DOCUMENT RESUME

ED 054 952	24	SE 012 392
AUTHOR	Scandura, Joseph M.	
TITLE	Mathematics and Structural Learning. Final Report.	
INSTITUTION	Pennsylvania Univ., Philadelphia. Graduate School of Education.	
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.	
BUREAU NO	BR-9-0277	
PUB DATE	May 71	
GRANT	OEG-2-9-480277-1045(010)	
NOTE	98p.	
EDRS PRICE	MF-$0.65 HC-$3.29	
DESCRIPTORS	Behavior; Creativity; Educational Psychology; Learning Motivation; Learning Processes; *Learning Theories; *Mathematics Education; Memory; Problem Solving; *Teacher Education	

ABSTRACT

This report contains four papers describing research based on the view of mathematical knowledge as a hierarchy of "rules." The first paper: "The Role of Rules in Behavior" was abstracted in ED 040 036 (October 1970). The second paper: "A Theory of Mathematical Knowledge" defends the thesis that rules are the basic building blocks of mathematical knowledge. These rules operate at different levels, for example: addition and subraction at one level, the idea of inverse operations at a higher level. Mathematical creativity then consists of combining rules to produce new results. The third paper: "Deterministic Theorizing in Structural Learning" describes experiments based on this hypothesis, in which a subject's performance on certain tasks is predicted with virtual certainty from his performance on the component rules. The role of the memory is also discussed, and experiments are described supporting the view of the mind as an information processor with a fixed capacity. The fourth paper: "A Research Basis for Teacher Education" argues that teachers need to know more of the aims of the mathematics they teach, more of the way mathematical knowledge is structured, and more of the way students learn. The author's research on structural learning (as described in the second and third papers) is then summarized, and possible developments are outlined. (MM)

BR 9-0277
PA 24
SE

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED054952

STRUCTURAL LEARNING SERIES

Mathematics
Education
Research
Group

Graduate School of Education,

UNIVERSITY of PENNSYLVANIA
Philadelphia

1

SE 012 392

ERIC

FINAL REPORT

Grant No. OEG-2-9-480277-1045(010)

MATHEMATICS AND STRUCTURAL LEARNING

Joseph M. Scandura

Graduate School of Education
University of Pennsylvania
Philadelphia, Pa. 19104

May 1971

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report
Grant No. OEG-2-9-480277-1045(010)

Mathematics and Structural Learning

Joseph M. Scandura

University of Pennsylvania

Philadelphia, Pennsylvania

May 1971

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TABLE OF CONTENTS

ERIC

SUMMARY OF

MATHEMATICS AND STRUCTURAL LEARNING

INTRODUCTION AND OBJECTIVES

This research was concerned with two basic problems.

1. The first objective can be stated as a question, "How can mathematical knowledge be characterized in a way which is at once behaviorally significant and compatible with what is known about mathematical structures?" We were interested in pursuing this research at two levels.

(a) We wanted to further clarify both the nature of rules and their role in behavior.

(b) We also wanted to explore the possibility of extending our rule formulation to provide a basis for characterizing more complex mathematical knowledge.

2. The second major objective involved the development of a psychological theory — of operational definitions and theoretical assumptions which were compatible with our rule based characterizations. More specifically, we wanted to consider the following problems.

(a) How can one operationally define what rule an individual is using in terms of the behavior it elicits? We had already proposed a preliminary version of such a definition and Levine had done this for the special case of discrimination learning but many details still needed to be worked out.

(b) The learner often has available several ways of accomplishing a particular task and why he uses the procedure he does use is not at all clear. Building on some experimental research we had already completed, we wanted to come up with some fundamental proposals that would be suitable for detailed experimental testing.

(c) We hoped to explore the fundamental question of how existing knowledge is combined to make new behavior possible.

To the extent that time and funds allowed, we also planned to conduct some pilot work to test our theoretical ideas.

RESEARCH OUTCOMES AND IMPLICATIONS

While we did not initially expect to fully achieve our aims during the course of this short project, things progressed more rapidly than we had dared hope.

Part 1.

An entire paper, "Role of Rules in Behavior: Toward an Operational Definition of What (Rule) is learned," now published in the Psychological Review, is devoted to the first problem. In this paper, a precise formulation of the notion of a rule in terms of sets and functions -- the Set Function Language (SFL) -- is proposed. In particular, the extension of a rule is viewed as a function, or set of S-R pairs. The rule itself involves a domain, an operation, and a range. It is argued that this molar formulation cannot be captured by networks of associations unless one allows associations to act on (other) associations. This formulation is then used as a basis for showing how rules are involved in decoding and encoding, symbol and icon reference, and higher order relationships. Decoding and encoding are shown to involve insertion into and extraction from classes, respectively. Reference is viewed in terms of rules which map equivalence classes of signs into the classes of entities denoted by these signs. Symbols are shown to involve arbitrary reference, whereas icons retain properties in common with the entities they denote. Higher order relationships are then expressed as higher order rules on rules. This is a direct generalization of associations on associations.

Furthermore, a partial solution is posed to objective (2a) -- the vexing problem of "what (rule) is learned." Given a rule-governed class of behaviors, "what is learned" is defined as the class of rules which provides an accurate account of test data. Empirical evidence is presented for a simple performance hypothesis based on this definition.

There are three major directions in which future research might proceed. First, the rule formulation (SFL) itself undoubtedly can be further improved. While we feel reasonably confident that the basic ideas presented in this paper would hold up under further analysis, additional detail must be added -- but only as much as is absolutely necessary to deal with behaviorally relevant aspects of the rule construct.

Second, the SFL might profitably be used as an analytical tool to help clarify what is involved in many kinds of structured learning and performance. Most of the SFL-based research conducted to date has concentrated on an analysis of what is being presented, the nature of the required outputs, what is being learned, and the interrelationships between them. While such analyses can, at least to some extent, be undertaken without the use of the SFL, or for that matter any other scientific language, the SFL seems to provide a useful framework for putting things into perspective and for helping to clarify difficult points. In the author's research a number of questions have been asked on mathematics learning which seem not to have been asked previously in any serious way. For example, we have found that what is learned in mathematical discovery can sometimes be identified and presented by exposition with equivalent results. Similarly, we were led, on the basis of an earlier finding, to the question of what in the statement of a mathematical rule leads to extrascope transfer.

The SFL needs to be applied more systematically in studies involving subject matters other than mathematics and, in particular, we need to determine where the SFL might profitably be used to formulate research and where not. There is reason to believe that the SFL may be applicable only to the extent that the classes of overt stimuli and responses involved can be viewed as discrete (i.e., nonoverlapping) and exhaustive entities. While these requirements are met throughout much of mathematics and other structured knowledge, this may not be the case in such areas as social studies, poetry, and even language, where synonymy does not necessarily imply equivalence. It is hoped that other investigators will apply the SFL to a wider range of tasks and thereby help to clarify further its relative strengths and weaknesses.

Third, theoretical assumptions need to be made and their implications need to be drawn out. Although this paper is concerned primarily with describing a new scientific language, it was not possible to completely avoid reference to theoretical assumptions. Thus, the proposed operational definition of "what is learned" would be behaviorally meaningless without an application assumption. Fortunately, there is considerable empirical support for the idea. While such an assumption is clearly not sufficient for a theory of structural learning, it might nonetheless come to play a central role. Whatever form additional theoretical assumptions might take, it seems almost certain that they would be more compatible with cognitive (rule-based) notions than with those based on neo-associationism. Nonetheless, any complete theory of structural learning will undoubtedly require reference to such things as the limited capacity of human subjects to process information. Without recourse to some such physiological capacity, I can see no way in which to explain memory or other aspects of information processing.

<u>Part 2</u>.

An effective answer to objective (1b) is provided by a second paper, entitled "A Theory of Mathematical Knowledge: Can Rules Account for Creative Behavior?"

In this paper, we proposed and defended the rather strong thesis that rules <u>are</u> the basic building blocks of <u>all</u> mathematical knowledge.

The main purpose of the paper was to indicate how complex mathematical behavior might be accounted for in terms of finite rule sets.

Every mathematical system consists of one or more basic sets of elements, together with one or more operations and/or relations and/or distinguished elements of the basic sets. By capitalizing on certain logical equivalences it is possible to reduce the characterizing elements to one basic set and one or more relations. Consider a simple example -- the system whose basic set consists of three "undefined" elements A,B,C, denoted $\{A,B,C\}$, with A being distinguished in the sense that it serves as an "identity," and whose defining relation is $\bigcirc = \{(A,A) \rightarrow A,\ (A,B) \rightarrow B,\ (B,A) \rightarrow B,\ (A,C) \rightarrow C,\ (C,A) \rightarrow C,\ (B,B) \rightarrow C,\ (C,C) \rightarrow B,\ (B,C) \rightarrow A,\ (C,B) \rightarrow A\}$.

What may be called an embodiment of a mathematical system results on assignment of meaning to the undefined elements. Thus, in the example just cited, the undefined terms might correspond to certain rotations with A corresponding to a rotation of 0°; B, to a rotation of 120°; and C to a rotation of 240°. In this case, the operation would simply be "followed by." For example, a rotation of 120° followed by one of 240° results in the same action as a rotation of 0°.

What kinds of behavior are implied by knowing systems and embodiments of this sort? And, how can such behaviors be accounted for in terms of rules?

First, knowing a system certainly implies the ability to compute within the system. Thus, for example, given the pair, A,B, the "knower" should be able to give the "sum," B. He should also be able to do more complex computations, like $((A \bigcirc B) \bigcirc A) \bigcirc C \rightarrow (B \bigcirc A) \bigcirc C \rightarrow B \bigcirc C \rightarrow A$, which involve combining individual facts (i.e., associations). In addition, the knower should be able to give "differences," i.e., given the sum and one of the "addends," he should be able to generate the other addend.

If these were the only kinds of behavior to be accounted for one could simply list the facts (rules) involved. But clearly any reasonable interpretation of "knowing a system" must also deal with relationships as well. For example, mastery of a system would surely include the ability to generate the subtraction (difference) rule from the addition rule, and vice versa. Knowing that $B + C = A$, for example, should be a sufficient basis for generating the corresponding subtraction fact, $A - B = C$.

Relational rules of this sort provide a simple way to account for such behaviors. Thus, instead of listing all of the subtraction facts separately it would be sufficient to know the addition facts together with the relational rule. That is, assuming, as is traditional in formal linguistics, that individual rules can be composed -- performed in succession.

The obvious way to account for such relationships -- the way taken by curriculum developers of the operational objectives persuasion -- is to simply add more rules to the characterization. There are, however, major problems with this approach. For one thing, listing a new rule for each kind of relationship would have a post hoc flavor not likely to add much in the way of understanding more creative behavior. For each new system (of the same type) considered, for example, there would be a new relational rule for each one in the original system. Even granting the economy obtained by eliminating inverses, and the like, the number of rules could grow large very fast. This would not be bad in itself assuming that this is the best one could do. The important question, however, is: Can one come up with a more efficient account which is at the same time more powerful -- and which allows for some measure of creative behavior?

To answer this question, first note that knowing how one or more systems are related to a given one may provide a basis for knowing how to compute in the new systems given how to compute in the original. The relationships of interest will generally be mathematical in nature, but they need not be limited to morphisms. For example, one system may be a simple generalization of another, as with cyclic 5 and cyclic 3 groups.

Because of the way particular relationships are defined, however, this advantage will generally be of a limited sort. With homomorphisms, for example, the ability to compute in the new system applies only to the defining operations themselves and not, say, to their inverses or to relationships between the operations.

A far more powerful and parsimonious characterization results by simply allowing rules to operate, not on just ordinary stimuli, but on other rules. Such rules may be said to be acting in a higher order capacity -- or, in short, to be higher order rules. Although functions on functions are common in various branches of analysis, and their formalization is routine, the idea seems not to have pervaded formal linguistics. The closest linguists have come in this regard has been to introduce the notion of a grammatical transformation between phrase markers, which closely parallels what are here called relational rules (e.g., between addition and subtraction).

Consider what higher order rules might suggest in the present situation. Suppose that a subject has learned a higher order rule which connects each operator (rule) with its inverse. Such a rule would connect not only, say, addition of numbers with subtraction, but composition of all sorts (e.g., of permutations, rotations, rigid motions, etc.) with the corresponding inverse operations. The defining operation of each system and its inverse may be thought of as being distinct rules which are mapped one on to the other by this higher order "inverse" rule. Assume, in addition, that the subject has learned how to add in system A, the relationship (e.g., a homomorphism) between system A and system B, and also how to form the composition of arbitrary rules (in the rule set).

In this case, there are all sorts of behaviors that the (idealized) subject would be capable of. For example, he would be able to subtract, not only in system A but in system B as well. To see this, one need only observe that the subject can form the composition of the rule between systems A and B and the higher order inverse rule. This composite (higher order) rule in turn allows the subject first to generate an addition rule in system B and then to generate a subtraction rule in system B. This subtraction rule, in turn, would allow the subject to subtract. Translated into more meaningful terms, a rule set of this sort would imply such abilities as finding inverses with rigid motions given only the ability to add numbers. But, then, isn't this just what is considered as creative behavior?

To summarize, this paper deals with what it means to know an existing body of mathematics. Relatively little is said about intellectual skills of the sort that must inevitably be involved in doing real mathematics. Nonetheless, it is shown that what appears to be creative behavior might well be accounted for in terms of growing rule sets. The key idea in making this a feasible and rather attractive possibility is that of the higher order rule.

Part 3.

The third paper, "Deterministic Theorizing in Structural Learning: Three Levels of Empiricism," deals in an integrated fashion with objectives (2a), (2b), and (2c). It also reports on some pilot experiments designed to test the basic theoretical hypotheses.

The foundations of three partial theories of structural learning are described and some relevant pilot data are reported. First, a partial theory of structured knowledge is proposed, in which it is argued that the knowledge had by any given subject may be characterized in terms of a finite set of rules. By allowing rules to operate on other rules (in the set) it is shown how new rules can be generated. Examples are also given to show how these new rules, in turn, can account for creative behavior. With the addition of several performance assumptions, this theory is extended so as to account for learning, performance, and motivation under idealized conditions where behavior is unencumbered by memory. Finally, we outline how memory and information processing might be dealt with, and report some preliminary data in favor of our main hypothesis.

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 us 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 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 the memory-free theory, the main task is one of introducing mechanisms of idealized performance, learning, and motivation, thereby extending the theory of knowledge so that it 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.

The theory deals with three basic kinds of situation: 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? 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 simple as it appears, this assumption has a number of important implications. One is that it provides an adequate basis for determining what might be called a subject's behavior potential, relative to a given class of rulegoverned behaviors. Briefly by applying this assumption to assessing individual behavior potential or individualized testing, we have been able to predict a subject's second test performance on individual items with a high degree of accuracy. (Precisely how this was done is described in detail in the paper.) In a total of 204 cases, utilizing a variety of tasks and subjects of greatly differing abilities and grade levels (from the preschool through graduate school), we have been able to predict second test performance 197 times or with 97% accuracy. The results of this research could be particularly useful for constructing refined diagnostic tests in many areas of psychological and educational testing.

In the second case, the subject has not explicitly learned a rule for achieving a given goal. He 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.

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, we assume that control reverts to still higher order goals. Theoretically, this process could continue indefinitely.

To complete things, a third hypothesis is needed 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.

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. Here, we assume that

ERIC

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 which is 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.

Preliminary empirical support for these hypotheses is reported in the paper.

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 case, the subject simply applies the newly learned rule.

On the basis of these assumptions, 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.

The third case is concerned with what happens where a subject has more than one way of achieving a given goal and we want to know which way he will choose. It was assumed in this case that the subject would use one of the available rules (Hypothesis (A)), but nothing was said about which one. It is our 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.

We worked on this problem for sometime, and at first we were not particularly pleased with our results. To be sure, our pilot 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 been able to determine what rules or parts of rules a subject selected with an accuracy rate of about 85%.

To recapitulate, it should be re-emphasized that everything which 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 does not imply that the theorizing 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 subject 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. 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.

The mechanisms of memory and information processing proposed in the paper are speculative and subject to revision. Nonetheless, they 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.

Finally, we 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 subjects in terms of rules to characterizing curricula in terms of operational objectives. Unlike the current list type of curricula, 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, we 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 which 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 of the necessary competencies. This breakdown corresponds directly to the hierarchies obtained in the usual manner by asking Gagne's 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.

Part 4.

A final paper, "A Research Basis for Teacher Education," goes beyond the immediate scope of the proposed research: It is directed to professional educators and attempts to provide a broader perspective concerning the problems and their possible resolutions. More specifically, the purpose of this paper is to (1) indicate why basic research in mathematics (and subject matter) education is badly needed (2) to identify some of the kinds of information which every good mathematics teacher needs (3) to describe some of the basic research which we have under way and also to mention some of the implications of this research for further development in mathematics education and behavioral research generally, and (4) to describe some of our current developmental activities in teacher education in mathematics.

ROLE OF RULES IN BEHAVIOR: TOWARD AN OPERATIONAL DEFINITION OF WHAT (RULE) IS LEARNED[1]

JOSEPH M. SCANDURA

University of Pennsylvania

A precise formulation of the notion of a rule in terms of sets and functions is proposed. It is argued that this molar formulation cannot be captured by networks of associations unless one allows associations to act on (other) associations. This formulation is then used as a basis for showing how rules are involved in decoding and encoding, symbol and icon reference, and higher order relationships. Decoding and encoding are shown to involve insertion into and extraction from classes, respectively. Reference is viewed in terms of rules which map equivalence classes of signs into the classes of entities denoted by these signs. Symbols are shown to involve arbitrary reference, whereas icons retain properties in common with the entities they denote. Higher order relationships are then expressed as higher order rules on rules. This is a direct generalization of associations on associations. Finally, a partial solution is posed to the vexing problem of "what (rule) is learned." Given a rule-governed class of behaviors, "what is learned" is defined as the class of rules which provides an accurate account of test data. Empirical evidence is presented for a simple performance hypothesis based on this definition.

During the past few years there has been a gradual shift of emphasis in psychology from the study of simple to complex learning. Even where investigators are still working primarily with simple tasks, such as the learning of paired-associate lists, the questions being asked seem to have broader significance.

This shift has not come, however, without attendant difficulties. While existing theories are clearly inadequate for dealing with complex structural learning, there are other, even more basic, problems which have not yet been adequately resolved. In particular, there has been no scientific language with which even to talk about many of the problems. The general question of the relative efficacy of discovery and expository learning (e.g., Gagné & Brown, 1961; Wittrock, 1963) provides a ready example. The research has not only been confounded by differences in terminology, but also by the frequent use of multiple dependent measures and vagueness as to what is being taught and discovered (Roughead & Scandura, 1968). Similar statements may be made about arguments for and against specific versus general training (e.g., see Scandura, Woodward, & Lee, 1967).

In trying to add precision to their formulations, most investigators to date have taken one of two paths. Some have chosen to elaborate on or to extend the S-R mediational language (e.g., Berlyne, 1965; Staats & Staats, 1963). Others have shamelessly preferred more cognitive, or rule-based, formulations (Bartlett, 1932, 1958; Mandler, 1962, 1965; Miller, Galanter, & Pribram, 1960).

Which approach is to be preferred is perhaps based more on a philosophy of science than on psychology per se. The former approach appeals more to those who want their theories and basic formulations grounded in empirical data. They have a precise language now, which relates spe-

[1] Portions of this article were presented at the meeting of the American Psychological Association, Washington, D. C., September 1967. The author would like to thank John H. Durnin for his general assistance in the preparation of this article.

