DOCUMENT RESUME

ED 040 036 SE 007 938

AUTHOR	Scandura, Joseph M.
TITLE	The Role of Rules in Behavior, Toward an Operational Definition of What (Rule) Is Learned.
INSTITUTION	Pennsylvania Univ., Philadelphia. Graduate School of Education.
PUB DATE	25 Nov 69
NOTE	44p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.30
DESCRIPTORS	*Cognitive Processes, Concept Formation, *Learning, Mathematical Concepts, *Mathematical Linguistics, Mathematics Education, Models, Research, Structural Analysis
IDENTIFIERS	Set-Function Language

ABSTRACT

This report describes research conducted regarding the development of a precise scientific language called the "Set-Function" Language (SFL) which was formulated in terms of sets and functions. The SFL retains many of the basic aspects of cognitive formulations but also provides more rigor than most of the other scientific languages. The SFL characterizes notions like rules, decoding and encoding processes, "chaining", reference mechanisms, and higher order rules in a precise manner. The report claims that the SFL is more adequate than the existing S-R and cognitive languages for formulating research on meaningful learning. Also, the author presents a partial solution to the problem of "what (rule) is learned." (FL)

November 25, 1969

MAY 12 1970

THE ROLE OF RULES IN BEHAVIOR[1]

Toward an Operational Definition of What (Rule) is Learned

Joseph M. Scandura

University of Pennsylvania

Abstract

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.

THE ROLE OF RULES IN BEHAVIOR[1]

Toward an Operational Definition of What (Rule) is Learned

Joseph M. Scandura

University of Pennsylvania

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 vs. general training (e.g., see Scandura, Woodward, & Lee, 1967).

---

[1] Portions of this article were presented at the APA convention in Washington, D. C., September 1, 1967. The author would like to thank Charles N. Cofer for his most valuable suggestions for the improvement of the manuscript, and John H. Durnin for his general assistance in the preparation of this article.

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

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 specifically to behavior, and don't 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 (Piaget), 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 utilize 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 is there 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.[2]

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.

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, (sign) reference, and higher-order relationships. Finally, with the addition of an extremely weak theoretical assumption about how Ss perform, we propose a partial solution to the important problem of "what (rule) is learned."

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, Greeno and Scandura had their Ss 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 Ss 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 Ss 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) Ss overlearned the list shown in Figure 1.

Insert Figure 1 about here

Prior to learning the list, both the Ss and the experimenter agreed on the relevant dimensions and values--si... (large-small), color (black-white), and shape (circle-triangle). The Ss were told to learn the pairs as efficiently as they could since this might make it possible for them to respond appropriately to the transfer stimuli. After learning, the Test One stimuli were presented and the Ss 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 Two stimuli were presented in the same manner. The results were clearcut. All but three of these Ss gave the responses "black" and "large" respectively to the Test One stimuli (see figure 1) and also responded with "white" and "small" to the Test Two 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 One 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 Ss discovered the two underlying principles during List One 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 a subject 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 and Scandura, 1968; Scandura, 1969b, 1967b; Scandura _et al._, 1967; Scandura and Durnin, 1968).

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[3]--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 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 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 fom 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 (i.e., domain, range, and connecting-rule). The denotation (i.e., S-R instances) of 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 domain, range, and connecting rule.

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[4] of overt stimuli (1 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.

The underlying rule, however, is 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 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 series in question. The domain of this rule is the set of all triples of values that the dimensions, A, L, and N, may take on (e.g., A = 1, L = 7, N = 4). These triples may be viewed as (composite) properties of the entities denoted by the overt stimuli (e.g., "1 + 3 + 5 + 7"). We may refer to these critical properties as response <u>determining</u> (D) properties. The range is the set 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 <u>sums</u>, 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 (i.e., the domain), and O to the combining <u>operation</u> or transformation by which the derived properties (of the responses) in the range (R) are derived from the properties in D.

We note parenthetically that accounting for such behaviors as adding arithmetic series in terms of rules is <u>not</u> 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 <u>classes</u> 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. We have 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 our 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 my knowledge has explicitly considered as stimuli what amount in a related context to S-R pairs (i.e., associations). (Note: Any given entity may serve 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 as a minimum have such connections as

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

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.

