Why certain people are able to solve some problems for which they have never learned a specific rule, whereas others cannot, is a question of paramount interest. We want to know exactly what is involved, and why subjects perform as they do.

TABLE 1

Paths	College Student A		College Student B		H.S. Student A		H.S. Student B	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
1	+	+	+	+	+	+	+	+
2	+	+	+	+	+	+	+	+
3	+	+	+	+	-	-	-	-
4	+	+	+	+	+	+	-	-
5	+	+	+	+	+	+	-	-
6	+	+	+	+	-	-	-	-
7	+	+	+	+	-	-	-	-
8	+	+	-	+	-	-	-	-

"+" indicates correct response, *"-" indicates incorrect response.*

As a first approximation at least, it again appears that a very simple mechanism may suffice, This mechanism may be framed as a hypothesis as follows :

(B) Given a goal situation for which the subject does not have a learned rule immediately available, control temporarily shifts to the higher order goal of deriving a procedure, which does satisfy the original goal condition.

With the higher-order goal in force, the subject presumably selects from among the available and relevant higher order rules in the same way as he would with any other goal. Furthermore, where no such ***higher order*** rules are available, one might suppose that control would revert to still higher order goals. Theoretically, this process could continue indefinitely, but I suspect that a subject would tire of it, or run out of higher order rules, as quickly as would a theorist attempting to describe what is happening.

To complete things, we need a third hypothesis which allows control to revert back to the original goal once the higher order goal has been satisfied. We can state this as follows:

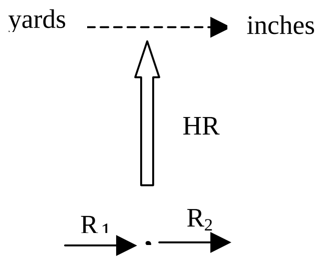
(C) If the higher order goal has been satisfied, control reverts back to the original goal.

When we say that a higher-order goal has been satisfied, of course, what we mean is that some new rule has been derived, such that that rule, when applied to the stimulus situation, satisfies the original goal criterion.

Although implicit in what has been said, it is important to note that each of the hypothesized mechanisms is assumed to work at all levels. For example, hypothesis (A) applies in higher order goal situations as well as in simple ones.

These assumptions provide an adequate basis for generating predictions in a wide variety of problem solving situations. Suppose, for example, that the problem posed to a subject is to convert a given number of yards into inches. Consider two possible ways in which a subject might solve the problem. The first is to simply know, and have available, a rule for converting yards directly into inches: "*Multiply the number of yards by thirty-six*". In this case, the subject need only apply the rule according to hypothesis (A). The other way is more interesting, and involves all the mechanisms described above. Here, we assume that the subject has mastered one rule for converting *yards* into *feet*, and another for converting *feet* into *inches*. The subject is also assumed to have mastered a ***higher order rule****, which* allows him to combine learned rules (in which the output of one matches the input of the other as is the case, for example, with rules for converting yards into feet and feet into inches) into single composite rules.

In a situation of this sort, the subject does not have an applicable rule immediately available, and, hence, according to hypothesis (B), he automatically adopts the higher order goal of deriving such a procedure. Then, according to the simple performance hypothesis (A), the subject selects the higher order composition rule and applies it to the rules for converting yards into feet and feet into inches. This yields a new composite rule for converting yards into inches. Next, control reverts to the original goal by hypothesis (C) and, finally, the subject applies the newly derived composite rule by hypothesis (A) to generate the desired response. This sequence of events is depicted in Fig. 3.



HYPOTHESES

- (B)
- (A) (HR applied to R_1 & R_2)
- (C)
- (A) (applied to composite of R_1 & R_2)

Fig. 3. A schematic representation of the hypothesized mechanism for problem solving. R_1 and R_2 represent rules for converting yards into feet and feet into inches, respectively. HR refers to the higher order rule for generating composite rules.

Although we are still in the process of refining our procedures and collecting more data, Lou Ackler and Chris Toy have run enough subjects under one condition to suggest that we are on the right track.