An unabridged version of the present paper can be obtained on request from the author.

cifically to behavior, and do not want to give it up without good reason. Presumably, they would rather improve it as to detail than to discard the whole idea. Cognitive formulations generally conform more closely to intuition about psychological processes, but they too have major disadvantages. On the one hand, more traditional cognitive theories (e.g., Bartlett, 1958; Flavell, 1963; Tolman, 1949) have been extremely vague as to their relationships to behavior. Precise languages have been almost nonexistent. Modern information processing theories (e.g., Hunt, 1962; Newell, Shaw, & Simon, 1958; Reitman, 1965), on the other hand, which use the computer as a model, have been formulated in precise terms (computer programs). The problem here is that it is not at all clear how specific aspects of programs relate to human behavior—if indeed they do at all. Most of what has gone into such programs exists as much for programming convenience as for modeling human behavior, and it is anyone's guess what are the really important ingredients. In order for a language to be maximally useful, it must be pruned of excess and possibly misleading notational baggage.[3]

Over the past several years, a precise formulation of the notion of a rule has evolved. Since this formulation involves sets and functions, and since these characterizing notions have been used by the author and some of his students in formulating research, the label Set-Function Language (SFL) has been used. The SFL retains many basic tenets of cognitive formulations, but like all scientific languages, is free of specific theoretical assumptions. In addition, the SFL is based on extremely basic, and highly general, notions (sets and functions), so that it deals only with essential aspects of the constructs and empirical phenomena involved.

[3] In this regard, Shaw (1970) has recently presented cogent arguments to the effect that understanding computer programs, which model human behavior, is likely to be just as difficult as understanding the human behavior itself. Computer simulation, in effect, is not an adequate substitute for theory construction in psychology.

The purpose of this paper is to describe this formulation (of a rule) and to show how it provides for a number of features involved in the learning of complex structured knowledge: decoding and encoding processes, ([illegible]) reference, and higher order relationships. Finally, with the addition of an extremely weak theoretical assumption about how subjects (*S*s) perform, a partial solution to the important problem of "what (rule) is learned" is proposed.

THE SET-FUNCTION LANGUAGE (SFL)

Two Preliminary Observations

During the summer of 1962, Greeno and Scandura (1966) found in a verbal concept learning situation that transfer occurred on the first presentation of a new item or not at all. Specifically, they had their *S*s learn common responses (nonsense syllables) to each stimulus exemplar (nouns) of varying concepts. After each S-R pair had been learned, a transfer list was presented containing one new instance of each concept from the first list together with a paired control. The *S*s either gave the correct responses to new concept exemplars on the first learning trial, or they learned the items at the same rate as their controls. The data were consistent with the hypotheses of all-or-none transfer.

It later occurred to Scandura that *S*s might also transfer on an all-or-none basis to new instances of rules in which the stimuli may be paired with different responses. In this case, one new instance of a rule could be used as a test to determine whether the rule is learned, thereby making it possible to predict the responses to other (new) stimuli associated with the rule.

To test this point, a number of pilot studies were conducted during 1963 (Scandura, 1966, 1967a, 1969a); in one experiment (Scandura, 1969a), a total of 15 (highly educated) *S*s overlearned the list shown in Figure 1. Prior to learning the list, both the *S*s and the experimenter agreed on the relevant dimensions and values—size (large-small), color (black-white), and shape (circle-triangle). The *S*s were told to learn the pairs as efficiently as they could, since this

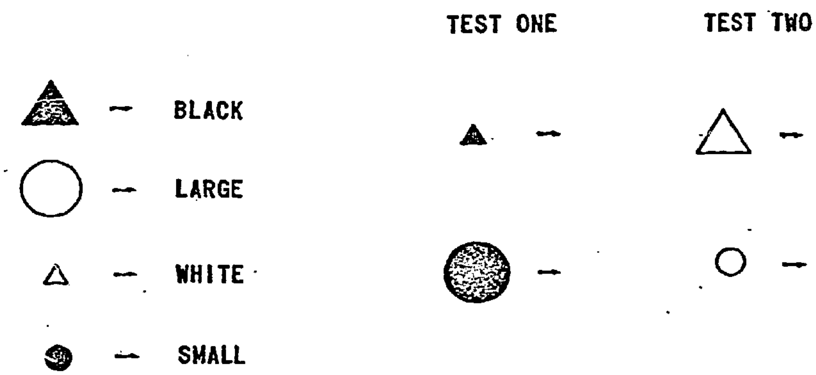


FIG. 1. Sample learning, assessment (Test One), and prediction (Test Two) stimuli and responses.

might make it possible for them to respond appropriately to the transfer stimuli. After learning, the Test 1 stimuli were presented and the *S*s were instructed to respond on the basis of what they had just learned. Positive reinforcement was given no matter what the response. Then, the Test 2 stimuli were presented in the same manner. The results were clear-cut. All but three of these *S*s gave the responses "black" and "large," respectively, to the Test 1 stimuli (see Figure 1) and also responded with "white" and "small" to the Test 2 stimuli.

On what basis could this happen? It was surely not a simple case of stimulus generalization; the responses did not depend solely on common stimulus properties. The first Test 1 stimulus, for example, is as much like the fourth learning stimulus as the first. Perhaps the simplest interpretation of the obtained results is that most of the *S*s discovered the two underlying principles during List 1 learning and later applied them to the test stimuli. These principles might be stated, "If (the stimulus is a) triangle, then (the response is the name of the) color" and "if circle, then size." In effect, whenever an *S* responded to the first test stimulus in accordance with one of these principles, he almost invariably responded in the same way to the second. Since this study was conducted, a relatively large amount of relevant data has been collected with essentially the same results (Roughead & Scandura, 1968; Scandura, 1967b, 1969b; Scandura & Durnin, 1968; Scandura et al., 1967).

The second observation was that each of Gagné's (1965) eight types of learning could be represented by a set of ordered stimulus-response pairs (Scandura, 1966, 1967a, 1968) in which each stimulus was paired with a unique response. That is, each type conformed precisely to the set-theoretic definition of the mathematical notion of a *function*. To see this, first recall Gagné's eight types of learning: (1) signal learning—the establishment of a conditioned response, which is general, diffuse, and emotional, and not under voluntary control, to some signal; (2) S-R learning—making very precise movements, under voluntary control to very specific stimuli; (3) chaining—connecting together in a sequence two (or more) previously learned S-R pairs; (4) verbal association—a subvariety of chaining in which verbal stimuli and responses are involved; (5) multiple discrimination—learning a set of distinct chains which are free of interference; (6) concept learning—learning to respond to stimuli in terms of abstracted properties like color, shape, and number; (7) principle (rule) learning[4]—acquiring the idea involved in such propositions as "If A, then B" where A and B are concepts—that is, a chain or relationship between concepts, internal representations (of concepts) rather than observables being linked; (8) problem solving—combining old principles so as to form new ones.

The first four types clearly involve a

[4] Gagné has not made a distinction between rules and principles.

single stimulus and a single response. (Chaining and verbal associations, of course, may involve intermediary steps.) Multiple discrimination simply refers to a set of discrete S-R pairings (possibly with intermediate steps), each of which may act independently of the others and, hence, must be represented as a separate entity. Knowing a concept, however, may involve any number of different stimuli (exemplars), and each of these stimuli is paired with a common (unique) response. In addition, rules involve multiple responses. The stimuli and responses, however, are not paired in an arbitrary way; each stimulus has a unique response attached to it (see Figure 1, for an example).

In effect a rule can be denoted by a function whose domain is a set of stimuli and whose range is a set of responses. The concept and the association become special cases. A concept can be represented by a function in which each stimulus is paired with a common response, while an association can be viewed as a function whose defining set consists of a single S-R pair.

What Gagné (1965) called problem solving involves a higher level of analysis. In particular, "combining old principles so as to form new ones" requires (higher order) rules which act on other rules. More generally, higher order rules may involve any number of combinations (sets) of old rules and any number of new ones, paired so that there is a unique new rule attached to each set of old ones. (Details are deferred to the section on higher order rules.)

Was this only a more formal way of expressing what psychologists have said all along—that responses are "functionally" dependent on stimuli? I could not help but feel that there was a deeper significance. Still, defining rules, concepts, and associations in terms of their denotative sets left me with the unsatisfactory feeling of not knowing what they really were; or, to put it differently, how to characterize the knowledge underlying the observables.

A Characterization of the Rule Construct

A function can be defined as a set of ordered pairs or as an ordered triple. The denotation of a rule, (i.e., class of S-R behaviors which can be generated by a rule) seems best characterized by the former type of definition, but the rule construct itself conforms more closely to the latter type of definition involving a set of inputs, a set of outputs, and a connecting operation.

Consider, for example, the task of summing arithmetic series (e.g., $1 + 3 + 5 + 7 + 9$). In this case, any one of an equivalence class[5] of overt stimuli (like the sign, "$1 + 3 + 5 + 7 + 9$") may represent the same number series (i.e., $1 + 3 + 5 + 7 + 9$). Each such equivalence class serves as an effective (functionally distinct) stimulus. Effective responses (sums) may similarly be thought of as equivalence classes of overt responses (e.g., "25"). The denotation of the rule, then, consists of the set of ordered pairs whose first elements are equivalence classes of representations of number series, and whose second elements are equivalence classes of representations of their respective sums.

Underlying rules are, however, probably more naturally thought of not as acting on effective stimuli (responses) themselves but on properties of the entities denoted by these effective stimuli. Thus, for example, the property of having "a common difference of two between adjacent terms" refers to the number series, $1 + 3 + 5$, and not to its name, "$1 + 3 + 5$." Note that a distinction is being made between the entity (e.g., number series) and the equivalence class of

[5] By an equivalence class of overt stimuli (responses) or an *effective* stimulus is meant a class of overt stimuli, each of which has the same set of defining properties. The term "effective" is used to emphasize that we are talking about the stimuli and responses "effectively" operating in the situation rather than the overt stimuli and responses themselves. Thus, for example, the stimuli "5" and "five" would, for most purposes, count as the same *effective* stimulus since they both represent the same number. The stimuli "5" and "6," on the other hand, would correspond to different effective stimuli. In previous papers, Scandura (1966, 1967a) used the term "functionally distinct."

The distinction between an entity and the sign used to represent it will also play a role in the present analysis. This distinction is first referred to in the following paragraphs and is explained more fully in the section on reference.

representations of that entity. However, since there is a one-to-one relation between equivalence classes of overt stimuli (the signs) and the abstract entities denoted, we can ignore the distinction, except in the section on reference, where it plays a central role. These properties, in turn, determine (via the rule) other properties (of the responses). One rule for summing arithmetic series, for example, may be represented by the expression, $[(A + L)/2]N$, where A refers to the first term, L to the last term, and N to the number of terms of the arithmetic series in question. The critical inputs associated with this rule are triples of values of the dimensions, A, L, and N (e.g., $A = 1$, $L = 7$, $N = 4$). These triples may be viewed as (composite) properties of the entities denoted by the stimuli. We may refer to these critical properties as response determining (D) properties. The set of outputs consists of response properties (numbers) derived from the properties in D. These properties (numbers) determine equivalence classes of number names (e.g., the number property, 16, which is the sum of the series, $1 + 3 + 5 + 7$, defines the equivalence class of all signs of the form "16"). (Notice, however, that these number properties may also be viewed as properties of the series themselves. In this role, the number properties are called sums, which just happen to be properties of arithmetic series which can be derived from other presumably more easily determined properties, like the first term and the number of terms.)

In effect, a rule may be defined as an ordered triple (D, O, R) where D refers to the determining properties of the stimuli, and O to the combining operation or transformation by which the derived properties (of the responses, R) are derived from the properties in D.

Parenthetically, note that accounting for such behaviors as adding arithmetic series in terms of rules is not the same as introducing mediating responses and response-produced stimuli. In the latter case, the basic idea is to provide a detailed account of the interrelationships involved in terms of (possibly complex) networks of associations. Rules treat such relationships at a more molar level. That is, rules by their very nature act on classes of effective stimuli and not on particular stimuli.

The basic question, of course, is which of these two alternatives better captures the essential characteristics of behavior on structured tasks. The first observation cited above, taken together with the relatively large amount of available data (e.g., Scandura, 1969a), indicates the behavioral reality of rules. Scandura found repeatedly that performance on any one instance of most structured tasks is directly related to performance on any other instance of the respective tasks. Behavior strongly tends to be either uniformly good or bad. (There is more that can be said on this point, but going into this here would detract from the main point.) Accordingly, it would seem that when an investigator is interested in working with structured tasks, the rule would seem to provide the more natural conceptual basis. Mediational accounts of such behavior tend to be ad hoc as well as complex and cumbersome. (In working with nonsense materials, on the other hand, where it is unclear as to what, if any, relationships exist among the instances, some resort to associations and their related theory may be more fruitful.)

This inadequacy of mediational accounts becomes one of principle unless one takes a more general view of stimulus and response than has generally been the case. In particular, no mediation theorist to the author's knowledge has explicitly considered as stimuli what amount, in a related context, to S-R pairs (i.e., associations). (Note: Any given entity may seve as either a stimulus or a response. What the entity is called in any particular situation depends solely on the role it is playing—Hocutt, 1967.) To see this, it is sufficient to consider the associative connections involved in generating sums and differences in arithmetic, together with those connections which relate addition and subtraction. In this case, we would have as a minimum such connections as

$$\begin{array}{c} 4 + 5 \rightarrow 9 \\ \downarrow \\ 9 - 5 \rightarrow 4 \end{array}$$

where the vertical arrow acts neither on the stimuli, $4 + 5$ and $9 - 5$, nor on the responses, 9 and 4, but rather on the associations themselves.

As a second and somewhat more subtle example, consider the task of adding "4" and "3" in column addition. If embedded in a problem like $\begin{array}{r} 41 \\ +32 \\ \hline \end{array}$, the tens digit in the sum is "7." However, if the problem involves carrying, like $\begin{array}{r} 47 \\ +35 \\ \hline \end{array}$, then the tens digit in the sum is "8." In effect, the response given to the complex "4, 3" depends on the context, in particular on the previous response. (In the first problem, the units digits "1" and "2" sum to "3" which does not involve carrying, whereas, in the second problem, the sum "12" of "7" and "5" does.) This implies that the effective stimulus in column addition includes not just the digits in a particular column but the previous response as well, specifically "carry" or "no carry." In effect, the stimulus in this case is a pair consisting of either "carry" or "no carry" paired with the tens digits "4" and "3." Thus, "carry, 4, 3" elicits the response "8," whereas "no carry, 4, 3" elicits "7." To see how these S-R pairs may be viewed as associations on associations, we need only observe that mediation theorists have no difficulty in talking about stimulus properties of responses (or, equivalently, in saying that the source of a given stimulus is the previous response). Hence, in this case, the stimulus properties of the response "carry," for example, may be thought of as eliciting the compound entity "4" and "3" as the response; it is the association "carry"→"4, 3," then, that serves as the stimulus (in the second problem) for the response "8."

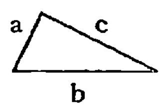
As unfamiliar as this view may seem, this is precisely the sort of assumption that Suppes (1969) had to make in proving that given any finite connected automaton (which for present purposes amounts essentially to a rule), there is a stimulus-response model that asymptotically becomes isomorphic to it. In order to account for rule-governed behavior, then, mediation theorists of necessity will have to generalize what to date has been the traditional view. The section that follows on higher order rules represents an important generalization of this idea. In particular, the view is taken here that "associations on associations" are nothing more than a special case of "rules on rules," such as those commonly involved in problem solving.

Decoding and Encoding Processes

The distinction we have made between overt stimuli and responses, on the one hand, and properties (of the entities denoted by these stimuli), on the other, raises the question of how the decoding and encoding "gaps" are to be filled. In particular, rules operate on properties of stimuli and not directly on overt stimuli (or, more accurately, on properties of the entities these stimuli denote). Similarly, they generate properties (of responses), but not the responses themselves. The rule, N^2, for example, operates on the "number of terms" (a property of number series) and (with certain number series) generates a number (a property of sets) called the sum. The question essentially is one of how to represent the process by which stimulus properties are determined from overt stimuli and how overt responses are determined from derived (response) properties.

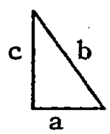
Fortunately, this can be accomplished quite naturally. Each stimulus property defines a class of overt stimuli (i.e., the class consisting of those overt stimuli which denote entities having that property). Hence, decoding may be viewed as a process or mapping which assigns overt stimuli to particular classes. The result of decoding an overt stimulus, then, can be viewed as a class of overt stimuli. For example, one decoding process involved in "perceiving" representations of arithmetic series is the map which assigns given (representations of) series to classes in a way that leaves all of the "essential" properties invariant (including, but not limited to, the first, last,

and number of terms). For example, "$1 + 3 + 5 + 7$" and "one plus three plus five plus seven" would be assigned to a common class, since they both represent precisely the same arithmetic series. Similarly, the stimuli



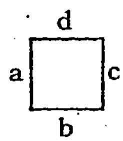
and $(24 + 16) \div 17$,

would almost certainly be viewed by educated adults as equivalent to



and $(24 + 16)/17$,

respectively, but not to



and $(38 + 17)\frac{1}{16}$.

A similar mechanism is required on the response side for encoding. Once the derived response properties have been determined, the question remains as to how the result is to be made observable. Consider a situation in which an S, after having determined the solution to a problem, is expected to write it down on paper. For simplicity, let the solution be the number five (a property of sets) and let the desired response be the numeral "5." Clearly, there are many variations in the way this numeral could be written which would have no effect whatsoever on the referent. Each of the allowed variations in sign refers to the number five. The encoding process simply amounts to constructing or identifying one of these signs. In effect, since each derived property in R defines a class of observables (i.e., overt responses), it would appear that the encoding process might be thought of as "selecting" one of the functionally equivalent overt responses in the defined class.

Normally the processes involved in perception (decoding) and encoding are very complex.[6] It is important to note, however, that the difficulties involved are of a practical nature and are not of principle. In principle, it is always possible to increase the depth of analysis further by introducing additional rules at the beginning of the initially given rules (for decoding) or at the end (for encoding). An initial rule, for example, may be used to derive a property used in a given rule from still more primitive properties. Thus, for example, the property, N, the number of terms in an arithmetic number series, which is used in the rule

$$\left(\frac{A + L}{2}\right)N,$$

may be derived from the more primitive properties, A, L, and D (the common difference) by means of the (initial) rule

$$\left(\frac{L - A}{D}\right) + 1.$$

The notion of a composite rule provides a ready means for representing multistage rules of this sort. Thus, if the rules, $r_1, r_2, \cdots, r_n$ represent n simple rules, such that the outputs of r_i may serve as inputs of r_{i+1} $(i = 1, 2, \cdots, n-1)$, then the rule $g = r_n \cdots r_2 r_1$ represents the composite rule. Complex procedures (e.g., see Groen, 1967; Suppes & Groen, 1967), which involve branching, can be handled in a similar fashion, but discussion here would be an unwarranted digression (for details, see Scandura, in press).

[6] It is worth noting that this complexity is intrinsic and is not unique to the present formulation. Thus, in S-R mediation language, decoding corresponds to $S(\text{overt}) - r_m$ and encoding, to $s_m - R(\text{overt})$. In effect, both formulations make a distinction between overt and effective stimuli, on the one hand, and overt and effective responses (i.e., s_m's which elicit overt responses), on the other. The difference is simply in how the indicated "gaps" are to be filled. Mediation theorists prefer to use associations both for connections between the observable world and internal events and between internal events. In the present formulation, each kind of connection is treated differently. The former involve "inserting observables into classes" or "extracting entities from them." Internal events are connected by rules.