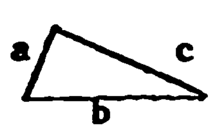
As a second and somewhat more subtle example, consider the task of adding "4" and "3" in column addition. If embedded in a problem like $\begin{array}{r}41\\+32\\\hline\end{array}$, the tens digit in the sum is "7". However, if the problem involves carrying, like $\begin{array}{r}47\\+35\\\hline\end{array}$, 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 <u>associations on associations</u>, 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" $\rightarrow$ "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.

<u>Decoding and Encoding Processes</u>. The distinction we have made between overt stimuli and response, on the one hand, and properties (of the entities denoted by these stimuli), on the other, raises the question of how

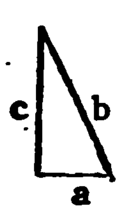
the decoding and encoding "gaps" are to be filled. In particular, rules operate on properties of overt stimuli and not directly on overt stimuli (or, more accurately, on properties of the entities these stimuli denote). Similarly, they generate properties (of overt responses) but not the responses themselves. 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

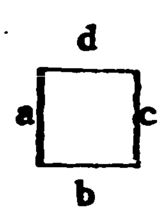


and $(24 + 16) \div 17$,

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 a 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.[5] 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 the 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 composite function provides a ready means for representing multi-stage rules of this sort. In this case, each of the constituent rules is represented by a simple function. Thus, if the functions, F_1, F_2, ..., F_n represent n simple rules, such that the range of F_i is the domain of F_{i+1} $(i = 1, 2, ..., n-1)$, then the composite function $G = F_n \ldots F_2 F_1$ represents the composite rule. Complex procedures (e.g., see Suppes & Groen, 1967; Groen, 1967), which involve branching, can be handled in a similar fashion but discussion here would be an unwarranted digression. (For details, see the author's Mathematics and Structural Learning. Englewood Cliffs: Prentice-Hall, forthcoming.)

Reference. - Although we avoided going into details above, the nature of our 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. (1) 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.) (2) 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 (given) 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 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 a S has acquired particular meanings? There seem to be at least two ways in which this might be done: (1) by determining whether or not the subject can paraphrase or otherwise describe the intended meaning, and (2) 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. A 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 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 (forthcoming).)

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 rule-like--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, we first note that we *can* represent the connection between signs and their referents 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. (That is, 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 <u>non</u>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, we first consider what might be called <u>elemental symbols</u>, 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. <u>The arbitrary nature of symbol reference has both limitations and advantages</u>. Perhaps its most important limitation is that symbol reference is non-generalizable. 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, Ss 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 our 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 ten; 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 ten" may be interpreted to mean "Let the elements in each set in the denotation correspond to ten elements in a corresponding set in the denotation of the units-digit." 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 first meaning class. Thus, corresponding to the set, {|||||}, in the units meaning class, we construct the set, {|||||||||| |||||||||| ||||||||||}, in the tens class, where each of the three bundles contains precisely ten vertical lines. For details on how such interpretative rules are constructed, the reader is referred to Scandura (forthcoming).

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 Ss were trained with neutral materials [e.g., 3 (5 + 4 (3 + 2))] until they could reliably work with parentheses. Then, half of the Ss 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 Ss could almost invariably apply rules, like [([X] + [Y])/[Z]], to instances once statements of these rules had been committed to memory. The Ss 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[6] have properties in common with the entities they denote; they denote in a non-arbitrary 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 (non-degenerate) 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 referent-like, icons retain 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,

<u>Insert Figure 2 here</u>

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

Insert Figure 3 here

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.[7]

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 (non-degenerate) 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 our 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 our 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 difference 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." (That is, first multiply by two and then by five.)

The point is that many people appear able to combine pairs of given rules into corresponding <u>composite</u> rules. Thus, for example, given rules like,

x yards $\longrightarrow$ 3x feet,

and

y feet $\longrightarrow$ 12y inches,

many <u>S</u>s can combine them to form composite rules, like

x yards $\longrightarrow$ 3x feet $\longrightarrow$ 12(3x) inches.

(Using arrows is a convenient way to represent the denotation of rules. Thus, for example, x yards $\longrightarrow$ 3x feet is interpreted to mean $\{$(x yards, 3x feet)$|$ x 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: $\{(R_1, R_2), R \mid R_1$ and R_2 are (equivalence classes of) rules, and R is the rule formed from R_1 and R_2 by composition$\}$.