What we did was to teach each *S* how to use two simple rules, comparable to those described above (e.g., for converting yards into feet). These rules are denoted r_{11} and r_{12} in Table 2. As shown in the table we were successful in teaching these rules to all of the children in the sense that they could apply them uniformly well to all instances (of the respective rules). Then, each subject was tested to see if he could solve a problem requiring for its solution the composite rule, denoted $r_{11} \circ r_{12}$. As shown, only one of the subjects was initially successful on this type of problem. Next, we taught the subjects with neutral materials how to combine pairs of simple rules such as the ones they had been taught. This time we were successful with all but one subject. (To accomplish this we also had to teach many of the subjects what it was they were trying to do—that is, find a rule that could be used to solve problems such as that requiring $r_{11} \circ r_{12}$ above. In short, we taught them a decision making capability for determining whether or not they had achieved the higher order goal. More details on this are given in Scandura, forthcoming.)

At this point, we taught each subject a new pair of rules (indicated by r_{21} and r_{22}) and then tested him to see if he could solve the corresponding composite problem, which required r_{21} o r_{22} for its solution. As can he seen in Table 2, all but one of the subjects succeeded on the test problem whereas only one of them had before. Furthermore, the one subject who failed was the same subject who had previously failed to learn the higher order rule when it was taught. This same pattern of teaching and testing was repeated two more times as shown, with precisely the same results.

TABLE *2*

Summary of Experimental Procedure and Results

	Age of Subject										
	6	**7**	**8**	**5**	**8**	**7**	**6**	**8**	**6**	**6**	**8**
r_{11}	+	+	+	+	+	+	+	+	+	+	+
r_{12}	+	+	+	+	+	+	+	+	+	+	+
$r_{11} \circ r_{12}$	**+**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
HR	+	-	+	+	+	+	+	+	+	+	+
r_{21}	+	+	+	+	+	+	+	+	+	+	+
r_{22}	+	+	+	+	+	+	+	+	+	+	+
$r_{21} \circ r_{22}$	+	-	+	+	+	+	+	+	+	+	+
r_{31}	+	+	+	+	+	+	+	-	+	+	+
r_{32}	+	+	+	+	+	+	+	-	+	+	+
$r_{31} \circ r_{32}$	+	-	+	+	+	+	+	-	+	+	+
r_{41}	+	+	+	+	+	+	+	+	+	+	+
r_{42}	+	+	+	+	+	+	+	-	-	+	+
$r_{41} \circ r_{42}$	+	-	+	+	+	+	+	-	+	+	+

Note: "+" indicates S reached criterion. "-" indicates S did not reach criterion.
Subjects were only tested (without training) on **composite rules in bold**.

While no empirical data are available, it has been possible to analyze a number of other, more complicated problem situations in very much the same way (Scandura, forthcoming), including problems taken from Polya's (1962) pioneering yet atheoretical discussion of mathematical problem-solving. This includes taking into account the role of heuristics. A very similar type of analysis can also be applied to discovery learning, and, indeed, even to simple association learning (Scandura, forthcoming). The situation is very much like problem solving in which there are a number of simple problems presented in sequence, rather than just one. It would be misleading to imply, however, that this is a routine undertaking. To the contrary, it seems to require a good deal of experience, familiarity with the subject matter, and good intuition about how Ss actually do things. Most important, it usually takes time to come up with a viable analysis **[2006 update: anticipating Structural Analysis]**. Nonetheless, I am satisfied that this can be done in principle; what remains is to test these analyses empirically to see if the three hypotheses introduced above are sufficient to account for the performance of actual Ss (under the idealized conditions required by the theory).

The important point of all this is that learning can be viewed as a problem-solving process. Subjects learn as a result of being exposed to problem situations, which require that they combine available rules in new ways. Once a problem has been solved, however, no further learning is assumed to take place upon repeated presentations of similar problems. In that ease, the subject simply applies the newly learned rule.

By systematic application of our simple principles (of performance), then, it would be possible to derive all kinds of implications about learning and performance. In particular, highly specific predictions might be made about individuals who enter the learning situation with given sets of rules and who are then subjected to particular sequences of problem situations. Such analyses would have obvious implications for instructional theory. (It must be remembered, of course, that all such predictions would necessarily be limited to empirical situations, which satisfy the conditions of level two [i.e., memory-free] theorizing.)