Reference

Although I avoided going into details above, the nature of the discussion forced a recognition of the distinction between equivalence classes of signs, on the one hand, and the entities denoted by these equivalence classes, on the other. This distinction came up both in discussing the rule construct itself and in discussing the decoding process. In the latter regard, we saw that there are two distinct senses in which (meaningful) stimuli may be viewed. (*a*) Signs may be interpreted in terms of what they represent. Thus, signs may be held equivalent if they have the same meaning. This view was emphasized, as it seems most appropriate in dealing with meaningful behavior. (In fact, one might possibly define "meaningful" stimuli to be stimuli which have clear referents.) (*b*) Signs, however, may also be thought of as (meaningless) entities in their own right (with properties of their own). In this case, signs are held equivalent according to whether or not they have certain properties in common. Even signs like "X P Z" and "* o +" which have no well-defined referents, for example, might be taken as equivalent, since each has three perceptually distinct parts.

The problem of reference, then, in the present view, is one of explicating the relationship between signs and their referents. As can readily be appreciated, this general question is extremely complex. All we can do here is to touch on two important aspects of the problem. Specifically, nothing is said about signs with ambiguous meanings.

First, if the meaning of signs is defined in terms of denoted entities, how are we to know when an *S* has acquired particular meanings? There seem to be at least two ways in which this might be done: (*a*) by determining whether or not the subject can paraphrase or otherwise describe the intended meaning, and (*b*) by seeing whether or not he can perform in accordance with the underlying meaning. The referent of (equivalence classes of signs like) "snake," for example, is defined as the class of (all) snakes. An *S* might demonstrate his awareness of the intended meaning, then, by describing what a snake is—"a hideous, long, thin, squirming animal, with no legs, which moves by . . . and whose bite is sometimes poisonous. . . ." He might also do this by reacting appropriately to a statement (sign complex) in which "snake" is embedded. Thus, if someone shouts "Snake!" during a hike in the outback, the listener is likely to evidence through his behavior an awareness of imminent danger. He knows the meaning! The meaning of the relational symbol "run," which refers to the class of all acts of running, might be determined in generally the same way. Apparently, this approach is in some ways similar to Osgood's (1953) S-R formulation, in which responses are viewed essentially as indicators that signs have certain referents. The present view is potentially more precise, however, in that with signs having highly structured meanings, the indicators of meaning can be made highly specific and unambiguous. Consider, for example, the rule statement "$[(A + L)/2]N$." In this case one can test for the meaning (a rule) by presenting particular arithmetic number series and seeing if the *S* can apply the rule so as to give the indicated sum (see below). (For more details, also see Scandura, in press.)

The second question is perhaps more central to the present discussion and deals specifically with the nature of the connection between equivalence classes of signs and their meanings. Specifically, is this connection rulelike—or would associative connections be adequate in all cases? A positive answer to this question would lend considerable additional support for adopting the rule as the basic unit of behavioral analysis. A negative answer would be a serious blow to any such conception.

To provide an answer, first note that the connection between signs and their referents can be represented as rules which map properties of signs into (other) properties. These latter properties, in turn, define classes of entities called referents. Thus, for example, "snake" or any other equivalent sign has certain properties which distinguish it from other signs. These invariant properties are precisely those which are mapped onto the properties which characterize (real) snakes (i.e., the latter properties are what define the class of snakes). The class of symbols equivalent to "run" is assigned to its meaning in precisely the same way.

Of course, we could also represent this type of connection directly in terms of associations. The real question, therefore, is whether or not connections exist which require for their characterization *non*degenerate rules. (Presumably, representation of such rules in terms of associations in the manner described by Suppes (1969) would be cumbersome and, in addition, would require a generalization of the notion of association—to include associations on associations.)

As it turns out, there are two fundamentally different kinds of reference in which nondegenerate rules are involved. One type involves signs that are abstract symbols, and the other, icons.

Before taking a look at symbol reference generally, first consider what might be called elemental symbols, symbols which are minimal indicators of meaning. (In the language of automata theory and formal systems, such symbols are called "letters of the alphabet.") Probably the single most important characteristic of elemental symbols is that they denote arbitrarily. The arbitrary nature of symbol reference has both limitations and advantages. Perhaps its most important limitation is that symbol reference is nongeneralizable. Thus, for example, there is no common way in which the numerals "5" and "6" refer. The meaning of each symbol must be learned separately; knowing that "5" denotes the number of elements in {00000} does not help in learning that "6" denotes the number of elements in {000000}. Any other symbol would be an equally valid candidate.

On the other hand, because symbols may be assigned arbitrary meanings, they can be used to represent highly abstract notions in a precise way. Thus, "five apples" refers to the class of all sets of five apples, whereas "five" refers to the class of all sets of five elements; but there is no loss of precision associated with the increasing degree of abstraction. For example, the symbol, "N" (the set of natural numbers), refers unambiguously to a still higher order collection. Abstract relations may be denoted by symbols with equal ease. Thus, the terms "taller than," "greater than," and "relationship between" refer to progressively more abstract relations with equal precision.

Obviously, not all reference is of this simple form. If it were, *S*s could learn the meaning of, at most, a finite number of different symbols and this clearly runs counter to what is known about language. In particular, there is no upper bound on the number of new statements in English (say) which can be understood by a mature knower of the language. What is needed, therefore, is some mechanism which is sufficiently rich to provide for this sort of capability.

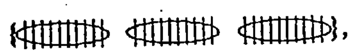
Rules would satisfy this requirement, of course, but it remains to be shown exactly how they might be involved. To make the discussion definite, consider the task of "generating" the meaning of arbitrary numerals like "35," "278," and so on. Clearly, composite numerals of this set have meanings, just as do simple numerals, like "5" and "6." But individuals do not have to learn each meaning independently. They presumably have rules available for figuring out the meanings of even new numerals which they have never seen before.

It is possible to construct a rule for interpreting numerals of arbitrary size, but we can make essentially the same point, and more simply, by considering numerals with no more than two digits. In this case, the following rule will work: "Give meaning to the units-digit (i.e., the first digit on the right); then give meaning to the tens-digit; next, "multiply" the meaning of the tens-digit by 10; finally, combine the meaning of the units-digit with the meaning of the transformed tens-digit." In order to interpret this rule properly, note the following: (*a*) Knowing the meanings of the digits 0 through 9 is basic to using the rule. (*b*) "Multiply by 10" may be interpreted to mean "Replace each element in each set in the denotation of the tens digit with 10 elements of the same kind." For example, consider the numeral, "35." In this case, we first give meaning to

"5," as above. The same is then done for "3." In carrying out the next step, we take into account sets in the second meaning class. Thus, corresponding to the set,

{|||},

we construct the set,



where each of the three bundles contains precisely 10 vertical lines. For details on how such interpretative rules are constructed, the reader is referred to Scandura (in press).

In general, then, it would appear that compound symbols may acquire meaning by referral to the meanings of the constituent symbols, together with a "meaning grammar" by which such meanings are combined to form rules for interpretation. General support for this contention was found in a recent study by Scandura (1967b). It was shown that where the "grammar" necessary for combining the meanings of constituent (minimal) symbols has been mastered, knowing the meaning of particular constituent symbols is both a necessary and also (essentially) a sufficient condition for applying a rule statement involving these particular symbols. In this case, the "grammar" involved the use of parentheses (i.e., "work from the inside out"). The originally naive *S*s were trained with neutral materials [e.g., 3 (5 + 4 (3 + 2))] until they could reliably work with parentheses. Then, half of the *S*s were trained on the meaning of unfamiliar signs, like [X], "the largest integer in X." Training continued until they could reliably give the "meaning" of arbitrary signs of the form [X] (e.g., [6.6], [7.0], [8.9], etc.). These *S*s could almost invariably apply rules, like [([X] + [Y])/[Z]], to instances once statements of these rules had been committed to memory. The *S*s who were not given this training on meaning were uniformly unable to apply the rule. Presumably, the ability to work with parentheses can be viewed as a highly encompassing rule of grammar, one which makes it possible to integrate the meanings of a wide variety of kinds of symbols. Once the meaning of the constituent symbols in a rule statement (involving parentheses) is made clear and is available to the *S* (in memory), the "grammar" combines these meanings into a unified whole. The statement, "name the color," provides a similar example. "Name" is a verb phrase which refers to a large number of acts of naming. "Color" simply indicates what is to be named. Intuitive semantics tells us how these meanings are to be combined. A task for the future will be to make such intuitions public.

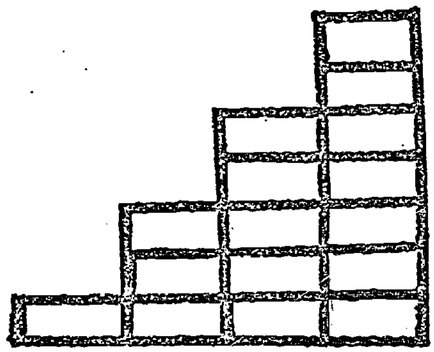
In contrast to symbols, icons [7] have properties in common with the entities they denote; they denote in a nonarbitrary way. This characteristic way in which icons denote has important implications. In the first place, some relations seem easier to denote using icons than others. Thus, proximity and relative size can be handled quite easily, but, as an example, the relationship between parents and children can only be dealt with indirectly. Insofar as mathematics is concerned, icons seem to be particularly well suited to representing geometric ideas where the relationships involved tend to vary continuously.

Second, and this is most important here, icon reference involves (nondegenerate) rules. The icons, "1," "11," "111," "1111," etc., for example, can all be mapped onto their meanings by a common rule. This is possible just because each icon can be put into one-to-one correspondence with the elements of the sets in the corresponding denotative class of sets. (That is, each set in the given denotative class contains the corresponding number of elements.) For a second example, it is sufficient to note that particular properties of relief maps correspond to features of the terrain they represent. These corresponding features provide a sufficient basis for constructing general rules for interpretation.

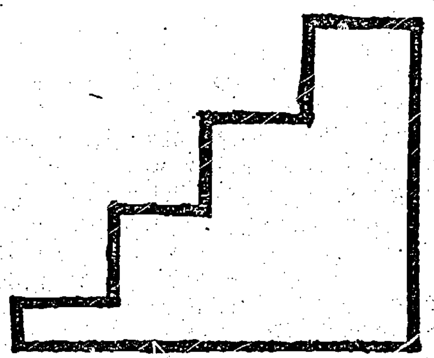
This ability of icons to refer in a generalizable way, however, is bought at a price. Because they are referentlike, icons retain

[7] Here, "icon" is used to refer to any still or moving picturelike representation. While still pictures may refer to "things" and certain kinds of "relations," moving pictures are required to represent action.

progressively more irrelevant information when used to represent increasingly abstract ideas. Thus, it is easy to find an icon that can be used to represent a particular finite arithmetic sequence of numbers in which the successive numbers increase by a common amount. The sequence 1, 3, 5, 7, for example, can be represented by the icon,



However, without the introduction of symbols of one sort or another, icons are not capable of representing arithmetic sequences in general. In this case, the icon would have to indicate that there is a common difference between successive terms and that both the relative size of the first term and the (common) difference between terms and the number of terms are irrelevant. Abstracting from the icon above, we observe that



would provide an adequate representation if it did not specify a relative size between the first jump and the successive jumps as well as a specific number of terms (i.e., 4). This information is irrelevant and, worse, misleading.[8]

Higher Order Rules

It has already been commented that rules can be represented in terms of associative networks, but only if we allow associations to act on other associations (viewed as stimuli) (cf. Suppes, 1969). Since associations in the present view are nothing more than special cases of rules, it seems reasonable to also ask whether there is any natural rule counterpart to associations on associations. In particular, if rules are as basic to complex learning as has been suggested, then one would suspect that there ought to be (nondegenerate) rules which act on classes of associations (rather than on single associations), or, even better, rules which act on classes of rules.

Notice that this observation provides us with another independent check of the power of the formulation. We have just seen how rules are involved in reference, and now we ask whether they are also involved in higher order relationships, which are analogous to associations on associations.

To prove the point, we need only demonstrate the existence of one such higher order rule. As a simple example, consider the rules involved in translating from one unit of measurement into another: yards into feet, gallons into quarts, quarts into pints, weeks into days, and so on. Clearly, there are close relationships among many such rules which obviate the need to learn all of them separately. Knowing how to convert yards into feet and how to convert feet into inches, for example, is often a sufficient basis for converting yards into inches. Furthermore, for most adults, it makes no dif-

[8] It should also be apparent that signs evident in the "real world" are like icons, only more so. Rather than being two dimensional, however, these signs have three dimensions. Because of this, the signs and their referents must have even more things in common. The rules defining reference, therefore, are even more general than with icons. Still, it should be emphasized that "real world" signs need not refer to identity. To the contrary, such signs almost invariably refer to broad classes. Thus, young children let blocks refer to automobiles, buildings, boxes, and so on. Even "John Smith," at a given instant in time, does not refer to identity—but, typically, to John Smith irrespective of when.

ference what the particular units are. If told that there are five "apps" in a "blug" and two "blugs" in a "mugg," it would be a simple task to also convert "muggs" into "apps" (i.e., first multiply by two and then by five).

The point is that many people appear able to combine pairs of given rules into corresponding composite rules. Thus, for example, given rules like, "x yards $\rightarrow 3x$ feet," and "y feet $\rightarrow 12y$ inches," many Ss can combine them to form composite rules, like "x yards $\rightarrow 3x$ feet $\rightarrow 12(3x)$ inches." (Using arrows is a convenient way to represent the denotation of rules. Thus, for example, x yards $\rightarrow 3x$ feet is interpreted to mean $\{(x \text{ yards}, 3x \text{ feet}) | x \text{ is a number}\}$.)

One can account for this type of ability by introducing a higher order rule, which says, in effect, "combine the rules so that the output of the first serves as the input of the second." More specifically, the higher order rule can be characterized by the triple, D = a set of pairs of actions (more accurately, a set of properties which define equivalence classes of pairs of actions), O = the higher order action of combining pairs of lower order actions, and R = the corresponding set of composite actions. The denotation of such a rule, then, can be represented: $\{(R_1, R_2), R | R_1$ and R_2 are (equivalence classes of) rules, and R is the rule formed from R_1 and R_2 by composition$\}$.

Ackler and Scandura are presently performing a study in the University of Pennsylvania laboratory which demonstrates, conclusively in the author's opinion, the behavioral reality of such higher order rules (Scandura, 1970). Given the necessary constituent rules, as above, Ss, ranging in educational level from kindergarten to postgraduate work, were able to solve problems involving the composite rule if and only if they also had available the necessary higher order rule for combining pairs of such rules. Specifically, if they had already mastered the higher order rule, or could be experimentally trained in its use, as judged by their ability to use it on neutral tasks (i.e., neutral rule pairs) to form composite rules, then they were able to solve the composite problems; otherwise, they were not. The amazing thing about these results is that they held up with essentially every S. It was not a question of averaging over individuals or tasks.

Two earlier studies also bear on this issue. The first (Scandura, 1967b) has already been discussed in the section on reference. Suffice it to say here that the rule by which the constituent meaning rules (i.e., rules which assign meanings to minimal symbols) were combined is a higher order rule.

In a second study, Roughead and Scandura (1968) were able to identify a higher order rule, of the sort Gagné and Brown (1961) had alluded to earlier, for discovering other rules. This higher order rule can be stated,

> . . . formulas for the sum of the first n terms of a series (Σ^n) may be written as the product of an expression involving n (i.e., $f(n)$) and n itself. The required expression in n can be obtained by constructing a three-columned table showing: (a) the first few sums, Σ^n, (b) the corresponding values of n, and (c) a column of numbers, $f(n) = \Sigma^n/n$, which when multiplied by n yields the corresponding values of Σ^n. Next, determine the expression $f(n) = \Sigma^n/n$ by comparing the numbers in the columns labeled n and Σ^n/n, and uncovering the (linear) relationship between them. The required formula is simply $\Sigma^n = n \cdot f(n)$ [Roughead & Scandura, 1968, p. 285].

This rule can also be analyzed in the same general way, but the analysis is not as simple as the examples given above. The main ideas are sketched and the reader is referred as before to Scandura (in press) for more details. (a) The inputs of the higher order rule are n-tuples of associations (i.e., degenerate rules) between particular series of a given form and their respective sums (e.g., $1 + 3 + 5 + 7$ is mapped into 16). (b) The output rules are also associations, this time between classes of series (e.g., $1 + 3 + 5 + \cdots + (2n - 1)$) and formulas in n (e.g., n^2) by which sums of particular series of the given form may be determined. In effect, the higher order rule maps n-tuples of specific number series–sum pairs of a given form (e.g., $1 + 3 \rightarrow 4$, $1 + 3 + 5 \rightarrow 9$, $1 + 3 + 5 + 7 \rightarrow 16, \cdots$) into output associations (e.g., $1 + 3 + 5 + \cdots + (2n - 1) \rightarrow n^2$).

As a final example, note that the inverse relation between addition (i.e., the rule) and subtraction is but one instance of a higher order rule by which any binary operation (e.g., multiplication) can be mapped onto its respective inverse (e.g., division).

In each case, higher order rules are in some sense orthogonal to the lower order rules on which they operate. Lower order rules act on classes of stimuli and map them onto classes of responses. Higher order rules map classes of rules (or n-tuples thereof) onto other classes of rules. Of course, there is no reason to stop at this second level, and one can easily envision rules which act on rules which act on rules . . . , and so on.

AN OPERATIONAL DEFINITION OF WHAT (RULE) IS LEARNED

The question of "what is learned" is tied inextricably to the question of transfer (e.g., Smedslund, 1953). In rule interpretations, the tendency has been to explain transfer in terms of "what (rule) is learned." Such interpretations, however, have been rightly criticized as lacking operational definition. On strictly logical grounds it is effectively impossible to define in terms of performance "what (rule) is learned" in any unique sense. There are typically many different routes to the same end. For another thing, rules frequently have an infinite number of instances; it is practically impossible in such cases to test for the acquisition of all but a relatively few.

On the positive side of the ledger, it does not appear necessary to know everything that an S knows in order to predict what he will do in a given situation. Much of the S's knowledge becomes irrelevant once a goal is specified. Even the lowliest rodent has a large number of behavioral capabilities (rules). What rules may be applied depends on what the organism is trying to do. In almost all experimental research (whether it is based on neo-associationistic or more cognitive notions), there is at least the implicit recognition that goals, as well as the stimulus context, are crucial to experimental outcomes. When an S fails to do what is expected of him, he is branded as uncooperative. Specifically, knowing an S's goal in any given stimulus situation is tantamount to specifying a class of rule-governed behaviors, that is, a class of behaviors which can be generated by a rule. (There may be more than one such rule for any given class.) Thus, for example, knowing that an S is trying to add (a given pair of numbers) defines the (rule-governed) class of all pairs consisting of (pairs of) numbers and their sums, denoted $\{[(m, n), (m+n)] \mid m, n \text{ are numbers}\}$. This class effectively partitions the set of rules an S has learned into two mutually exclusive subsets, one including those rules which can be used for adding pairs of numbers and the other including those rules which cannot be so used.

Equally important, an increasing amount of evidence (Levine, 1966; Levine, Leitenberg, & Richter, 1964; Scandura, 1966, 1967a, 1969a) suggests that the relevant knowledge which underlies mathematical and other meaningful behavior can often be specified with a fair degree of precision.

These observations place important restrictions on the form a truly adequate operational definition of "what (rule) is learned" might take. First, it is essentially impossible to define "what *rule* is learned" in any unique sense. Second, an operational definition of what is learned must be formulated relative to a given class of rule-governed behaviors. Third, any such definition must be based on performance on a small, finite number of instances, and, if possible, should be applicable no matter how many test instances are employed.

In view of these restrictions, any attempt to define operationally what particular rule is learned seems a priori doomed to failure. What appears to be needed is a definition which takes into account all feasible underlying rules. Such a definition can be given by specifying what is learned up to a class of rules. Thus, given a class of rule-governed behaviors and that a particular stimulus in that class elicits the corresponding response, "what is learned" can be defined as that class of rules whose denotations all include the given S-R pair. This definition may be interpreted to mean that at least one of the rules in the class has been used in responding to the test item.

The problem remains of adapting the

definition to include any number of test instances. Fortunately, this can be accomplished directly. Given a particular rule-governed class, n test instances, and a performance capability summarized by success on m of the n test instances ($m \leq n$) and failure on $n - m$ of these test instances (and assuming that no learning takes place during testing), then "what (rule) is learned" is defined as that class of rules which provides an adequate account of the test data. In particular, a rule is included in the class if and only if its denotation (i.e., set of S-R instances) includes all of the test instances on which success is obtained, but none of those involving failure. That is, the characterization of "what is learned" includes all of the rules which might possibly account for the fact that S succeeded on some of the items but not others.

The definition says nothing, however, about which rules S may have used to generate his failures. It is also worth noting that if a given rule is in the class "what is learned," and is equivalent in generating power to some finite connected automaton, then there is a way of determining whether or not the S can actually use that particular rule (i.e., whether or not the rule is *really* learned). This can be seen at once by recalling that any such rule can be represented in terms of a finite set of associations. While the total number may be large, it is possible in principle, at least, to test for the acquisition of each and every constituent association.[9]

To see how this definition applies, consider the (rule-governed) class consisting of the arithmetic number series and their respective sums. Let us first suppose that an S has demonstrated his ability to find the sum (2,500) of the arithmetic series $1 + 3 + \cdots + 99$. The definition tells us that the class "what is learned" includes all and only those rules which provide an adequate account of this behavior. In this case, the class would include, among possibly other rules, each of the following: Sequential addition (applied to arithmetic number series); the general rule for summing arithmetic series, denoted $\left(\frac{A + L}{2}\right)N$; the rule N^2, which applies to all arithmetic series of the form $1 + 3 + \cdots + (2N - 1)$; the direct "association" between the series, $1 + 3 + \cdots + 99$, and its sum, 2,500. Thus, "what is learned" might be denoted by the class,

$$\left\{\text{direct association}, N^2, \left(\frac{A + L}{2}\right)N, \text{sequential addition}, \cdots\right\}.$$

As more test information is obtained about an S's performance capability, it will be possible generally to eliminate rules from this class. Suppose, for example, that an S is successful in determining the sum not only of the original test series, but also (say) of the series, $1 + 3 + \cdots + 47$. Then the size of the class "what is learned" is reduced accordingly to

$$\left\{N^2, \left(\frac{A + L}{2}\right)N, \text{sequential addition}, \cdots\right\}.$$

According to the definition, the direct association would no longer be allowed, since it does not apply to the second series. If the S is successful on still another test instance, say, on the series $2 + 4 + \cdots + 100$, then the class "what is learned" is further reduced to the set

$$\left\{\left(\frac{A + L}{2}\right)N, \text{sequential addition}, \cdots\right\};$$

The rule N^2 is eliminated since it is not applicable to the third test series (i.e., $2 + 4 + \cdots + 100$). Suppose, on the other hand, that the S is successful on the first two test stimuli (i.e., $1 + 3 + \cdots + 99$, and $1 + 3 + \cdots + 47$), but not the third (i.e., $2 + 4 + \cdots + 100$). Then, according to

[9] In practice, it is usually not necessary to go to this extreme. The only essential thing is that the rule in question be represented in terms of a (finite) set of operating and decision rules, each of which has a *finite* domain (cf. Scandura, in press). Although this point is implicit in what has been said, it is perhaps not obvious, and I would like to thank Gerald Goldin for raising the question.

the definition, not only would the direct association be eliminated as a feasible rule, but so would the more general rules $\left(\frac{A+L}{2}\right)N$ and sequential addition. In effect, the class "what is learned" would include only N^3, together with possible other unidentified rules which also provide an adequate account of the behavior.

This definition provides a basis for determining the behavior potential (i.e., the class of behaviors that an S is actually capable of) of individual Ss relative to given rule-governed classes. To see this, we first note that the rules in the defined class "What is learned" can frequently be used to generate behaviors in the given rule-governed class, other than the initial test instances. Knowing what rules are learned (i.e., in the defined class), then, might well be used as a basis for making predictions about performance on other instances in the rule-governed class of behaviors. To make such predictions, the only theoretical assumption about performance which seems necessary is that if an S has one or more rules available, which apply in a given test situation, then he will use at least one of them. As trivial as this assumption may seem, it is an assumption. There is no guarantee that just because an S wants to achieve a particular goal and he knows one or more rules which apply, that he will necessarily use one of them. Furthermore, it is an assumption which may well prove to be fundamental to any formal, predictive theory based on the rule construct (cf. Scandura, in press).[10]

The really basic question, of course, is whether or not the actual behavior potential of particular Ss is compatible with this view. Fortunately, Scandura and his associates have collected a fairly substantial body of data over the past few years which suggests that this is the case (Roughead & Scandura, 1968; Scandura, 1966, 1967b, 1969a; Scandura & Durnin, 1968; Scandura et al., 1967). Whenever the response given by an S to one unfamiliar test stimulus was in accord with a particular class of rules, so was the response to a second test stimulus which was of the same "general type" as the first. It was generally possible to predict second test behavior with anywhere between 80% and 95% accuracy. It is encouraging that other investigators have also found this sort of assessment procedure useful. Levine et al. (1964), for example, have used performance on nonreinforced trials to predict performance on reinforced trials with a high degree of success.

Furthermore, the results of the Scandura and Durnin (1968) study suggest that actual behavior potential can often be determined in a systematic manner. It was found that successful performance with two stimuli, which differed along one or more dimensions, implied successful performance with new stimuli which differed only along these dimensions. In particular, success on two instances in a rule-governed class, which differ simultaneously along all possible dimensions, implied success on any other test instance in the rule-governed class.

This whole approach undoubtedly oversimplifies what is an extremely complex problem, but all things considered, it does seem to provide a reasonably adequate first approximation. The ultimate objective, of course, will be to devise a systematic procedure for determining behavior potential on any class of tasks by using a finite testing procedure of some sort. In fact, substantial progress has recently been made in this direction (Scandura, 1970; in press; Scandura & Durnin, 1970).

Summary and Needed Research

A precise formulation of the notion of a rule in terms of sets and functions was proposed. It was argued that this molar formulation cannot be captured by networks of associations unless one allows associations to act on (other) associations. This formu-

[10] I originally felt that a stronger assumption of this sort was needed—in particular, that S will continue using the same rule as long as his goal remains unchanged and feedback otherwise indicates that he is responding in an appropriate manner (Scandura, 1969b). While this Einstellung-type assumption may still have some merit, it is not a necessary requisite for making predictions about behavior potential.

lation was then used as a basis for showing how rules are involved in decoding and encoding, symbol and icon reference, and higher order relationships. Decoding and encoding were shown to involve insertion into and extraction from classes, respectively. Reference was viewed in terms of rules which map equivalence classes of signs into the classes of entities denoted by these signs. Symbols were shown to involve arbitrary reference, whereas icons retain properties in common with the entities they denote. Higher order relationships were then expressed as higher order rules on rules. This was a direct generalization of associations on associations. Finally, a partial solution was posed to the vexing problem of "what (rule) is learned." Given a rule-governed class of behaviors, "what is learned" was defined as the class of rules which provides an accurate account of test data. Empirical evidence was presented for a simple performance hypothesis based on this definition.

There are three major directions in which future research might proceed. First, the rule formulation (SFL) itself undoubtedly can be further improved. While I feel reasonably confident that the basic ideas presented in this paper would hold up under further analysis, additional detail must be added—but only as much as is absolutely necessary to deal with behaviorally relevant aspects of the rule construct. (There should be emphasis on this point to dissuade computer enthusiasts from adopting the language of computer science wholesale (e.g., automata theory) without careful consideration of which aspects are important in human behavior and which are not.) Work in this direction is currently underway and will be reported in Scandura (in press).

Second, the SFL might profitably be used as an analytical tool to help clarify what is involved in many kinds of structured learning and performance. Most of the SFL-based research conducted to date (Roughead & Scandura, 1968; Scandura, 1966, 1967a, 1967b, 1969a; Scandura et al., 1967) has concentrated on an analysis of what is being presented, the nature of the required outputs, what is being learned, and the interrelationships between them.[11] While such analyses can, at least to some extent, be undertaken without the use of the SFL, or for that matter any other scientific language, the SFL seems to provide a useful framework for putting things into perspective and for helping to clarify difficult points. In the author's research a number of questions have been asked on mathematics learning which seem not to have been asked previously in any serious way. For example, Roughead and Scandura (1968) found that what is learned in mathematical discovery can sometimes be identified and presented by exposition with equivalent results. Similarly, Scandura and Durnin (1968) were led, on the basis of an earlier finding (Scandura et al., 1967), to the question of what in the statement of a mathematical rule leads to extrascope transfer.

The SFL needs to be applied more systematically in studies involving subject matters other than mathematics and, in particular, we need to determine where the SFL might profitably be used to formulate research and where not. There is reason to believe that the SFL may be applicable only to the extent that the classes of overt stimuli and responses involved can be viewed as discrete (i.e., nonoverlapping) and ex-

[11] I am of the opinion that insofar as structural learning is concerned, it may be possible, in fact, desirable, to first concentrate on understanding what kinds of behaviors might be involved and to give a distinctly subordinate role to such things as latency and exposure time. Precious little is known about what an S might be able to do when placed in a mathematical situation without complicating the matter further by trying to predict how rapidly he can do it or to determine the precise exposure time needed to bring the behavior about. In effect, what I am proposing is that ecological thinking needs to be brought more directly into theory construction in psychology.

This general type of approach has proved useful in other sciences. In the early development of chemistry, for example, it was of considerable interest to know what kinds of compounds one might expect to get by mixing various combinations of elements. Questions as to the precise values of the boundary conditions of temperature, pressure, and the like needed for such reactions to take place were something which could reasonably be postponed. The first step in theory construction in structural learning might well follow this path (see Scandura, 1970).

ERIC

haustive entities. While these requirements are met throughout much of mathematics and other structured knowledge, this may not be the case in such areas as social studies, poetry, and even language, where synonymy does not necessarily imply equivalence. It is hoped that other investigators will apply the SFL to a wider range of tasks and thereby help to clarify further its relative strengths and weaknesses.

Third, theoretical assumptions need to be made and their implications need to be drawn out. Although this paper was concerned primarily with describing a new scientific language, it was not possible to completely avoid reference to theoretical assumptions. Thus, the proposed operational definition of "what is learned" would be behaviorally meaningless without the application assumption. Fortunately, there is considerable empirical support for the idea. While this assumption is clearly not sufficient for a theory of structural learning, it might nonetheless come to play a central role. Whatever form additional theoretical assumptions might take, it seems almost certain that they would be more compatible with cognitive (rule-based) notions than with those based on neo-associationism. Nonetheless, any complete theory of structural learning will undoubtedly require reference to such things as the limited capacity of human *S*s to process information (Miller, 1956). Without recourse to some such physiological capacity, I can see no way in which to explain memory or other aspects of information processing. (For elaboration, see Scandura, in press.)

REFERENCES

BARTLETT, F. C. *Remembering: A study in experimental and social psychology.* New York: Macmillan, 1932.

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

BERLYNE, D. E. *Structure and direction in thinking.* New York: Wiley, 1965.

FLAVELL, J. H. *The developmental psychology of Jean Piaget.* Princeton, N. J.: Van Nostrand, 1963.

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

GAGNÉ, R. M., & BROWN, L. T. Some factors in the programming of conceptual learning. *Journal of Experimental Psychology,* 1961, 62, 313–321.

GREENO, J. G., & SCANDURA, J. M. All-or-none transfer based on verbally mediated concepts. *Journal of Mathematical Psychology,* 1966, 3, 388–411.

GROEN, G. J. *An investigation of some counting algorithms for simple addition problems.* (Tech. Rep. No. 118) Stanford, Calif.: Institute for Mathematical Studies in the Social Sciences, 1967.

HOCUTT, M. On the alleged circularity of Skinner's concept of stimulus. *Psychological Review,* 1967, 74, 530–532.

HUNT, E. B. *Concept learning: An information processing problem.* New York: Wiley, 1962.

LEVINE, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology,* 1966, 71, 331–338.

LEVINE, M., LEITENBERG, H., & RICHTER, M. The blank trials law: The equivalence of positive reinforcement and non-reinforcement. *Psychological Review,* 1964, 71, 94–103.

MANDLER, G. From association to structure. *Psychological Review,* 1962, 69, 415–427.

MANDLER, G. Subjects do think: A reply to Jung's comments. *Psychological Review,* 1965, 72, 323–326.

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., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston, 1960.

NEWELL, A., SHAW, J. C., & SIMON, H. A. Elements of a theory of human problem solving. *Psychological Review,* 1958, 65, 151–166.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford University Press, 1953.

REITMAN, W. R. *Cognition and thought: An information-processing approach.* New York: Wiley, 1965.

ROUGHEAD, W. G., & SCANDURA, J. M. "What is learned" in mathematical discovery. *Journal of Educational Psychology,* 1968, 59, 283–289.

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

SCANDURA, J. M. The basic unit in meaningful learning—association or principle? *The School Review,* 1967, 75, 329–341. (a)

SCANDURA, J. M. Learning verbal and symbolic statements of mathematical rules. *Journal of Educational Psychology,* 1967, 58, 356–364. (b)

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

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

SCANDURA, J. M. New directions for research and theory on rule learning: III. Analyses and theoretical direction. *Acta Psychologica*, 1969, 29, 205–227. (b)

SCANDURA, J. M. *Deterministic theorizing in structural learning: Three levels of empiricism.* (Rep. No. 55) Philadelphia: University of Pennsylvania, Mathematics Education Research Group, 1970.

SCANDURA, J. M. *Mathematics and structural learning.* Englewood Cliffs: Prentice-Hall, in press.

SCANDURA, J. M., & DURNIN, J. Extra-scope transfer in learning mathematical rules. *Journal of Educational Psychology*, 1968, 54, 350–354.

SCANDURA, J. M., & DURNIN, J. Assessing behavior potential. Paper presented at the Structural Learning Meeting, sponsored by the Mathematics Education Research Group, Philadelphia, April 1970.

SCANDURA, J. M., WOODWARD, E., & LEE, F. Rule generality and consistency in mathematics learning. *American Educational Research Journal*, 1967, 4, 303–320.

SHAW, R. Cognition, simulation, and the problem of complexity. In J. M. Scandura (Ed.), Special Issue of *Journal for Structural Learning*, 1970, in press.

SMEDSLUND, J. The problem of "What is learned?" *Psychological Review*, 1953, 60, 157–158.

STAATS, A. W., & STAATS, C. K. *Complex human behavior: A systematic extension of learning principles.* New York: Holt, Rinehart & Winston, 1963.

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

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

TOLMAN, E. C. Is there more than one kind of learning. *Psychological Review*, 1949, 56, 144–155.

WITTROCK, M. Verbal stimuli in concept formation: Learning by discovery. *Journal of Educational Psychology*, 1963, 54, 183–190.

A THEORY OF MATHEMATICAL KNOWLEDGE: CAN RULES ACCOUNT FOR CREATIVE BEHAVIOR?

JOSEPH M. SCANDURA[1]
Associate Professor of Education
University of Pennsylvania
Philadelphia, Pennsylvania

Mathematics is perhaps the most highly organized body of knowledge known to man. Yet, in spite of its clarity of structure, most of the research done on mathematics learning and behavior has been strictly empirical in nature. To be sure, there has been a fair amount of research in the area and the amount seems to be growing rapidly, but there has been no superstructure, no framework within which to view mathematical knowledge and mathematical behavior in a psychologically meaningful way.

A number of psychologists feel that the mechanisms involved in language, mathematical, and other subject-matter behavior may be accounted for within the confines of S-R mediation theory. This may be possible in principle (e.g., see Millenson, 1967; Suppes, 1969a), but the networks of S-R associations required to do the job would almost certainly be so complex as to provide little intuitive guidance in formulating research on complex mathematical learning. For arguments pro and con, see Arbib (1969), Scandura (1968, 1970b, 1970d), and Suppes (1969b).

As a way around these problems, linguists, like Chomsky (1957, 1965), have introduced rules and other generative mechanisms to account for (idealized) language behavior. Although many details still need to be worked out, most generally agree that some sort of analysis in terms of rules will prove adequate to account for most language behavior.[2]

During the past few years, the author has been attempting to develop a similar approach to mathematics learning (Scandura, 1966, 1967a, 1968, 1969b). No comprehensive scheme for classifying mathematical

1. This paper is based on a talk given at a Symposium on Structural Learning in the Subject Matter Disciplines at the AERA Convention in Los Angeles, Thursday, February 6, 1969. The author would like to thank Leon Henkin, Henry Hiz, Dag Prawitz, Marshall S[illegible], and Patrick Suppes for their helpful comments on an earlier draft and John Durnin for his general assistance in the preparation of this paper.

2. Nonetheless, some linguists (e.g., Hiz, 1967) do not feel that *all* important aspects of language can be dealt with in this way.

behaviors has been proposed, however, and most (but not all) of the experimental research has been based on relatively simple mathematical tasks (cf. Scandura, 1969b). The basic supposition has been that an understanding of what is involved in such tasks will provide a better position for explaining more complex mathematical learning. While there has been increasing support for this contention among behavioral scientists (e.g., Bartlett, 1958; Gagné, 1965; Miller, Galanter, & Pribram, 1960), some mathematics educators have been skeptical. Presumably, the position is that any interpretation of complex mathematical learning in terms of simple rules will surely be inadequate.

In reaction, the author proposes and defends the rather strong thesis that rules *are* the basic building block of *all* mathematical knowledge and that, if looked at in the right way, *all* mathematical behavior is rule-governed. More specifically, it is proposed that the mathematical behavior any given individual is *potentially* capable of, under *ideal* conditions of performance, can be accounted for precisely in terms of a finite set of rules.

This statement is clearly meant to imply more than just a *post hoc* account of a given finite corpus of behaviors. If limited to this, the claim would be trivially true since any given subject during his lifetime is necessarily limited to a finite number of behaviors. (A finite number of behaviors can obviously be generated by a finite number of rules.)

Furthermore, this is not a thesis to be proved since it is basically empirical in nature. The problem is that there is no operational way of determining the behavior potential of a subject independently of the rules used to characterize his knowledge.[3] Unfortunately, it would be extremely difficult and time-consuming to obtain an adequate sample of mathematical behaviors to work with under the ideal conditions envisioned—that is, where the subject is unencumbered by memory or his limited capacity to process information.

To compensate for this difficulty, the author suggests the proposal and evaluation of alternative characterizations of given finite corpora of behavior in terms of their relative powers and/or parsimony. That is, given a large class of behaviors, such as those associated with mastery of a given school curriculum, the idea is not only to come up with a finite set of rules which characterizes the curriculum but to come up with the best possible set. (Loosely speaking, power refers to the diversity of behaviors which the characterization accounts for; parsimony refers to the number

3. If there was some way of knowing, then Church's thesis would provide a natural basis for deciding whether or not the behavior (potential) is rule-governed. Church's thesis (Rogers, 1967, 20–21) is that partial recursive functions (which can be defined formally) are precisely those which can be computed by algorithm (which is an informal notion). Thus, the proposal would be true or false depending on whether the class of potential behaviors is or is not partial recursive.

and intuitive simplicity of the rules in the characterizing set.) Such criteria, of course, have been an essential part of formal linguistics ever since Chomsky's (1957) influential *Syntactic Structures* was published.

In order for a characterization to have maximal relevance to psychology, however, these criteria alone are not sufficient. It is also important that a theory of knowledge (i.e., a characterization) be compatible with the mechanisms which govern human learning and performance. Specifically, it is important, in addition to specifying finite rule sets, to also specify *how* the constituent rules may be *combined* to generate behavior. It is these "rules of combination" which must find parallels in the way learned rules are put to use in particular situations. This question of relationships between different levels of theorizing is an extremely important one. For further discussion, see Scandura (1970c).