Suppose now that a S has more than one way of achieving a given goal and that we want to know which way he will choose. As suggested above, this problem of rule selection is basically one of motivation. To see this, we ask what theorizing about motivation involves, and how this relates to our earlier discussion. We might be tempted to define the task of motivation theory as one of explaining and/or predicting which goals subjects will adopt in given situations and let it go at that. This would not be sufficient, however, for that would not tell us where such goals come from in the first place, nor how they relate to the situation at hand.

In any given situation, the observer almost always has some idea of what a given S is trying to accomplish. Thus, for example, he may not know what sort of building an architect will design, but he can be quite sure that it will be a building, under certain circumstances at least. Similarly, he can usually be fairly certain that the next move made by a chess master will be a good one, although he may not know what the specific move will be. He can also be reasonably confident that, faced with a simple theorem, a competent mathematician will come up with a valid proof, but generally speaking, he will not know what kind of proof it will be. An analogous statement may be made about a competent fifth-grader on simple addition problems. The observer may not know, say, how quickly the sums will be given, but he will generally know that they will be correct (cf. Suppes and Groen, 1967). **[2006 Update: Automation is now accommodated via levels in ASTs.]**

Looked at in this way, the motivation theorist's task is to say something additional about what a *S* will actually do in any given situation, whether this involves explaining why the architect designed the building he did, why the chess master made his particular move, or why the mathematician used an indirect proof, or the child, a certain shortcut in addition. More generally, the key question for motivation theory is to explain (and/or predict) why the *S* took (will take) the path he did (does). (In retrospect, it appears that we have already proposed an answer to this question in that special case where the *S* has no rule immediately available for achieving the initial goal. In that case, it was hypothesized (B) that Ss adopt the higher-order goal of deriving a procedure that does satisfy the initial goal.)

The problem comes in where the *S* has *more* than one rule available for achieving the initial goal. It was assumed in this case that the *S* would use one of the available rules (Hypothesis A), but nothing was said about which one. It is my contention that the answer to this question of "which one" lies at the base of what we normally think of as motivation, especially as it is realized in structural learning and performance.

Unfortunately, space does not permit anything approaching the discussion that this problem warrants. (The problem is discussed at length in Scandura (forthcoming) and will be the subject of subsequent papers.) For present purposes, it is sufficient to assume that ***Ss are systematic in their selections***. I do not believe that people are intrinsically unpredictable, even in so complex a field as motivation.

If this is true, it would seem that perhaps one could gain insight into what a person might do in the future on the basis of what he has done in the past. But, then, don't we do just this almost every day? With experience, we begin to sense the way in which particular people are likely to behave in given situations, and may therefore tend to act accordingly. We frequently know ahead of time, for example, how the boss will react to a request for a raise, or what kind of activity Jani will engage in during free play, or what kind of homework will be left undone until last.

The task of the motivation theorist is to translate such intuitions into empirically testable hypotheses. A doctoral student, Francine Endicott, and I have been working on this problem for several months now, and at first were not particularly pleased with our results. To be sure, the data almost always supported our hypotheses in a gross probabilistic sense, but they could hardly be called deterministic. By using past selections as a guide, we have been able to do much better and have recently obtained an accuracy rate of about 85% correct predictions. What we did in these experiments was to provide each *S* with an opportunity to learn two distinct procedures (Rules A and B) in the same manner as was done in the assessment (of behavior potential) study. The stimuli were identical but the responses generated by the two procedures could easily be discriminated. After learning both procedures, each *S* was presented with a general goal, which could be satisfied by using a path of either procedure. For testing purposes, stimuli on which *S* had precisely the same choice to make between paths were viewed as equivalent. As in the assessment study, *S* was tested twice on each equivalence class. According to our assumption, it was hypothesized that *S* would select the same paths on corresponding Test 1 and Test 2 stimuli.

TABLE 3

Results of Rule Selection Study

		Test 2: Rule A	Test 2: Rule B
Test 1	Rule A	52	2
	Rule 2	13	64

The results are summarized in Table 3. This table shows that whenever a *S* selected a path of Rule A on Test 1, he almost invariably (52 times out of 54) selected a path of Rule A on Test 2. Rule B selections were consistent with the hypothesis 64 times out of 77.

To recapitulate, it should be re-emphasized that everything that has been said so far about learning, performance, and motivation only applies in situations where memory and the limited capacity of human subjects to process information do not enter. The proposed mechanisms have all assumed an information processor with an essentially unlimited ability to process information, and with perfect memory for previously acquired knowledge.

This definitely clues ***not*** imply that the theorizing so far is of little value. That conclusion would be wrong on at least two counts. First, there are many practical situations in structural learning where memory is of minimal concern. In problem solving, for example, the *S* is almost always given all of the paper, pencils, and other memory aids that he needs. Typically, we also do our best to insure that the necessary lower-order rules are readily available, even to the extent of making textbooks available. The concern is generally with whether or not the individual can integrate available knowledge to solve problems. Considerations such as whether he can do it in his head or not, time to solution, and so on, are of secondary concern (cf. Scandura, 1967). Second, questions of memory can usually be eliminated in experimentation by insuring that relevant rules and memory aids are available to the subject. This can normally be accomplished by training.

TOWARD A THEORY OF MEMORY & INFORMATION PROCESSING

Any fully adequate theory of structural learning, of course, must deal with more than just idealized behavior. In particular, such a theory must as a minimum take both (long-term) memory and information processing into account. Insofar as memory is concerned, there must be mechanisms for storing and retrieving information in long-term memory. In addition, hypotheses are needed to deal with the processing of information, and particularly the limited amount of information that human beings can process at any given time. Thus, for example, an adequate theory should make it possible to account for the differential ability of human subjects to perform mental arithmetic, even where the Ss know perfectly well how to compute.

In most theorizing about memory, there has been an unfortunate tendency to confound these two kinds of problem. Much of the more recent work, for example, has been heavily influenced by a technique for measuring retention, which was introduced by Peterson and Peterson (1959). The basic idea of their experiment was (1) to present CCC nonsense syllables, (2) have the S count back ward by threes or fours, and (3) test him to see if he could remember the given nonsense syllable after some intervening period ranging from about 0 to 30 seconds, Contrary to the then prevailing expectation of most psychologists, they found that retention decreased rapidly over this short period, and psychologists had a new game to play. The basic paradigm is still in wide use today.

The difficulty with this type of study is that it does not distinguish operationally between mechanisms associated with the storage and retrieval of information from long-term store, on the one hand, and the limited ability of human Ss to process information, on the other. Thus, in a Peterson and Peterson type experiment, a *S* may attempt to retain a nonsense syllable, say, either by continuing to process the information, by a process typically referred to as rehearsal, or by storing it in long-term memory. Under these circumstances, it is difficult to say anything definitive about either type of mechanism as a result of the experimental data obtained.

For present purposes, it would obviously be desirable to have a theory of structural learning which deals with the two kinds of problems raised above, and which at the same time is compatible with our earlier theorizing. Specifically, we need to ask how the memory-free theory may be supplemented so as to take both (long-term) memory and information processing into account. No hard answers, unfortunately, are available at the present time, particularly insofar as memory is concerned. All that can be done here is to sketch one approach to the problem that seems to hold some promise.

Insofar as long-term memory is concerned, nothing basically new seems to be required; the basic mechanisms of the idealized theory appear to be adequate as they are. What does need to be done is to increase the domain of applicability of these mechanisms. Specifically, rules are needed for ***storing*** and for ***retrieving*** information. Storing rules act on observables, as do other rules, but the outputs of such rules are strictly internal. Retrieving rules, on the other hand, act on stored (internal) units of knowledge (which serve as stimuli) and generate observables.

What these rules do is to relate new knowledge with knowledge that has been acquired previously. For example, in order to store (i.e., give the ***correct meaning*** to) the statement, "*any function continuous on a closed interval is uniformly continuous*", *S* must clearly know ahead of time what continuous functions, closed intervals, and uniformly continuous functions are. The storing rule combines these meanings into a new meaning, which corresponds to the statement taken as a whole (Scandura, 1970b). This has the effect of tying in (i.e., locating) the desired meaning with previously acquired knowledge.