The basis of the present argument is that, given suitable rules of combination, much of what normally goes under the rubric of creative behavior can be accounted for in terms of finite rule sets. In order to limit the scope, this paper will deal primarily with those kinds of rules which are more properly associated with mathematical or logical content—specifically, with mathematical systems and axiomatic theories. In each case, one begins with a mathematical characterization and then shows what it means to know the underlying mathematics in a behavioral sense.

Relatively little attention is given to so-called mathematical processes.[4] Thus, for example, inference rules are discussed, but relatively little is said about heuristics and other higher order rules by which inference rules may be combined in constructing proofs. This does not imply, however, that such processing skills cannot be formulated in terms of rules. To the contrary, it is basically a simple matter to formulate such heuristics as "organize (arrange) the data" and "work backward from the unknown" (cf. Polya, 1962) as rules. What *is* hard is to show explicitly how these rules may be combined with other rules to solve problems. Even this problem is not insurmountable, however, and some illustrative analyses of this sort have been worked out (Scandura, 1970b).

What Is a Rule?

Before continuing, it is necessary to define what is meant by a *rule*. In spite of an increasing amount of research on the subject, it is perhaps surprising that the term has no clearly defined meaning among behavioral scientists.

As a first step it is necessary to make a sharp distinction between underlying rules—or generative procedures composed of rules and rule-

4. For a taxonomy of such processes and an introductory discussion, see the author's forthcoming book, *Mathematics: Concrete-Behavioral Foundations* (M:CBF). New York: Harper & Row, 1971, (in press).

governed (RG) behavior. Intuitively speaking, a class of behaviors is said to be RG if the behavior can be generated by a common algorithmic (generative) procedure of some sort. This means, in effect, that a person who has mastered *any* underlying procedure should, ideally speaking, be able to generate each and every response, given any particular stimulus in the class of stimuli.

More specifically, RG behavior involves the ability to give the appropriate response in a class of functionally distinct responses to each stimulus in a class of functionally distinct stimuli. (The term "functionally distinct" refers to the fact that each effective (i.e., functionally distinct) stimulus (response) corresponds to a class of overt and "functionally equivalent" stimuli (responses).) The class of S-R pairs, defined in this way, are called S-R instances. To see what this means, consider simple addition. The proposed definition says that the behavior is RG if each pair of numbers is attached to a unique number called the sum. Thus, for example, any overt representation of the number pair (5, 4) can be paired with any overt representation of the number 9 but not with any representation, say, for the number 6.

Ideally, then, RG behavior corresponds precisely to the notion of a function in the mathematical sense. That is, *every* stimulus is paired with a unique response.[5] When looked at in this way it is clear that what psychologists call *concepts* and *associations* can be viewed as special cases of rules (Scandura, 1968, 1969a, 1969b). Concepts are simply rules in which each stimulus in a class is paired with a common response. Associations are further restricted to a single stimulus-response pair.

In its simplest form, a rule can be viewed as an ordered triple, (D, O, R), where D is the set of (n-tuples of) stimulus properties which determine the responses, and O is the operation or generative procedure

5. As indicated above, of course, the behavior of human beings is not always *ideal*. People make mistakes. There are two conceptually different ways in which errors may occur. First, the rule(s) learned by a subject may only apply to a subclass of S-R instances (of the given RG class). Thus, for example, young children are frequently unable to add numbers which involve "carrying" although they can perform perfectly well on those that do not. In this case, following the notion of partial function in recursion theory, one may refer to such behavior as partial RG behavior. Partial RG behavior is rule-governed but not (necessarily) by rules associated with the given RG class. The other way in which errors may arise is due to the limited capacity of human subjects to process information (Miller, 1956). There is, in effect, an important difference between knowing a rule and being able to use it (Chomsky & Miller, 1963). Thus, a person may know *how* to add any pair of numbers but be quite unable to perform the necessary operations mentally when the numbers are large. In the present discussion, the author assumes throughout that all rules can be used perfectly.

Note (parenthetically) that the abstract notion of a *functor* is sufficiently flexible to capture either or both senses of incompleteness. (Roughly, a functor is a structure preserving function between two categories, the categories being analogous to classes of functionally distinct stimuli and responses.) Whether there is any real significance to this fact or not, however, the author cannot say (cf. Scandura, forthcoming).

ERIC

by which the responses in R are derived from the critical properties in D (Scandura, 1966, 1967a, 1968). More of the detail involved can be represented by adopting ideas taken from recursion theory. In particular, a generative procedure is a sequence consisting of at most four kinds of rules:

1. decoding rules by which essential properties of stimuli are put into store,
2. transforming rules by which things in store are transformed into something else in store,
3. encoding rules by which things are taken out of store and made observable,
4. rules for selecting other rules given the desired goal and the output (which is in store) of some previous rule.

Church (1936) has proposed that any set of behaviors which mathematicians would be willing to classify as partial recursive can be generated by a procedure composed of just these four types of rules. In general, this would include just about all of the mathematical behaviors one normally expects of the school-age child, the ability to perform arithmetic computations, to construct geometric figures with ruler and compass, etc.

Characterization of Mathematical Knowledge

The main purpose of this paper is to indicate how complex mathematical behavior might possibly be accounted for in :ms of finite rule sets.

Mathematical systems. Every mathematical *system* consists of one or more basic sets of elements, together with one or more operations and/or relations and/or distinguished elements of the basic sets. By capitalizing on certain logical equivalences it is possible to reduce the characterizing elements to one basic set and one or more relations. Consider a simple example—the system whose basic set consists of three "undefined" elements A, B, C, denoted $\{A, B, C\}$, with A being distinguished in the sense that it serves as an "identity," and whose defining relation is $\bigcirc = \{(A, A) \to A, (A, B) \to B, (B, A) \to B, (A, C) \to C, (C, A) \to C, (B, B) \to C, (C, C) \to B, (B, C) \to A, (C, B) \to A\}$. This is a system in which the distinguished element A "maps" every element it is paired with into itself. When B is combined with B, the result is C and when C is combined with C, the result is B. Finally, B combined with C in either order results in A. Notice that no meaning is specified for either the elements A, B, C, or the operation. They are "undefined terms."

What may be called an *embodiment* of a mathematical system results on assignment of meaning to the undefined elements. Thus, in the example just cited, the undefined terms might correspond to certain rotations with A corresponding to a rotation of 0°; B, to a rotation of 120°; and C to a rotation of 240°. In this case, the operation would simply be "followed by." That is, the result of combining two rotations is that single rotation which results in the same action as first doing one rotation and then the other. For example, a rotation of 120° followed by one of 240° results in the same action as a rotation of 0°.

These definitions of systems and embodiments say something about the nature of the objects we are studying and in that sense they are extremely important. They do not, however, tell very much about their psychological nature.

What kinds of behavior are implied by knowing systems and embodiments of this sort? And, how can such behaviors be accounted for in terms of rules?

First, knowing a system certainly implies the ability to compute within the system. Thus, for example, given the pair, A, B, the "knower" should be able to give the "sum," B. He should also be able to do more complex computations, like $((A \bigcirc B) \bigcirc A) \bigcirc C \rightarrow (B \bigcirc A) \bigcirc C \rightarrow B \bigcirc C \rightarrow A$, which involve combining individual facts (i.e., associations).[6] In addition, the knower should be able to give "differences," i.e., given the sum and one of the "addends," he should be able to generate the other addend.

If these were the only kinds of behavior to be accounted for one could simply list the facts (rules) involved. But clearly any reasonable interpretation of "knowing a system" must also deal with relationships as well. For example, mastery of a system would surely include the ability to generate the subtraction (difference) rule from the addition rule, and vice versa. Knowing that $B + C = A$, for example, should be a sufficient basis for generating the corresponding subtraction fact, $A - B = C$.

Relational rules of this sort provide a simple way to account for such behaviors. Thus, instead of listing all of the subtraction facts separately it

6. These facts correspond to *non-degenerate* rules in the various embodiments of the system. For example, in the illustrative embodiment, the fact, $B \circ C = A$, corresponds to a rotation of 120° "followed by" one of 240°. The rule (operator) of doing one and then the other applies to *all* pairs of rotations, not just one (pair). In addition, knowing a concrete display corresponding to this embodiment involves being able to perform the various rotations on whatever concrete objects (e.g., an equilateral triangle) might be involved and whatever its position or orientation. As anyone who has worked with young children knows, this is not something which can automatically be assumed. (One thing which can easily be overlocked in analyzing behaviors, for example, is that these "rotations" are actually equivalence classes of rotations, and that these equivalence classes may be different for child and observer.) While such things may not be important in mathematics, strictly speaking, they are relevant in science and, in the opinion of the author, ought to be dealt with as an integral part of the elementary school mathematics curriculum.

In an important sense, then, knowing a concrete embodiment (or a corresponding display) may involve a different type of knowledge than knowing the same amount about a corresponding system. This observation could have relevance to a number of recent results (Dienes & Jeeves, 1965; Scandura & Welis, 1967; Suppes, 1965) and should be taken explicitly into account in designing future studies.

would be sufficient to know the addition facts together with the relational rule. That is, assuming, as is traditional in formal linguistics, that individual rules can be composed—performed in succession.

The obvious way to account for such relationships—the way taken by curriculum developers of the operational objectives persuasion—is to simply add more rules to the characterization. There are, however, major problems with this approach (Scandura, 1970a). For one thing, listing a new rule for each kind of relationship would have a *post hoc* flavor not likely to add much in the way of understanding more creative behavior. For each *new* system (of the same type) considered, for example, there would be a new relational rule for each one in the original system. Even granting the economy obtained by eliminating inverses, and the like, the number of rules could grow large very fast. This would not be bad in itself assuming that this is the best one could do. The important question, however, is: Can one come up with a more efficient account which is at the same time more powerful—and which allows for some measure of creative behavior?

To answer this question, first note that knowing how one or more systems are related to a given one may provide a basis for knowing how to compute in the new systems given how to compute in the original. The relationships of interest will generally be mathematical in nature, but they need not be limited to morphisms. For example, one system may be a simple *generalization* of another, as with cyclic 5 and cyclic 3 groups.

Because of the way particular relationships are defined, however, this advantage will generally be of a limited sort. With homomorphisms, for example, the ability to compute in the new system applies only to the defining operations themselves and not, say, to their inverses or to relationships between the operations. It is worth noting, nonetheless, that knowing even a relatively simple set of interrelated rules such as this would make possible a certain degree of creative behavior—what might be called "analogical reasoning." For example, suppose that a subject has learned how to add in system A and that he knows the homomorphism which connects A to system B (i.e., that he can generate the elements in B which correspond to those in A). Then, the subject should be able to add in system B without ever being told how. Consider the homomorphism to be one-to-one (i.e., an isomorphism), system A to be the embodiment of the illustrative 3 group above, and system B to be the illustrative system itself, then one might generate a sum in the abstract system B by (*a*) using the isomorphism to determine the corresponding elements in A, (*b*) adding in A, and (*c*) using the isomorphism in reverse direction to determine the element in B corresponding to the sum (in A). Notice that this follows only if our rules of combination allow for combination (of rules).

A far more powerful and parsimonious characterization results by simply allowing rules to operate, not on just ordinary stimuli, but on other rules. Such rules may be said to be acting in a higher order capacity—or, in short, to be *higher order rules*. Although functions on functions are common in various branches of analysis, and their formalization is routine, the idea seems not to have pervaded formal linguistics. The closest linguists have come in this regard has been to introduce the notion of a grammatical *transformation* between phrase markers (Chomsky, 1957), which closely parallels what are here called *relational* rules (e.g., between addition and subtraction).

There are two reasons why this has probably not been done in the past. First, grammatical transformations have so far resisted mathematical treatment (Nelson, 1968) insofar as this relates to computer science and, second, no existing approach to psychology (known to the author) provides any real motivation for introducing them.

This is unfortunate since there is a very simple and intuitively sound reason for including higher order rules. The main one is just this: The idea of allowing rules (in rule sets) to operate on other rules is compatible with the following intuitively appealing hypothesis concerning performance. If a subject does not have a rule available for achieving a desired goal, then he typically will try to construct a rule which does work (cf. Scandura, 1970c). There is a good deal of introspective evidence in favor of this hypothesis, and some empirical support for it has been collected. In a recent study (Scandura, 1967b), it was found that the ability to "use parentheses" was a sufficient basis for combining learned rules so as to solve the given tasks which involved interpreting new statements of mathematical rules. Later analysis of these tasks showed that use of parenthesis may be viewed as a higher order rule (Scandura, 1970b). The author is currently involved in research in which success in generalizing this result to a number of different kinds of situations and populations has been achieved (Scandura, 1970c).

Allowing rule sets to act in this way makes it possible for them to "grow" in ways not possible by just forming simple compositions (of rules). Thus, (higher order) rules may generate completely new kinds of rules, and these rules, in turn, may be used to generate still other rules.

Consider what higher order rules might suggest in the present situation. Suppose that a subject has learned a higher order rule which connects each operator (rule) with its inverse. Such a rule would connect not only, say, addition of numbers with subtraction, but composition of all sorts (e.g., of permutations, rotations, rigid motions, etc.) with the corresponding inverse operations. The defining operation of each system and its inverse may be thought of as being distinct rules which are mapped one on to the other by this higher order "inverse" rule. Assume, in addition, that the subject has learned how to add in system A, the relationship (e.g., a homomorphism) between system A and system B, and also how to form the composition of arbitrary rules (in the rule set).

In this case, there are all sorts of behaviors that the (idealized) subject would be capable of. For example, he would be able to subtract, not only in system A but in system B as well. To see this, one need only observe that the subject can form the composition of the rule between systems A and B and the higher order inverse rule. This composite (higher order) rule in turn allows the subject first to generate an addition rule in system B and then to generate a subtraction rule in system B. This subtraction rule, in turn, would allow the subject to subtract. Translated into more meaningful terms, a rule set of this sort would imply such abilities as finding inverses with rigid motions given only the ability to add numbers. But, then, isn't this just what is considered as creative behavior?

Axiomatic theories. There is clearly more to knowing systems than simply knowing the rules and interrelationships within these systems. This amounts to internal knowledge of the systems but it says nothing about the systems in the descriptive sense.

Axiomatic theories are concerned with properties of systems. As an example of one such property, notice that in the illustrative system it does not make any difference in which order two elements are combined. The system satisfies the commutative property; in fact, it satisfies all of the axioms (i.e., properties) of a commutative group of order three.

In order to define precisely what is meant by an axiomatic theory, the next thing to observe is that a set of axioms or properties defines a family of systems, namely that family consisting of all, and only, those systems which have each of the given properties. Therefore, an *axiomatic theory* may be defined to be the set of properties which holds in the family of systems defined by a given set of axioms. The set of axioms, of course, belongs to the set of properties.

Paralleling the discussion of systems, consider the question: "What kinds of behavior are involved in knowing axiomatic theories and what kinds of rules are needed to account for these kinds of behaviors?" Due to the complexities involved, the discussion will be restricted largely to lower order rules.

The *sine qua non* of mastering a theory is to know the axioms and theorems of that theory. In behavioral terms, this ability may be thought of as being able to give on demand the conclusions associated with each set of premises. Thus, as with knowing the particular "addition" facts of the illustrative system, one might be tempted to characterize knowledge of particular theories as sets of discrete associations. This would be wrong, however, on two counts. First, the number of theorems associated with any given theory (including trivial ones) is infinitely large, so that they could not all be learned in this way. (Of course, the number of important theorems is usually much smaller.) Second, and more basic, such a characterization, while feasible in part, would not be very parsimonious or powerful. Many more rules would be needed than might be desired and important relationships would simply be ignored

One problem has to do with *not* knowing proofs of the theorems but there is more to it than just that. Proofs can be learned in a strictly rote fashion and being able to generate one may signify little more than simply knowing the theorem itself.

The kind of rule in mind may act not only in any given theory, or even in any class of theories, but these rules may act in any theory whatsoever—indeed, in any situation at all. They are closely related to *inference* rules of formal logic but they do not act on strings of symbols nor do they generate strings of symbols. Neither do they *all* map properties of systems into properties of systems as one might suspect in view of the relationship between formal systems and axiomatic theories. (Strings of symbols of formal systems correspond to properties of mathematical systems.)

Some inference rules are of an entirely different sort. Instead of operating on properties of systems and generating new properties, what have been called *suppositional* inference rules map logical arguments into properties. Some work has been done in this area under the label "natural deductive systems," e.g., see Kalish & Montague (1964), Prawitz (1965), but little has been done with behavioral questions in mind. In present terminology, the *suppositional* inference rules correspond to rules which map instances of other inference rules, or combinations thereof, into properties. For example, from *any* specific argument, in which property B follows directly from property A, one can infer the property, $A \supset B$. In an important sense, then, suppositional inference rules correspond to what is referred to above as relational rules, and transformations, and not to higher order rules—since they do not operate on other rules, but on instances of other rules.

The stimuli of RG behavior may be viewed as families of systems and the responses as derived properties of these families, called theorems. Thus the RG behavior associated with any particular logical procedure involves a class of families of systems and a class of corresponding theorems of these various families. If the procedures are sufficiently unique, e.g., as in proving many non-trivial theorems, the class of RG behaviors may be quite small, indeed it could include only one instance.

In effect, a logical procedure may act on corresponding properties of *different* families of systems, and produce other properties of the respective families, called theorems. Some idea of the way complex logical

ERIC

procedures operate can be obtained by considering familiar rules of inference. Modus ponens provides a simple illustration. Suppose that the statements "If G is a finite group and S is a subgroup of G, then the order of S divides the order of G" and "G is a finite group and S is a subgroup of G" are properties of one family of systems (actually, of pairs of systems) and "If a function is continuous over a closed interval of the real line, then it is uniformly continuous" and "The function is continuous over a closed interval of the real line" are properties of another family. Then application of the logical rule (of inference) modus ponens tells us that "the order of S divides the order of G" and "the function is uniformly continuous" are also properties of the respective families. The corresponding premises and conclusions are quite different but the (logical) rule of inference by which they are related is identical.[7]

The same general idea may be extended to more complex logical procedures. In this case, decoding rules involve accepting, or rejecting, properties, axioms and theorems, of families as appropriate to given goals the subject might have. Rules of inference correspond to transforming rules (type two rules) and stating theorems, to encoding (type three). Branching rules (type four) may also be involved in logical procedures, as, for example, when repeated applications of a rule of inference is required. For example, the conclusion "D" can be inferred from the premises "$A \supset (B \supset (C \supset D))$," "$A$," "$B$," and "$C$" by repeated application of modus ponens.

Since inference rules and the generative procedures which may be constructed from them apply in all conceivable situations (i.e., to properties of situations), it may be that they might be discovered at an early age from instances—in the same way as many other rules. That is, (learning) *deduction may be viewed as induction on a logical rule*. If this is true, it could have important implications both for the study of mathematical reasoning and for teaching it.

Of course, no one individual has mastered, or ever will, all of the logical procedures that might be constructed. Such knowledge constitutes an ideal which can only be approached. The behavior involved in proving any non-trivial class of theorems is necessarily partial. According to

7. Note, the proposed definition of RG behavior as a function, has been questioned. The comment has been made that "the futility of trying to think of rules of inference (even) as functions is already evident once one considers substitution of equals." However, careful thought should convince one that the input of such an inference rule maps pairs of the form $y = b$, $P(b) = K$ into elements of the form $P(y) = K$ where y is allowed to vary. Thus, the form corresponds to a class of functionally distinct stimuli (e.g., $a_1 = b$, $P(b) = K$; $a_2 = b$, $P(b) = K$; $\cdots$) and so it is not surprising that one can generate any number of different responses (e.g., $P(a_1) = K$; $P(a_2) = K$; $\cdots$).