Retrieval rules, on the other hand, provide the subject with a basis for regenerating knowledge from the recall cues—for example, from a statement like "*what can be said about functions which are continuous on closed intervals?*"

Difficulties in recall are explained either in terms of what is (or is not) stored or the availability (or lack) of appropriate retrieval rules. For example, if a *S* memorizes a statement like that given above, without understanding it, and is asked at recall to explain the idea in his own words, then no one would reasonably expect the *S* to succeed. Similarly, if the *S* stored the meaning and was asked to repeat the statement verbatim, he would not likely be able to do more than come up with a rough paraphrase. Without adequate storing rules in the first place, of course, recall would be completely lacking according to this view. Even where a *S* has definitely learned (stored) something, he may still not be able to "recall" it because he lacks the necessary retrieval rules. Young children, for example, are frequently able to do things, like solve arithmetic problems, indicating that they have learned how, but be quite incapable of describing what they did. Although we cannot go into the problem here, this sort of analysis appears to provide relatively simple explanations for a number of well-known phenomena, such as retroactive inhibition and reminiscence. (Details are given in Scandura, forthcoming.)

It should be emphasized, however, that the theory is essentially deterministic, and applies only where one is dealing with highly structured materials, where one can make reasonable assumptions about the kinds of rules used in storage and retrieval. The theory is not designed to handle data from typical short-term memory experiments. (Even here, however, it can be suggestive (Scandura, forthcoming).) Rather, the theory calls for quite a different kind of memory experimentation -- experimentation with relatively complex and more highly structured materials, where explicit attention is given to the goal conditions imposed on the S and the kinds of storage and retrieval capabilities with which he enters the situation.

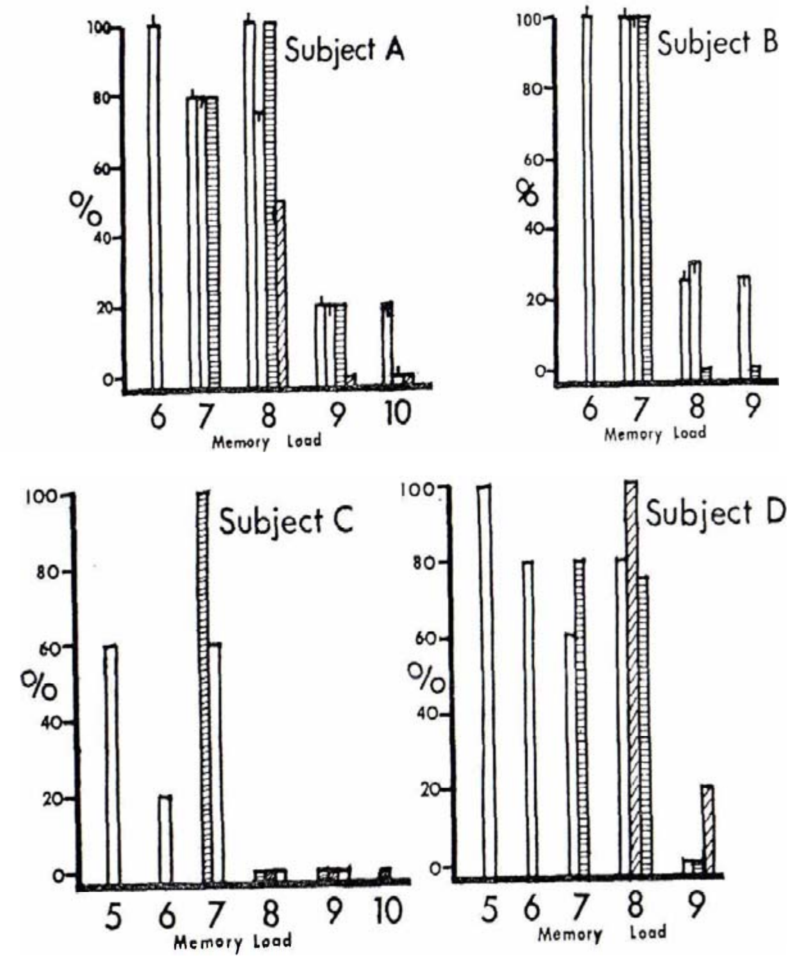
The only fundamentally new hypothesis involves ***information, processing***. The basic assumption is that each individual subject has a ***fixed finite capacity*** for processing information. While this capacity may vary somewhat over individuals, it is assumed to be of the order, 7 ± 2 ***units*** of information. (The term ***bits*** of information is avoided since it implies a connection with information theory which is not intended.) The classic work of Miller (1956) is obviously related, but his results were based largely on averages and relatively simple tasks. (It is not clear just how, or whether, Miller's work on card sorting is related to information processing in the sense described.) It is important that these results be extended to individuals and generalized to more complex tasks. We assume that this capacity remains constant for individuals, whether one is adding numbers, storing information, or solving problems -- as well as in repeating strings of digits, as Miller had his subjects do.

Demonstrating this to be the case, however, is not a trivial problem. Another student, Donald Voorhies, and I have been working on the problem trying to refine our experimental procedures to the point where we can get a fair test of the hypothesis. We still have some way to go but the results of our pilot data were reasonably good almost from the beginning and this, in retrospect, is probably what kept us going. In each case, after a certain degree of complexity was reached there was a sharp "drop off" in performance. Even this, however, required meticulous attention to detail. First, the procedure in question had to be broken down into its basic states and operators. Space does not permit going into details (this will be done elsewhere), but the basic idea is closely related to Suppes' (1969) S-R characterization of finite automata and my reinterpretation in terms of rules (Scandura, 1970b). Second, we had to get each *S* to use this procedure exactly as prescribed. The smallest of deviations could materially affect the results.

Another major roadblock was that we could not tell ahead of time with a new task where the “dropoff” would occur. What was needed was a general scheme for calculating memory load for any given rule – but developing one did not prove to be a simple task. We have recently come up with something that seems promising, however, and about a week ago, our data reached about 80% level of predictability, which may be about as good as can reasonably be expected with this type of task.[3]

Unfortunately, we have so far been unable to test any of our volunteer *Ss* (graduate students) on all of the tasks we have devised. The data available at the time of this writing are summarized in Fig. 4.

[3] Basically, the technique involves calculating for each step of the given algorithm: (1) the number of states needed to determine future states, (2) the number of operators needed to determine future operators, and (3) the number of subsets of the needed states and operators which must be distinguished in completing the procedure. Details will be published in a future article.



─ Repeating Digits

┴ Repeating Digits -- 1 after

┬ Repeating Digits -- 1 before

═ Addition - no carry

// Addition – carry

FIG. 4. The performance of four subjects on the indicated tasks with percentage of perfect responses plotted against memory load. For comparative purposes, repeating n digits had a calculated memory load of n; repeating n digits and then saying "1" had a memory load of n + 1; repeating n digits after saying "1" had a memory load of n + 2; addition of two 2, 3, and 4 digit numbers without carrying had memory loads of 7, 8, and 9, respectively; with carrying, the memory loads were 8, 9, and 10.

CONCLUDING COMMENTS AND IMPLICATIONS

The foundations of three partial theories of structural learning have been described and some relevant data have been reported. First, a partial theory of structured knowledge was proposed, in which it was argued that the knowledge had by any given *S* may be characterized in terms of a finite set of rules. By allowing rules to operate on other rules (in the set), it was shown how new rules could be generated. Examples were also given to show how these new rules, in turn, could account for creative behavior. With the addition of several performance assumptions, this theory was extended so as to account for learning, performance and motivation under idealized conditions where behavior is unencumbered by memory. Finally, we outlined how memory and information processing might be dealt with, and reported some preliminary data in favor of our main hypothesis. Even the most encompassing theory, however, does not deal with a number of behavioral phenomena, specifically the ultra short-term after images reported by Sperling (1960), Averbach and Coriell (1961), and others. Whether the theory might be extended further to account for these phenomena is difficult to say. But, in any case, this might well be left until later given the large number of questions raised by the theory as it presently exists.

The theory itself represents a sharp departure from existing theories of cognitive behavior, although it does have some things in common with existent competence and information-processing theories. The differences even here, however, are not minor, but have a fundamental effect, both on theoretical adequacy and on the very kinds of empirical questions one asks. Probably the most basic departure is the idea of introducing different levels of empiricism, and the possibility of deterministic theorizing at each of these levels. According to this view, it is possible to do behaviorally relevant empirical research at at least three quite distinct levels. Although all competence models, such as those proposed by Chomsky in linguistics, purport to deal with knowledge, concern traditionally has been limited primarily to the so-called mature speaker or hearer who effectively knows all there is to know about the language. In the present formulation, it is just as reasonable to talk about the knowledge had by different individuals, naive ones as well as mature. This is an extremely important characteristic in dealing with subject matters like mathematics, science, or even language, where knowledge is not a static thing, but grows with experience.