Church (1936), there exist classes of theorems for which no generative procedure can possibly exist. This does not necessarily mean, however, that theorems which belong to such classes can never be proved. Some procedure might exist for deriving any particular theorem; Church's thesis is simply that no *one* procedure will do for the entire class.

Nonetheless, many logical procedures, even reasonably complex ones, are apt to be common to a number of different theories. The number of more or less unique procedures in any particular theory is likely, according to the present view, to be relatively small. Hence, assuming prior mastery of most "standard" logical procedures, a skilled mathematician may gain mastery of a new theory in relatively short order by concentrating on those procedures associated with some of the deeper theorems of the theory. Note that logical procedures correspond roughly to proof schemas—that is, to classes of proofs of the same general form.

In order to prove most theorems, indeed to successfully engage in complex deductive reasoning of any sort, a subject must know more than just rules of inference, or even a large number of relatively complex logical procedures. The subject must also have higher order rules available by which he can combine known inference rules and other logical procedures into new forms—that is, so that he can *create*. One type of higher order rule that is frequently used in constructing proofs is closely associated with the heuristic: "Work backward from the conclusion." In this case, the learner attempts to derive a procedure for generating the conclusion from the premises, i.e., to construct a proof, by first selecting an inference rule which yields the conclusion and then trying to derive a logical procedure, by using this or other higher order rules, which yields the input of the first rule selected. Presumably, the subject continues in this way until he either succeeds or the whole approach breaks down. The widely used technique of proving theorems indirectly by assuming that the conclusion is false provides a particular example of a higher order rule generated by application of this heuristic (a still higher order rule). In this case, the problem reduces to one of constructing a proof of $\sim A$ from $\sim B$. The final step in constructing such a proof just amounts to selecting what might be called the contrapositive inference rule by which the theorem $A \supset B$, can be inferred from the argument from $\sim B$ to $\sim A$.

More could be said about such things as formal systems and metamathematics but space does not permit. In the first case, it suffices to say that formal systems are easier to work with than axiomatic systems. Nothing new is required, except that the allowable inference rules are specified, and no decoding rules are needed. The axioms and theorems are themselves the stimuli and responses. Metamathematics turns out to be nothing more than an axiomatic type of theory in which only non-controversial rules of inference are allowed.

Concluding Comments

In conclusion, this paper has dealt primarily with what it means to *know* an existing body of mathematics. Relatively little has been said about intellectual skills of the sort that must inevitably be involved in *doing* real mathematics. Nonetheless, it has been shown that what appears to be creative behavior might well be accounted for in terms of growing rule sets. The key idea in making this a feasible and rather attractive possibility is that of the higher order rule. Although space limitations have made it necessary to ignore many details, and there obviously are still a good many important questions left unanswered, the author feels that enough has been said to convince the reader that the basic conjecture must be taken seriously: all mathematical behavior is a rule-governed activity and the basic underlying constructs are rules.

REFERENCES

Arbib, M. A. Memory limitations of stimulus-response models. *Psychological Review*, 1969, 76, 507–510.

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

Chomsky, N. *Syntactic structures*. The Hague: Mouton Press, 1957.

Chomsky, N. Formal properties of grammars. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. II. New York: Wiley, 1963.

Chomsky, N. *Aspects of the theory of syntax*. Cambridge, Mass.: M.I.T. Press, 1965.

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, pp. 271–321.

Church, A. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 1936, 58, 345–363.

Corcoran, J. Discourse grammars and the structure of logical reasoning. In J. M. Scandura (Ed.), *Approaches to structural learning*. Englewood Cliffs, N.J.: Prentice-Hall, in press.

Dienes, Z. P., & Jeeves, M. A. *Thinking in structures*. London: Hutchinson Educational, 1965.

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

Hiz, H. Methodological aspects of the theory of syntax. *Journal of Philosophy*, 1967, 64, 67–74.

Kalish, D., & Montagne, R. *Logic: Techniques of formal reasoning*. New York: Harcourt, Brace & World, 1964.

Millenson, J. R. An isomorphism between stimulus-response notation and information processing flow diagrams. *The Psychological Record*, 1967, 17, 305–319.

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., Galanter, E., & Pribram, K. H. *Plans and the structure of behavior*. New York: Holt, 1960.

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

Polya, G. *Mathematical discovery*. Vol. I. New York: Wiley, 1962.

Prawitz, D. *Natural deduction: A proof theoretical study*. Stockholm: Almquist & Wiksell, 1965.

Rogers, H. J. *Theory of recursive functions and effective computability*. New York: McGraw-Hill, 1967.

Roughead, W. G., & Scandura, J. M. "What is learned" in mathematical discovery. *Journal of Educational Psychology*, 1968, **61**, 581–591.

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. The basic unit in meaningful learning—association or principle? *The School Review*, 1967, **75**, 329–341. (a)

Scandura, J. M. Learning verbal and symbolic statements of mathematical rules. *Journal of Educational Psychology*, 1967, **58**, 356–364. (b)

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

Scandura, J. M. New directions for research and theory on rule learning. II. Empirical research. *Acta Psychologica*, 1969, **29**, 101–133. (a)

Scandura, J. M. New directions for research and theory on rule learning. III. Analyses and theoretical direction. *Acta Psychologica*, 1969, **29**, 205–227. (b)

Scandura, J. M. A research basis for mathematics education. *High School Journal*, 1970, **53**, 264–280. (a)

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. Deterministic theorizing in structural learning: Three levels of empiricism. Report No. 5[illegible], Mathematics Education Research Group, Structural Learning Series, Graduate School of Education, University of Pennsylvania, Philadelphia, 1970. (c) (Also in *Journal of Structural Learning*, 1971, in press.)

Scandura, J. M. Theoretical note: S-R theory of automata? (A reaction to Suppes' reaction) Report No. 58, Mathematics Education Research Group, Structural Learning Series, Graduate School of Education, University of Pennsylvania, Philadelphia, 1970. (d)

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

Scandura, J. M. *Mathematics and structural learning*. Englewood Cliffs, N.J.: Prentice-Hall, in press.

Scandura, J. M., & Durnin, J. Extra-scope transfer in learning mathematical strategies. *Journal of Educational Psychology*, 1968, **54**, 350–354.

Scandura, J. M., & Wells, J. N. Advance organizers in learning abstract mathematics. *American Educational Research Journal*, 1967, **4**, 295–301.

Scandura, J. M., Woodward, E., & Lee, F. Rule generality and consistency in mathematics learning. *American Educational Research Journal*, 1967, **4**, 303–320.

Suppes, P. Psychological foundations of mathematics. Technical Report No. 80, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California, 1965.

Suppes, P. Stimulus-response theory of finite automata. *Journal of Mathematical Psychology*, 1969, **6**, 327–355. (a)

Suppes, P. Stimulus-response theory of automata and tote hierarchies: A reply to Arbib. *Psychological Review*, 1969, **76**, 511–514. (b)

ERIC

Deterministic Theorizing in Structural Learning:

Three Levels of Empiricism

Joseph M. Scandura

MERG and Graduate School of Education

University of Pennsylvania

In spite of the diversity which 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.

ERIC

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 <u>individuals</u> 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 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 which 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 determnistic 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

ERIC

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 psychology, today, is that very little attention has been given to specifying conditions under which theories are <u>not</u> presumed to hold. To date, the sole approach to this problem has been an <u>ad hoc</u> 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 <u>not</u> deal.

This is <u>not</u> true in science, however, where the <u>ultimate</u> aim may be to devise theories which 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 <u>certain</u> of these phenomena to facilitate constructing what might be called a <u>partial</u> 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 which 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 <u>not</u> 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, 483-488). In such cases, it will generally be in the theorist's interest to know just what aspects of reality his present theory does <u>not</u> consider. Stated differently, he must know what boundary conditions must be satisfied in order for

ERIC

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.)

Foundations of a Theory of Knowledge

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 the 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, 27 + 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 has been done by curriculum constructors who have followed this approach. The curricula consist essentially of long lists of rules for achieving the (operational) objectives, one rule for each objective.

ERIC

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, of course, 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 which 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 well.[1] More specifically, allowing rules to 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) $\{x, y \rightarrow x \circ y\} \Rightarrow \{x, y \rightarrow x \circ' y\}$,

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 <u>any</u> 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) $\{x, y \rightarrow x \circ y\} \Rightarrow \{\underline{x}, \underline{y} \rightarrow \underline{x} \circ \underline{y}\}$ (Note: x, y, o are different from $\underline{x}$, $\underline{y}$, $\underline{o}$, respectively)

(5) $\{(x \rightarrow y), (y \rightarrow z)\} \Rightarrow \{x \rightarrow z\}$

Rule (4) may be 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 of (B) 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 well. To see this, we need only observe that application of rule (5) to rules (4) and (2) yields rule

(6) $\{x, y \rightarrow x \circ y\} = \{\underline{x}, \underline{y} \rightarrow \underline{x} \circ' \underline{y}\}$.

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

(7) $\underline{a}, \underline{b} \rightarrow \underline{a} +' \underline{b}$ or $\underline{a}, \underline{b} \rightarrow \underline{a} - \underline{b}$ where $+' = -$.

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

In summary, the essentials of the theory of knowledge as outlined are just these. (1) the knowledge of <u>any given</u> individual at any given stage of learning can be characterized in terms of a <u>finite</u> 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. (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]

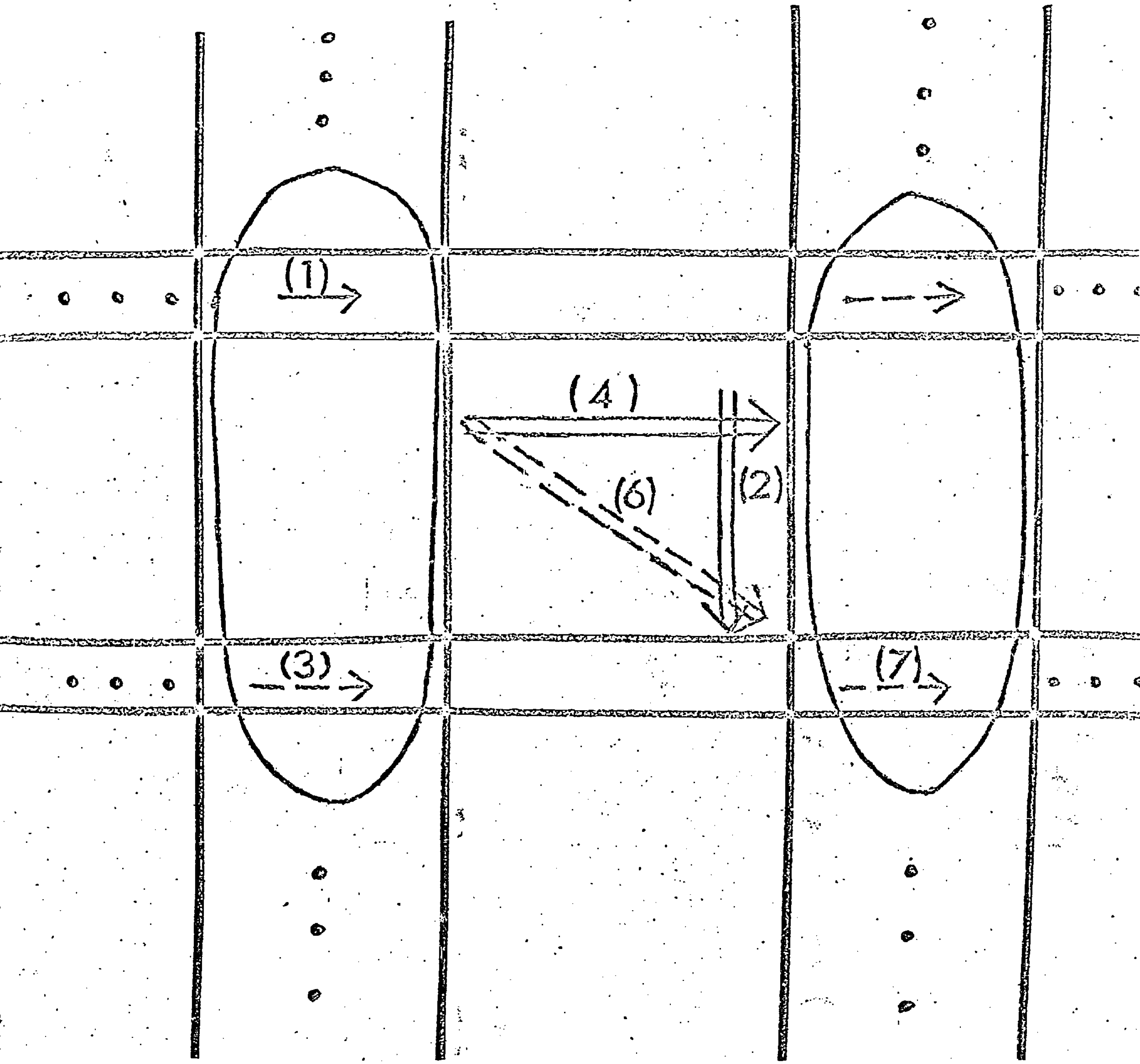


Figure 1: A schematic representation of the basic relationships described in the text. Solid arrows refer to (pre)learned rules and dotted arrows to derivable rules. Rule (5) by which rules (4) and (2) are combined to give (6) is not represented since this would require a third dimension and would complicate 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 which 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 <u>all</u> 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 <u>not</u> 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 which 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 in Figure 2.

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, 1966, 1969). 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 Figure 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

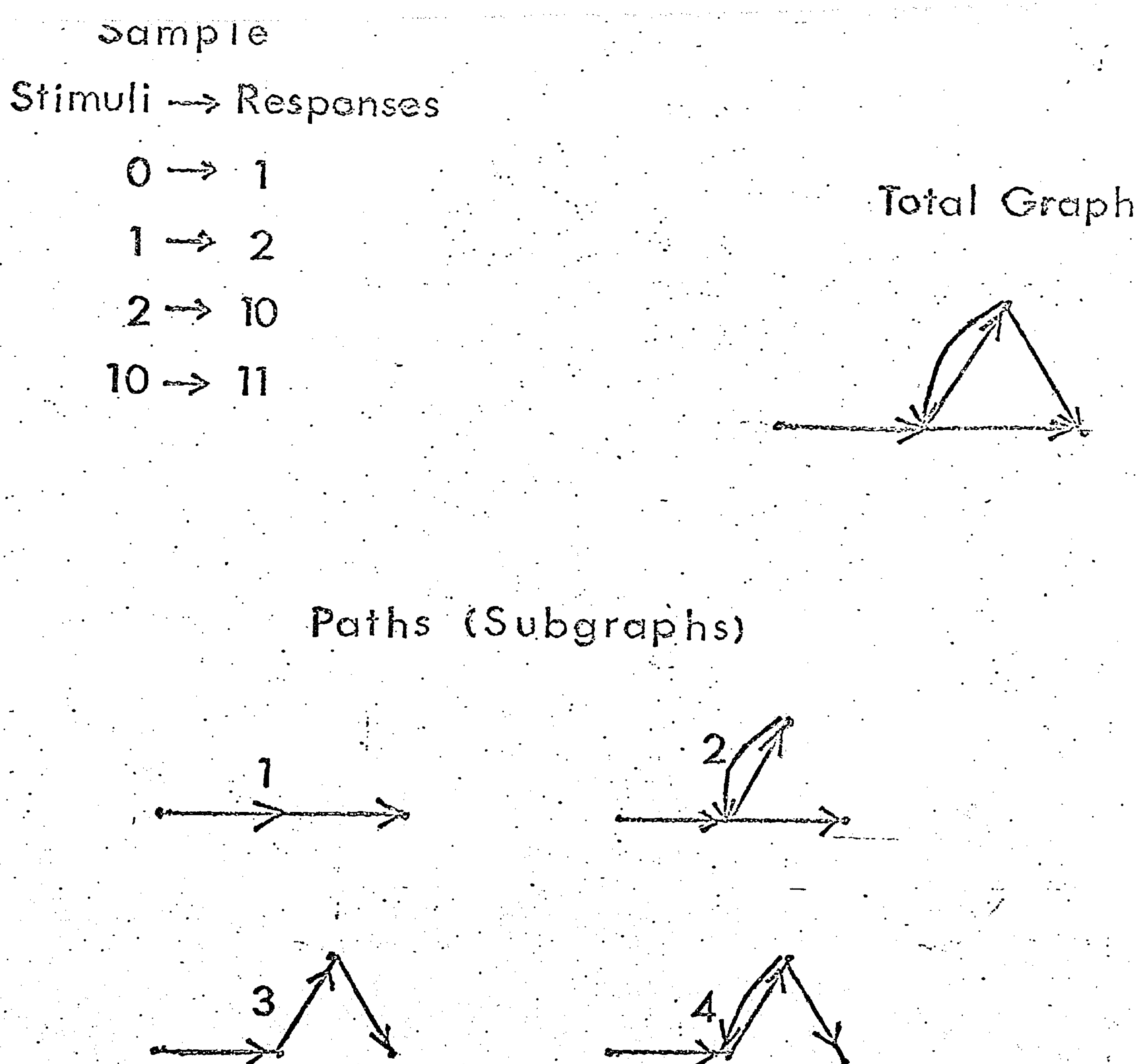


Figure 2: Sample stimuli and responses for the task of generating the next numeral in Base Three arithmetic, together with the (total) graph of a procedure for generating the behavior, and four graphs representing the four behaviorally distinguishable paths through this procedure.

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.

Table 1

Paths	College Student A Test 1	College Student A Test 2	College Student B Test 1	College Student B Test 2	High School Student A Test 1	High School Student A Test 2	High School Student B Test 1	High School Student B Test 2
1	+	+	+	+	+	+	+	+
2	+	+	+	+	+	+	+	+
3	+	+	+	+	-	-	-	-
4	+	+	+	+	+	+	-	-
5	+	+	+	+	+	+	-	-
6	+	+	+	+	-	-	-	-
7	+	+	+	+	-	-	-	-
8	+	+	-	+	-	-	-	-

Note: "+" indicates correct response.
"-" indicates incorrect response.

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.

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 add 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

ERIC

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 which is 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 Figure 3.

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 $r_{11}r_{12}$. 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 which could be used to solve problems such as that requiring $r_{11}r_{12}$ 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 $r_{21}r_{22}$ for its solution. As can be 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.