An even more basic departure is allowing rules to act on other rules. This seems to me to be the only real hope we have at present with which to account for creative behavior within an algorithmic framework. There is a good deal more detailed work to be done, but so far the main roadblocks appear to be ones of detail and not of principle.

The distinction between idealized theorizing and related empiricism, on the one hand, and the more complete theory, including memory, on the other, is equally basic. By ignoring the effects of memory and information processing capacity, for example, it has been possible to deal with quite complex behavior, such as problem solving and motivation, in a very precise way -- and even more important, in near deterministic fashion. In addition, the proposed mechanisms of memory and information processing are simpler and potentially more precise than those of existing information processing theories. Furthermore, the theory is designed primarily to apply to memory and information processing with complex structured materials, and not just with the short-term memory of lists of nonsense syllables, simple words, or sentences, as has been the case with most modern memory research.

Let me finally mention some of the most promising areas of application of this work in education. Insofar as curriculum construction is concerned, it is sufficient to simply reemphasize that it is a small conceptual step from characterizing knowledge of individual *Ss* in terms of rules to characterizing curricula in terms of operational objectives. Unlike the current list type curricula (Lipson, 1967), however, explicit attention might be given to the identification of higher order relationships. As simple as this change may seem, its importance cannot be overemphasized. It makes it possible not only to build a good deal of transfer potential directly into a curriculum, but also to capture, I think, what subject matter specialists almost uniformly feel has been missing in current curricula of the operational objectives variety—the creative element. We have a pilot project underway at Penn at this time, in which we are attempting to apply these ideas to teaching mathematics to elementary school teachers. It is too soon to say how things will actually turn out, but so far things have been going extremely well and we hope that we will be able to teach more sophisticated mathematics in this way, and to teach it more effectively.

A second major implication has to do with testing, particularly that sort of testing used to determine mastery on the objectives that go to make up curricula of the sort indicated. Here, the groundwork has been all but completed, and application would seem to be a rather straightforward operation. In fact, we are actually utilizing these ideas in another small-scale developmental project aimed at diagnosing difficulties urban youngsters are having with the basic arithmetical skills. Another phase of this project has to do with remediation of these difficulties. In this regard, we are using our own home-grown version of hierarchy construction. What we do, in effect, is simply to identify the particular algorithm (rule) we want to teach the child, and break it down into atomic sub-rules. Each sub-rule, in turn, is broken down in the same way, until we reach a level where we can be sure that all of our subjects have all the necessary competencies. This breakdown corresponds directly to the hierarchies obtained in the usual manner by asking Gagne's (1962) often quoted question, 'What must the learner be able to do in order to do such-and-such?" Unlike the traditional approach, however, ours provides a natural basis for constructing ***alternative*** hierarchies (since any number of procedures may be used to generate the same class of behaviors). Possibilities also exist in such areas as teaching problem solving, but our work to date has been limited to testing basic hypotheses.

REFERENCES

AVERBACK, E. and CORIELL, A. S. Short-term memory in vision. *Bell System Technical Journal,* 1961, **40,** 309-328.

CHOMSKY, N. *Syntactic structures.* The Hague : Mouton, 1957.

CHOMSKY, N. *Language and Lind.* New York : Harcourt, Brace, and World, 1968.

GAGNE, It. M. The acquisition of knowledge. *Psychological Review,* 1962, 69, 355-365.

GAGNE, R. *M. The conditions of learning.* New York : Holt, Rinehart & Winston, 1965,

GODEL, KURT. Uber formal unentscheidbare Satze der *Principia Mathematica* und verwandter Systeme. *Monatschrifte fur Mathematik and Physik,* 1931, 38, 173-198.

LIPSON, *J.* The development of a mathematical curriculum for individualized instruction. In *J. M.* Scandura (Ed.), *Research in mathematics education.* Washington: National Council of Teachers of Mathematics, 1967.