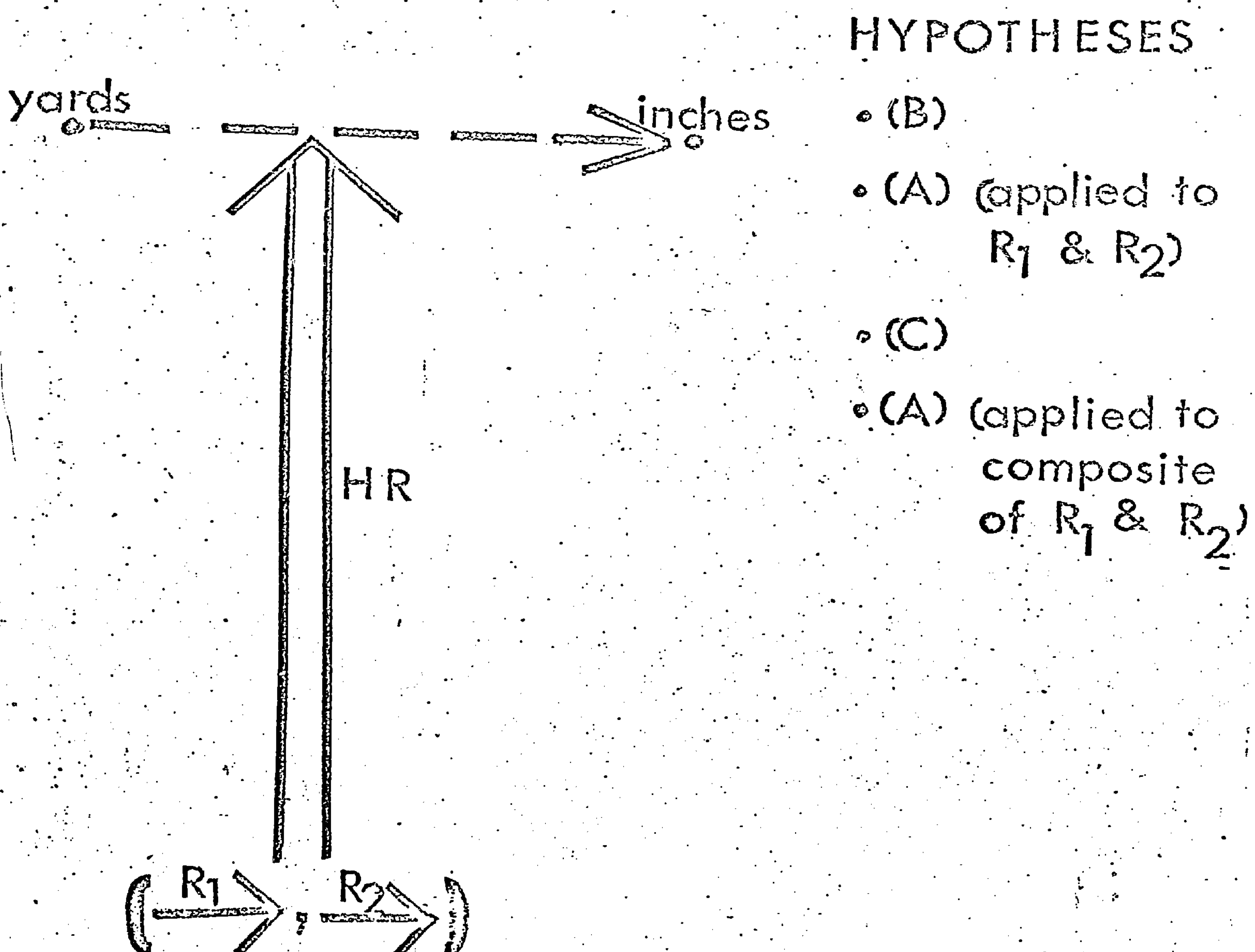


Figure 3: A schematic representation of hypothesized sequences involved in problem solving. R_1 and R_2 represent rules for converting yards into feet and feet into inches, respectively. HR refers to the higher order rule for generating composite rules.

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}	+	+	+	+	+	+	+	+	+	+	+
O $r_{11}r_{12}$	+	-	-	-	-	-	-	-	-	-	-
HR	+	-	+	+	+	+	+	+	+	+	+
r_{21}	+	+	+	+	+	+	+	+	+	+	+
r_{22}	+	+	+	+	+	+	+	+	+	+	+
O $r_{21}r_{22}$	+	-	+	+	+	+	+	+	+	+	+
r_{31}	+	+	+	+	+	+	+	+	+	+	+
r_{32}	+	+	+	+	+	+	+	+	+	+	+
O $r_{31}r_{32}$	+	-	+	+	+	+	+	+	+	+	+
r_{41}	+	+	+	+	+	+	+	+	+	+	+
r_{42}	+	+	+	+	+	+	+	+	+	+	+
O $r_{41}r_{42}$	+	-	+	+	+	+	+	+	+	+	+

Note: "+" indicates S reached criterion.
"-" indicates S did not reach criterion.
"O" indicates that S was tested only (without training).

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. 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 case, 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 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

ERIC

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).

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 which 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 which 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, do not 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 Janie 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 we 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.

The results are summarized in Table 3.

Table 3

Results of Rule Selection Study

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

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 which 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 does 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 and 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 which human beings can process at any given time. Thus, for examsle, 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 new 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 backward 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 which 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 stores (internal) units of knowledge (which serve as stimuli) and generate observables.

What these rules do is to relate new knowledge with knowledge which has been acquired previously. For example, in order to store (i.e., give *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, howeve[illegible] 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 which seems quite promising,[3] however, and about a week or so ago, our data reached about the 80% level of predictability, which may be about as good as can reasonably be expected with this type of task.

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 Figure 4.

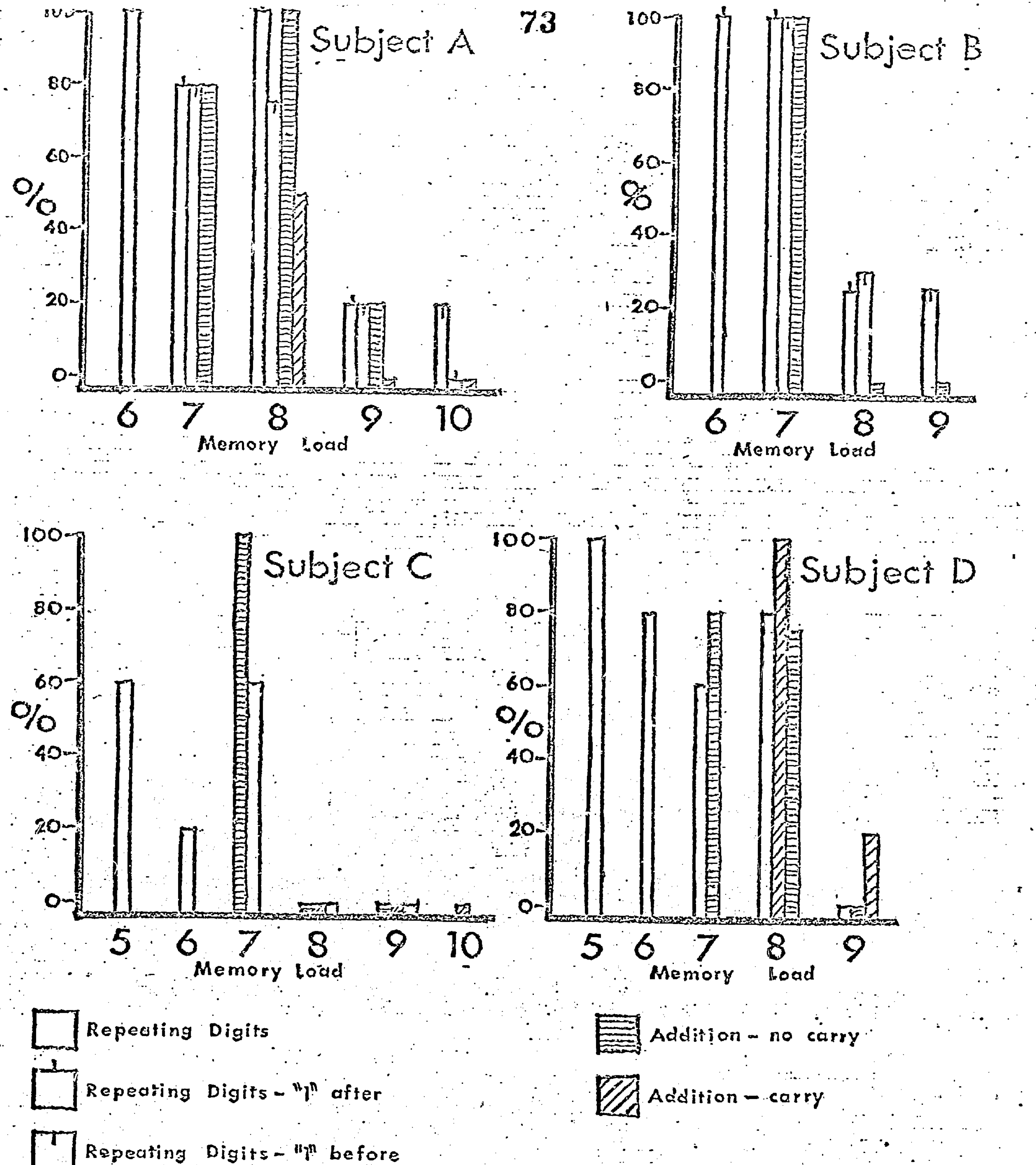


Figure 4: The performance of four subjects on the indicated tasks with percentage of perfect responses plotted against memory load. For comparative purposes, repeating n digits had a calculated memory load of n; repeating n digits and then saying "1" had a memory load of n + 1; repeating n digits after 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 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 just 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 which 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, two of my students (Jeannine Grammick and Debra Whitely) 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 Gagné'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 mind. New York: Harcourt, Brace, and World, 1968.

Gagné, R. M. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagné, R. M. The conditions of learning, New York: Holt, Rinehart, and Winston, 1965.

Gödel, Kurt. über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. Monatschrifte für Mathematik und 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, R. H. P[illegible] and the structure of behavior, New York: Holt, 1960.

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

Peterson, L. R., and Peterson, M. J. Short-term 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 Psychologica, 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, N.J.: Prentice-Hall, forthcoming.

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, (Ed.), Research in mathematics education. Washington: National Council of Teachers of Mathematics, 1967, 35-43.

Wohlwill, 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 as rules on rules and Miller, Galanter, and Pribram's (1960) TOTE hierarchies come close, however.

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]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, Gödel'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.

[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.

A Research Basis for Teacher Education

JOSEPH M. SCANDURA*

University of Pennsylvania

FOR the past fifteen years, most of the research activity in mathematics education has been concentrated in curriculum development. During that time, only a small handful of people have been actively engaged in basic research on mathematics learning.

Today, the situation is changing rapidly. While there are still only a few centers actively engaged in fundamental research on mathematics learning, mathematics educators are turning, more and more, to basic research as a basis for further development. This is, it should be noted, in direct contradistinction to the earlier views of some of those engaged in curriculum development—those who made a special point of downgrading the importance of basic research.

In this paper, I shall: (1) attempt to explain and account for the new interest in research in mathematics education during the last few years, (2) identify some of the kinds of information which every good mathematics teacher needs to know, (3) describe some of the basic research currently underway or being planned at the University of Pennsylvania, and (4) describe some of the teacher education materials we have developed which are based largely on this (basic) research.

Unfortunately, space prohibits discussion of any of these topics in the depth they really deserve. I do, however, hope to convey two main points which recur repeatedly throughout this paper: (1) the tremendous breadth and complexity of the problems which are involved and (2) the great promise of current and future research as a basis for solving these problems.

*I would like to thank Christopher Toy for his general assistance in the preparation of this paper. His participation was made possible by a graduate research training grant to the author by the U.S. Office of Education.

(1) WHY THE INCREASED INTEREST IN BASIC RESEARCH IN MATHEMATICS EDUCATION?

Let us first turn to the question of why the increased interest in basic research in mathematics education. We can answer this question best, I think, by tracing the history of basic and developmental research in this country over the past four decades.

It seems that successful scientific and technological development requires the presence of two vital ingredients: first, an adequate scientific base usually obtained through basic research; second, adequate financial and social support, most frequently by governmental agencies.

As a case in point, consider the space program. Here, a well worked-out scientific base goes as far back as Newton's Theory of Mechanics and includes, as well, Goddard's more recent work (1922) on liquid-fuel rocket technology. Yet, in spite of the ready availability of this ground-breaking work, full-scale development of space technology didn't come about for many years. Such development began only after considerable governmental pressure and concomitant economic pressure were brought to bear. Thus, the U.S. did not move into space seriously until the launching of Sputnik by the Soviet Union was coupled with the supposed "missile gap" of the 1960's. In short, it was public pressure and continued Soviet accomplishments that prompted President Kennedy's pledge for a moon landing before 1970.

And today, almost a decade later, after we have seen the spectacular achievement of Apollo 11, it is questionable whether the nation is willing to commit itself singlemindedly to landing men on Mars in the near future, despite the technological feasibility of such a goal. At the present time, domestic issues have taken priority with a large percentage of the American public and, as a result, much of the money that would normally be earmarked for space projects will undoubtedly be diverted to other areas.

In the late nineteen-fifties, the situation in mathematics education was much the same as it was in space research. Sputnik also gave realization to the American people that mathematics education in this country was woefully inadequate.

Furthermore, fundamental advances in mathematical research in the previous fifty to one-hundred years did, in this instance, provide a more than adequate scientific base for revolutionizing the content of school mathematics. (It is worth noting in this regard that the so-called "new mathematics" was *not* an invention of curriculum developers of the last decade.)

More recently, though, major improvements in mathematics education have come more slowly; it has been much harder to come up with new programs which are really better than what we already have. At best, the most recent programs have simply been refinements of other programs and, at worst, they have been unrealistic and/or philosophically indefensible. Some educators, for example, would have us teach high school students the same material that is at the present time offered in the junior and senior year to students at our better universities. Clearly, if one is dealing with extremely gifted students, this idea is completely feasible and perhaps even desirable. But, it can only be applied to the teaching of our more numerous "top twenty percent" by overemphasizing the relative role of mathematics in the high school curriculum.

One obvious reason for the slowdown is the economic pressure of the Vietnam war. This has resulted in greatly decreased support for research and development and has made it extremely difficult, if not impossible, to maintain the pace of the late fifties and early sixties. Perhaps the most fundamental reason, however, is that we lack an adequate base in the behavioral sciences (and educational philosophy) for making significant further improvements in curriculum development and, more specifically, here in teacher education.

In effect, when adequate basic knowledge is available, the relative gains from developmental activity are likely to be greater than those from basic research. However, after development has progressed for a period of time, the payoff from basic research is likely to be much greater. Figuratively speaking, developmental activity, without basic research, is much like living it up on past savings without any concern for the future. This is the situation in which we find ourselves today in mathematics education. We have largely exhausted our reservoir of knowledge about mathematics teaching and must, in my opinion, begin to build up a new body of knowledge before we can expect further breakthroughs in development of the sort to which we have become accustomed. This is particularly true, I think, in the areas of teaching methodology and the assignment of values to various objectives which might be included in a mathematics curriculum.

The situation is not unlike that in the field of atmoic energy where on the basis of the knowledge provided by Einstein's Theory of Relativity and the work of other pioneering physicists, like Rutherford and Fermi, scientists and engineers were able to produce in the late 1930's the world's first sustained chain reaction. Later, given the added impetus of World War II, scientists were able to create the atom bomb. Later still, on the basis of the same basic know-how, they were able to produce the hydrogen bomb and even to go on to harness the atom for peaceful purposes. However, when it came to harnessing hydrogen power (the fusion process) the situation was quite different. In spite of the billions of dollars that have been spent on development, the field is at a relative standstill and many scientists believe that we will not succeed in taming hydrogen power until we know much more about the processes operating inside the nucleus of the atom. In other words, development, here too, is dependent upon basic research.

(2) IDENTIFYING THE INFORMATION A GOOD MATHEMATICS TEACHER NEEDS TO KNOW

We now turn to some of the basic questions which must inevitably be asked (and answered through research) if teacher education in mathematics is to progress from its present sorry state. To conserve space, I shall not attempt to deal with the important subject of practical experience.

I will focus, instead, on the intellectual aspects of teacher education. This does not mean that I am suggesting (or feel) that practice teaching, micro-teaching, or classroom observation, for example, can be dispensed with, but rather that, while work in these areas is being actively pursued, there is very little work going on which deals with the conceptual aspects of teacher training. So, I will put my emphasis there.

The basic question we want to ask here is, what is it that the teacher needs to know? Now, it is obvious that the teacher needs to know something about mathematics—but, what mathematics? What should the level be? The emphasis? Should we teach arithmetic or number theory? Geometry or topology? Questions like these, about which we all have intuitive feelings, have, I fear, hardly begun to be dealt with in a systematic way.

Clearly, the teacher also needs to know something about teaching methodology. In this case, we are even worse off, however, because it is not merely a question of what methods to teach but whether there is really anything now known about methods which is worth teaching.

In the recent past, we have tended to emphasize such things as the history of modern curriculum development in mathematics or tended to make vacuous statements about how to motivate children or how to use audio-visual aids. About the best we have been able to do is to talk about particular approaches to the teaching of specific topics such as the multiplication of signed numbers to specific kinds of students. The alternative has been to espouse misleading doctrines concerning discovery teaching, inquiry training and the like.

While I do not quarrel with most of these ideas themselves, I do question whether this sort of approach really helps teachers to understand what mathematics education is all about. I have observed far too many classroom teachers who talk about such things as commutative and associative laws without having the foggiest notion of what all of this might mean in the broader context of mathematics education. Many of our more recent graduates—even our so-called better trained elementary school teachers, are merely replacing one set of terminology with another. And, this is certainly not mathematics teaching at its best. I feel that we must do better if we are ever going to provide our children with the kind of education we would all like them to have.

To do this, the first thing we must do, in my opinion, is to avoid the archaic and artificial dichotomy which separates instruction in mathematics, per se, from the methods one might use to teach mathematics. (This dichotomy originally developed because of poor communications between departments of mathematics and departments of education at our universities and colleges. Too frequently neither department has been willing or qualified to do the work of the other. However, I am happy to say that as more and more well-trained

ERIC

mathematics educators enter the field, this communication gap is slowly closing.)

It is of as little value, to give elementary school teachers empty statements about nonexistent theories of instruction as it is to teach them high level mathematics which they cannot fathom. And it is still worse to separate the mathematics that teachers learn from the methods they will use to teach others. For one thing, teachers who learn mathematics in isolation from methodology frequently have considerable difficulty translating the material into a form which can be taught to children.

Furthermore, there is simply not much that one can say, at present, about methods which do not include specific content. Even if general principles are found—and I think that some will be—the teacher must still learn how to apply those principles to specific mathematical topics.

What we need to do is to conceive of the teacher's job in more operational terms than we have in the past. We should, in fact, reformulate our question of what the teacher should know to read, "What are the capabilities which the teacher needs in order to teach mathematics effectively?"

In this context, consider the subject matter. In my opinion, the teacher must know, in fairly explicit terms, what kinds of mathematical behavior to expect of her students. She needs to know, too, something about the competencies required to elicit these behaviors. This will involve such things as the ability to compute (in arithmetic), the ability to make simple discoveries and, even the ability to prove a simple theorem. At the present time, there is relatively little that is known about these things.

For instance, the question of how best to characterize the knowledge which underlies some given universe of behaviors associated with a particular subject matter like mathematics has hardly been asked. It is true that some promising beginnings have been made in the field of formal linguistics, but, again, almost nothing has been done in mathematics or in any of the other school subjects.

In addition to knowing something about the general nature of mathematical knowledge and mathematical behavior, I feel that the skilled elementary school teacher also needs to know the objectives of elementary school mathematics and to know how they relate to the rest of the curriculum. For another thing, the teacher should have some rationale or general set of criteria for selecting and modifying the objectives in any given curriculum. The goals for a good mathematical education in affluent suburbs may be very different from the objectives in the inner city slums of our metropolitan centers or in farm communities across the country. In short, what might be a sound curriculum in Palo Alto, California, or Scarsdale, New York could be untenable in North Philadelphia or Zap, North Dakota. The basic question is how and why these objectives should change. We need a general philosophical framework within which to make such changes and modifications in curriculum. Without such a framework, we can only state opinions and are unable to subject those opinions to the scrutiny of others.

Teachers also need to know more than a grab-bag of teaching techniques and interest-getting devices. They need to acquire abilities which pertain to the teaching and the learning of mathematics generally. And, they need practice in the application of these principles to a wide variety of mathematical topics.

For example, teachers should know something about how to identify operationally the objectives in text materials and exercises. Many textbooks simply do not coordinate their content with the exercises they offer. While they may talk a good deal about teaching students *how* to discover, for example, their exercises often amount to little more than simple applications of *what* is discovered.