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.

MILLER, G. A., and CHOMSKY, N. Finitary models of language users. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology.* Vol. 2. New York : Wiley, 1963. Pp. 419-491.

MILLER, G. A., GALANTER, E., and PRIBRAM, K. H. *Plans and the structure of behaviour.* New York: Holt, 1960.

NELSON, R, *J. Introduction to automata.* New York: 1968.

PETERSON, L, R., and PETERSON, M. J. Short-terrn retention of individual verbal items, *Journal of Experimental Psychology,* 1959, 58, 193-198.

POLYA, G. *Mathematical discovery.* New York: Wiley, 1962.

SCANDURA, *J.* M. Precision in research on mathematics learning: The emerging field of psycho-mathematics, *Journal of Research in Science Teaching,* 1966, 4, 253-274.

SCANDURA, J. M. An alternative approach to research on mathematics learning—a reaction to Suppes's paper. *Journal of Research and Development in Education,* 1967, 1, 33-43.

SCANDURA, J. M. New directions for theory and research on rule learning: I. A set-function language. *Acta Paychologica,* 1968, 28, 301-321.

SCANDURA, J. M. New directions for theory and research on rule learning: II. Empirical research. *Acta Psychologica,* 1969, 29, 101-133.

SCANDURA, J, M. A theory of mathematical knowledge: Can rules account for creative behavior? University of Pennsylvania, Mathematics Education Research Group: Technical Report No. 52, 1970. (a) (to appear in *Journal for Research in Mathematics Education).*

SCANDURA, J. M. The role of rules in behavior: Toward an operational definition of what (rule) is learned. *Psychological Review.* 1970, 77, 516-533. (b)

SCANDURA, J. M. *Mathematics and structural learning,* Englewood Cliffs, NJ: Prentice-Hall, forthcoming. **[This reference is to Scandura, J.M. Structural Learning I: Theory & Research, NY/London: Gordon & Breach, 1973]**

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs,* 1960, 74 (11).

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

SUPPES, P. and GROEN, G. Some counting models for first-grade performance data on simple addition facts. In J.M. Scandura (Eds.), *Research in mathematics education*. Washington: National Council of Teachers of Mathematics, 1967, 35-43.

WHOLWILL, J.F. The emerging discipline of environmental psychology. *American Psychologist*, 1970, 25, 303-312.

FOOTNOTES

[1] Higher order rules on rules are common in various branches of mathematics, where they go under the label of functions on functions, but the idea seems not to have generally pervaded either computer science or formal linguistics. In formal linguistics, for example, where the goals closely parallel ours, no one seems to have seriously proposed the use of higher order rules, The closest linguists have come in this regard has been to introduce the notion of grammatical *transformation* between phrase markers (Chomsky, 1957). Rather than higher order rules, transformations correspond more closely to what we have here called *relational* rules (see Scandura, forthcoming).

There are two good reasons why this has probably not been done in the past. First, even grammatical transformations have so far resisted mathematical treatment (Nelson, 1968), and, second, no existing approach to psychology known by the writer provides any real motivation for introducing them. Gagne's (1965) view of problem solving in terms of rules and Miller, Galanter, and Pribram's (1960) TOTE hierarchies come closest.

This is unfortunate, since there is a very simple and intuitively sound reason for allowing rules to operate on (classes of) rules. The main one is just this: There is a very simple and intuitively compelling performance mechanism by which higher and lower order rules may be combined so as to generate completely new kinds of behavior. Furthermore, as shown in the next section, some empirical support for this mechanism has already been obtained.

[2] Extrapolating from Godels (1931) incompleteness theorem, and translating the results into behavioral terms it seems reasonable to conclude that there will always be behaviors which cannot he generated by *composing* any finite set of rules. Even if we allow rules to operate on rules, there will undoubtedly always he certain behaviors that no individual will be capable of performing, no matter how knowledgeable he may be.

[3] Basically, the technique involves calculating for each step of the given algorithm: (1) the number of states needed to determine future states, (2) the number of operators needed to determine future operators, and (3) the number of subsets of the needed states and operators which must be distinguished in completing the procedure. Details will be published in a future article.