Another ability which all teachers should have is the ability to identify the prerequisites for learning any particular task. The teacher should be able to determine in logical, systematic steps what it is that the student needs to know before he can be expected to learn or perform the desired task. Figuratively speaking, the teacher ought to be able to answer the question of "why Johnny can't multiply". What we want, then, is a superdiagnostician who can identify the source difficulties under any and all circumstances.

In addition, teachers should have the ability to assess (determine) the knowledge that a given student has. Unless the teacher is able to construct test items which will get precisely at what particular children know, she can never determine for sure whether her students have the necessary requisites for achieving what she wants them to achieve. Indeed, she will be unable to determine whether, in fact, they have learned what she intended to teach them.

The final ability we shall consider is the ability to motivate, the basic requisite for all learning. To date, the best we have been able to do in this area is to give the teacher specific techniques for teaching specific topics or, more specifically for motivating children to learn these topics. Sometimes we give the teacher general advice like, put the mathematics in the context of a game or, make up problems which deal with things in which the child has an interest. Although advice like this sometimes succeeds, it fails as often.

There is a real need, I think, to talk more about basic principles and, even more important, to provide the teacher with a general frame of reference for thinking about motivation. Hopefully, the teacher will then be in a better position to make up her own mind and to evaluate her own techniques for motivating children to learn in the variety of situations which confront her everyday.

On too many occasions, I have heard professionals (like myself) throw up their hands in dismay and say something like, what we need in mathematics education is to develop "teacher proof" texts. This is nonsense, of course. Good teachers can save even the worst texts and poor teachers frequently render the best texts incomprehensible. Since we are a long way off from replacing the teacher—if, indeed, we ever should—we must train teachers to use textbooks effectively so that the textbooks do not use them.

(3) NEW DIRECTIONS FOR BASIC RESEARCH IN MATHEMATICS EDUCATION

Obviously, we do not have all of the necessary information. There is much we need to know. I am happy to say, however, that progress is being made and that the future looks quite promising indeed.

Rather than attempt to review all of the research which bears directly or indirectly on mathematics education, I will limit myself in this paper to the research in which I have taken part. I have discussed elsewhere my completed research on the learning of mathematical rules, the ability treatment-interaction question and problem solving so there is no need to go into that here.* Instead, what I would like to do is to describe some of the things that we are now doing and plan to do as part of our new research project in mathematics and structural learning.** While I can do no more here than simply scratch the surface, I would like to call particular attention to the new directions we plan to follow.

One of the major concerns of the project is the problem of how to characterize mathematical knowledge. By that I mean that we want to know how to account for the behaviors which knowing mathematics makes possible. And, we want to account for these behaviors in a way which is both (1) mathematically viable—that is, compatible with the way mathematicians think about their subject—and (2) behavioral in nature.

While very little research has been done on this problem directly, there is much to be learned, I think, in this regard by considering some of the limitations of related research. Current work on operationalizing objectives, for example, obviously deals with behaviors, but the research has a distinctly *post hoc* flavor and does not provide a viable characterization of contemporary mathematics. This is true not only because of the almost exclusive emphasis on what might be called "rote rules"—but, also because very little attention has been given to such questions as how to characterize the higher order intellectual skills involved in actually doing mathematics or even how to characterize knowledge of the obvious relationships which we know exist between different mathematical rules. I shall say more on this below. Furthermore, no distinction has been made between the behaviors a subject might elicit, and the knowledge which might account for that behavior. This is an extremely important distinction in planning a curriculum because there are many different ways in which a person might learn to elicit any given class of behaviors. The specific form of the curriculum itself will depend, for example, on whether we want to teach children to subtract, say, by borrowing

* For a summary of much of this research see my recent articles in Acta Psychologica: New Directions for Theory and Research on Role Learning. I. A Set-Function Language. 1968, *28*, 301–321. II. Empirical Research 1969, *29*, 101–133. III Analyses and Theoretical Direction 1969, *29*, 205–227.

**This research is being supported through a grant from the U.S. Office of Education to a committee of the National Academy of Sciences for basic research in education.

or by the method of equal additions. Similarly, we can teach the so-called three cases of percentage separately or as simple variants on a common theme.

A second line of research related to the problem of characterizing mathematical knowledge, is associated with formal linguistics. The methods that linguists use to analyze language, that is, to identify the competencies underlying language behavior, are very similar to the kind of methodologies we need to use to identify the competences underlying mathematical behavior. The specific kinds of competencies which are most critically involved in language behavior, however, are simply not the same as those associated with mathematics. For example, although linguists such as Chomsky have argued the need for higher order competencies, which deal with relationships like that between passive and interrogative forms of a sentence, there is by no means uniform agreement on the point. One could hardly find a mathematician on the other hand, who did not feel that such higher order relationships form the very essence of mathematics—not to even mention what appear to be important differences in the nature of these higher order competencies.

To make matters worse, linguists have expended most of their research efforts at the syntactic level, that is, with grammatical relationships between, for example, words and sentences treated as entities themselves, and have almost disregarded the level of semantics or meaning. In ordinary mathematics, one can hardly get started without having to deal with both levels.

Thus, while it is true that we can gain many valuable insights into our problem from the methods of research used in formal linguistics, it is equally clear that we must be ready to extend and even depart from that tradition where necessary.

There is also much to be learned from work in the foundations of mathematics but, here, we must be even more careful. Specifically, in the foundations of mathematics the basic problem is one of clarifying the nature of the relationships between different kinds of mathematical objects. The *sine qua non* of research in the area is a formal and precise characterization of the mathematical objects in question. Now, this characterization, as it is usually expressed, is designed primarily to make (doing) the (meta-) mathematics easier

ERIC

and has little to tell us about the way in which one might characterize mathematical knowledge.

Although I have been working in this area for a number of years, it is only recently that I have begun to feel that we are finally beginning to get at the heart of the problem. Stated simply it is how can we account for complex mathematical behaviors; more directly to the point, how can we evaluate alternative characterizations of the same behaviors? The truth of the matter is that there may be any number of ways of accounting for the same behaviors. The important thing is that certain accounts are undoubtedly better than others, and our goal, obviously, is to come up with the best one possible. For example, one might conceivably characterize a given mathematical curriculum by listing all of the separate competencies which can be identified. One competency, for example, might make it possible to add and, another, to find the areas of certain geometric figures. This is basically the approach taken by those who would reduce curricula to sets of *discrete* operational objectives. There may be other ways of accounting for the same behaviors, however, which are in some sense both more powerful and/or more parsimonious. By "more powerful", I mean that the set of competencies makes it possible to generate more of the desired behaviors than the original (set). Even where two different sets of competencies account for all of the initial behaviors, one of them may be more powerful in the sense that it accounts, in addition, for behaviors outside the initial class. And by "parsimonious", I mean that the characterizing list of competencies is in some way shorter or more highly organized and interrelated.

Suppose, for example, our original aim was to account for the ability to add and the ability to subtract. In this case, one might simply identify one competence which accounts for addition and another for subtraction. Another way of dealing with the same behaviors, however, might be to introduce the same competency for addition and a higher order competency which deals with the relationship between any binary operation of which addition is just one example, and its inverse. Here, we would get subtraction "free", so to speak, since it can be generated by applying the higher order competency to addition. More important, given any other binary operation whatever, such as performing one permutation on a set of objects followed by another, one can automatically generate (account for) the inverse capability as well. In effect, while each characterization contains two competencies, the latter is far more powerful since it allows one to generate behaviors outside the original class in a way which the former never could.

Although we have hardly begun this line of research, it is already clear that current approaches to curriculum development, based on operationalized objectives, are entirely too fragmented. Most such curricula consist of nothing more than long lists of discrete objectives. While we must be careful not to underestimate the difficulty of the task, I would contend that curriculum development and teacher education, to name just two areas, cannot help but gain from systematic research along the lines just described.

While on the subject, I might also discuss some of our work on identifying mathematical processes. Let me first say what I mean by a mathematical process. A mathematical process is a (usually very general) intellectual skill (or competency) which while essential in *doing* mathematics is not normally considered to be part of mathematics (content). The fact that such skills act behaviorally in very much the same way as other competencies was shown in a recent study by Roughead and Scandura.* In that study, it was shown that the skills subjects learn by discovery can sometimes be specified in details and presented directly by exposition with equivalent results. I would also like to say something about a scheme we have developed for classifying such processing skills but space does not allow and I must refer you to my forthcoming book which deals with the problem: *Mathematics: Concrete Behavioral Foundations for Teachers.* Suffice it to say here that it appears that all such skills may be classified in one of four bidirectional and mutually exclusive categories.

The first is the ability to detect regularities and the inverse ability of particularization. The second is the subject's ability to describe what he knows and its inverse of interpretation. The third is deductive reasoning. This category involves drawing inferences, on the one hand, and, on the other, axiomatization—the ability to identify key ideas (in some class of ideas) from which all others may be deduced.

*"What is learned" in Mathematical Discovery. *Journal of Educational Psychology*, 1968, Vol. 59, No. 4, 283–289.

Finally, we consider storage and retrieval processes which are associated with memory. We are just beginning a major project in this area aimed at identifying and determining the utility of specific processing skills and I hope to have more to say on the subject in the next couple of years.

Another area in which research is badly needed concerns the development and application of criteria for selecting from among the large number of available topics in mathematics those which are most appropriate for any given group of students. This is a very difficult problem area in which to work because, in addition to raising mathematical questions about internal consistency, and behavioural questions concerning feasibility, it raises some very basic questions of priority and value. We are just beginning to explore this problem, but one thing has already become clear. There are many paradoxes that we are going to have to deal with. For example, we are going to have to resolve such dilemmas as this: Although it would obviously be desirable to both maximize the learnability of the material we present and, at the same time, maximize the generality of the same material, these two goals are not compatible. They clash head-on with some of our own research findings* which tell us that in order to maximize learnability one must increase specificity, and vice versa. Therefore, any decisions we make along these lines will necessarily be value decisions. Though this is still largely virgin territory, it is an area of great importance and we are hoping to interest "enlightened" educational philosophers in our work.

All things considered, we are probably most deeply engaged in work in still a third major problem area. Simply, the problem is to determine how mathematical knowledge is put to use and, how such knowledge is learned in the first place. Here, again, I break sharply with tradition. I do not feel that applied educational research, for example, while of proven value in dealing with certain kinds of problems, is going to tell us very much about how mathematics is learned, no matter how well designed the studies may be. As a basis for this statement, I can refer to my dissertation: *The Teaching-Learning Process: An Exploratory Investigation of Exposition and*

*Scandura, J. M., Woodward, E., and Lee, F. Rule Generality and Consistency in Mathematics Learning. *American Educational Research Journal*, Vol. 4, May 1967, No. 3.

Discovery Modes of Problem Solving Instruction, *Syracuse University*, 1962. There, it was shown that very minor within-method differences can have a greater effect on experimental results than differences in the basic methods themselves. Even more important, the basic variables that seemed to be involved are not subject to manipulation in the usual way. Different research methods are called for.

I reject, as well, current theorizing in so-called "academic" psychology. For one thing, the type of theory that people have been concerned with in this area is essentially probabilistic—that is, rather than attempting to make specific predictions about the way individuals will act in given situations, such theories talk, instead, about the probability of occurrence of various events. While this is worthwhile information to have, it is not the sort of thing we need most to know, say, in trying to teach students how to solve quadratic equations.

The problem, as I see it, is that theories of this sort treat everyone on the same plane. First, assumptions are made about how individuals do things. Then after introducing random variables into the picture, predictions are made concerning group statistics, such as means and variances. Next, experiments are run with groups and the results are compared with the predictions. So far, so good. It is the last step that causes the difficulty. As a result of the comparison, inferences are typically drawn about how individual subjects actually do things.*

While this may be a reasonable sort of assumption to make with "naive" rodents or with human subjects, when dealing with simple reflex behaviors, it simply does not apply to the learning of more complex material like mathematics. What humans do on any nontrivial situation depends in a very direct sense on what they already know and can do. Any attempt at theorizing which ignores or otherwise bypasses individual differences, cannot, in my opinion, hope to provide a viable model for education.

My own present approach, goes in quite a different direction. Rather than treating individual differences as unwanted or uncontrollable experimental variables, I have, during the past year and a half,

*This basic fallacy is widely recognised in psychology, but for one reason or another, few psychologists seem sufficiently concerned to try to do anything about it. Apparently, it has been the sort of thing which is much easier to sweep under the rug and forget.

been heavily engaged in deterministic theorizing about individual processes. Obviously, this is not the place to go into detail about such matters but let me at least outline some of my concerns. One problem area in which one of my doctoral students, John Durnin, and I have been deeply engaged concerns assessing the behavior potential of individual subjects. More particularly, we are attempting to determine which items in a given class a subject will be able to perform successfully on and which he will not, on the basis of his (prior) performance on a small finite number of test items. At a strictly theoretical level, we have been able to prove theorems about the number of test items needed, with respect to any given class, and, even more important, how to go about selecting these test items. At the time of this writing we have just begun empirical work on the problem but the results, so far, seem extremely promising.

Another line of research is concerned with the problem of explaining and predicting "what a subject will do next". More particularly, given some general contest, relative to which a subject has learned a number of relevant rules, the question is why does he select the rule that he does use. This turns out to be a very complex problem and, oddly enough, it turns out to have much in common with the perennial problem of motivation. Although our work* in the area is also just beginning, preliminary evidence suggests that, in general, subjects tend to select the path of "least resistance". The major problem, of course, will be how to characterize a path of least resistance.

The final line of research I will mention concerns the mechanisms that people use to solve problems. To date, most of my work in the area has been of a theoretical nature, but by a stroke of luck, it turns out that one of my earlier experiments** bears more or less directly on the major hypothesis involved. Loosely speaking, the hypothesis is that if a subject does not have a rule available for achieving a given goal, control automatically shifts to the higher-order goal of defining such a rule. While it is basically a very simple idea, this hypothesis has provided an adequate base for analyzing some fairly involved

*Francine Endicott is working with me.

**Scandura, J. M., Learning verbal and symbolic statements of mathematical rules. 1967 (*Journal of Educational Psychology*, 1967, *58*, 356–364.)

ERIC

mathematical problems, such as those discussed by Polya in his many books on mathematical problem solving. Needless to say, we are planning further work in the area.

In doing this research, I have found it useful to distinguish between what might be called "memory free theorizing", such as that described above, and theorizing which involves, as well, the limited capacity of human subjects to process information. This distinction has helped lead me to the beginning of what seems to be a rather simple (but apparently adequate) theory of complex learning and behavior. Although it is much too soon to determine the ultimate value of such theorizing, I am hopeful that research of this kind will have very important implications, both for the sort of future research that will be done in psychology and more importantly, for educational practice in our schools.

(4) SOME MATERIALS FOR TEACHER EDUCATION DERIVED FROM RESEARCH

While we have barely scratched the surface in this research, our theorizing has motivated several very practical projects in teacher education. Our first project is aimed at content for elementary school teachers. The materials we have developed are different from existing materials in that they emphasize what might be called the concrete behavioral foundations of mathematics. Furthermore, they integrate certain aspects of methodology with the content in a rather unique way. In particular, they deal with what might be called mathematical processes or higher-order intellectual skills, which are involved in *doing* mathematics, but which are not typically associated with mathematical content.

The processes we have singled out fall into three of the general categories mentioned above. The first is the ability to detect regularities. The second is the ability to reason deductively and the third, is the ability to interpret verbally presented information—or to learn by exposition. The text we have prepared develops each of these processing skills individually in the first chapter. Then the reader is asked to apply these skills in working with the mathematical content in the remainder of the book.

Chapter 2	Fundamental Mathematical Ideas—sets, relations, functions and special cases thereof
Chapter 3	Logic and Set Operations
Chapter 4	Mathematical Systems, Theories, and Relationships Between Systems
Chapter 5	The System of Natural Numbers
Chapter 6	Systems of Numeration and Arithmetical Algorithms
Chapter 7	The System of Positive Rationals
Chapter 8	The System of Integers
Chapter 9	The System of Real Numbers and Further Extensions

The final chapter ends, showing how algebra, the study of discrete entities, meets up with goemetry, the study of continuous variation, through the limit process.

This text is essentially complete and is being used in one of our classes experimentally before the final version is published by Prentice-Hall. We are, at the moment, involved in fine-tuning the text and developing exericses for it.

Most of the material in the book is traditional and, yet, the book is not traditional in many ways. For one thing, it is the first attempt that I know of to discuss mathematical processes in a systematic way. Furthermore, it is also the first attempt to apply such processes to a wide variety of mathematical content. The text and the exercises are designed to have the reader learn the mathematical processes which are useful in the various content sections by actually identifying and using such processes. There is in the book, too, a strong emphasis on the relationship of mathematics to the real world, an emphasis that does not exist in other texts for elementary school mathematics teachers.

We have also tried to emphasize the relationship between the various mathematical topics treated in the book. That is, we have attempted in our persentation to build on the similarities and differences between the various topics. This, I think, not only gives the reader deeper insight into the material, but it also provides for continual review by encouraging the comparison of new material to material which has already been learned.

It should be remembered, though, that this is only a small beginning and certainly not the full flowering which I envisage in this area. In fact, we are now planning a far more ambitious project which will attempt to deal systematically with each of the areas I mentioned in section (2). In particular, we want a text which deals with the nature of mathematical knowledge. We have already done most of the needed research,* our problem is to translate these ideas into a form that will be suitable for teachers. Another goal is to identify the specific objectives of the elementary school mathematics curriculum and to devise a sound rationale for their existence and/or modification. Finally, we hope to identify, describe, and illustrate general teaching principles which the teacher will be able to use in a wide variety of situations. In this work, we plan to draw heavily on our own theorizing about complex structural learning. Since teaching refers to what one does to promote learning, however, we shall have to translate this theorizing (which only tells us things about learning) into a practical form that teachers can use. In particular, we need to ask what one must do in order to *promote* learning. In this sense, we need to look at teaching, you might say, as a mirror image of learning.**

In outline form, *then*, what I have attempted to do in this paper is to (1) indicate why basic research in mathematics education is badly needed (2) to identify some of the kinds of information which every good mathematics teacher needs (3) to describe some of the basic research which we have under way and also to mention some of the implications of this research for further development in mathematics education and behavioral research generally, and (4) to describe some of our current developmental activities in teacher education in mathematics.

In conclusion, I want to emphasize that our developmental activity has been motivated by two things: first, a concern for needed improvements in courses in teacher education and second, the conviction that basic research in mathematics and structural learning can and will provide the basis for such improvements.

*To be reported in a monograph I am editing on *Mathematics and Structural Learning*, to be published by Prentice-Hall next year.

**See also Gage, N. L., "Theories of Teaching." *Theories of Learning and Instruction*. (Edited by Ernest R. Hilgard.) Sixty-third Yearbook, Part 1, National Society for the Study of Education. Chicago: University of Chicago press, 1964, pp. 268–285 and Scandura, J. M., Teaching—Technology or Theory. *American Educational Research Journal*, 1966, 3, pp. 139–146.

ERIC

(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, No. 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 Paedegogica 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 on mathematics learning--a reaction to Suppes' 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
5, 461-469.)

34. Scandura, J. M. and 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 "Straw man") 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, 19[illegible], 4, 31-34; 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.)
42. 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 education. 1968. (School & Society, 1969, April, 227-228.)
47. Scandura, J. M., Barksdale, J., Durnin, J. and McGee, R. An unexpected relationship between failure and subsequent mathematics learning. 1968. (Educational Studies in Mathematics, 19[illegible], [illegible], 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. (Highschool Journal, 1970, in press.)
50. Scandura, J. M. The role of rules in behavior; Toward an operational definition of what (rule) is learned. November 25, 1969.
51. Scandura, J. M. & McGee, R. An exploratory investigation of basic mathematical abilities of kindergarten children. December 1, 1969.