Lou Ackler and I have a study now underway in the Penn Laboratory which demonstrates, conclusively I think, the behavioral reality of such higher-order rules (Scandura, 1970). Given the necessary constituent rules, as above, Ss, ranging in age from kindergarten to post-graduate, 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, 1967) 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: (1) the first few sums, Σ^n, (2) the corresponding values of n, and (3) 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 $\Sigma^n = n \cdot f(n)$." This rule can also be analyzed in the same general way, but the analysis is not as simple as the examples given above. We simply sketch the main ideas and refer as before for more details to Scandura (forthcoming). (a) The inputs of the higher order rules 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 ouput rules are also associations, this time between classes of series (e.g., 1 + 3 + 5 + ... + (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 → 4, 1 + 3 + 5 → 9, 1 + 3 + 5 ... ...) into output associations (e.g., 1 + 3 + 5 + ... + ...

As a final example, we simply poin... ...rse relation between addition (i.e., the rule) and ... ...ne instance of a higher order rule by which an ... ...e.g., multiplication) can be mapped onto it... ...e.g., division).

In ... ...rder rules are in some sense orthogonal to the ... ...on which they operate. Lower order rules act on classes of ... ...nd 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 <u>everything</u> that a <u>S</u> knows in order to predict what he will do in a given situation. Much of the <u>S</u>'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 a <u>S</u> fails to do what is expected of him, he is branded as uncooperative. Specifically, knowing a <u>S</u>'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 a <u>S</u> 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 a <u>S</u> 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, our 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. (It says nothing, of course, about which rules S may have used to generate his failures.)

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 a S has demonstrated his ability to find the sum (2500) of the arithmetic series 1 + 3 + ... + 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 $(\frac{A + L}{2})N$; the rule N^2, which applies to all arithmetic series of the form $1 + 3 + \ldots + (2N - 1)$; the direct "association" between the series, 1 + 3 + ... + 99, and its sum, 2500. Thus, "what is learned" might be denoted by the class,

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

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

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

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 + ... + 100, then the class "what is learned" is further reduced to the set

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

The rule N^2 is eliminated since it is not applicable to the third test series (i.e., 2 + 4 + ... + 100). Suppose, on the other hand, that the S is successful on the first two test stimuli (i.e., 1 + 3 + ... + 99 and 1 + 3 + ... + 47), but not the third (i.e., 2 + 4 + ... + 100). Then, according to 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^2, 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 a 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 a 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 a 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, forthcoming).[8]

The really basic question, of course, is whether or not the actual behavior potential of particular Ss is compatible with this view. Fortunately, my students and I have collected a fairly substantial body of data over the past few years which suggests that this is the case (Scandura, 1966, 1967b, 1969a; Scandura, Woodward, & Lee, 1967; Scandura & Durnin, 1968; Roughead & Scandura, 1968). Whenever the response given by a 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, Leitenberg, & Richter (1964), for example, have used performance on reinforced trials to predict performance on non-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, forthcoming; 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 formulation 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*. (My emphasis on this point is 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 (forthcoming).

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 (Scandura, 1966a, 1967a, 1967b, 1969a; Roughead and Scandura, 1968; 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.[9] 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 our own research we have been led to ask a number of questions 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 extra-scope 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., non-overlapping) 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 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 Ss 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 (forthcoming).)

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. Technical Report No. 118, Stanford: Institute for Mathematical Studies in the Social Sciences, 1967.

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, 331-38.

Levine, M., Leitenberg, H., & Richter, M. The blank trials law: The equivalence of positive reinforcement and non-reinforcement. Psychological Review, 1964, 71, 94-103.

Hocutt, M. On the alleged circularity of Skinner's concept of stimulus. Psychological Review, 1967, 74, 530-532.

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, 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.

Osgood, C. E. On understanding and creating sentences. American Psychologist, 1963, 18, 735-751.

Reitman, W. R. Cognition and thought: An information-processing approach. New York: Wiley, 196_.

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, 101-103. (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. University of Pennsylvania, Mathematics Education Research Group Report No. 54, 1970.

Scandura, J. M. Mathematics and structural learning. Englewood Cliffs: Prentice-Hall, forthcoming.

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 meeting of the Mathematics Education Research Group, Philadelphia, April 4, 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., & Jenkins. Thought experiments for organisms too complex to be models. Paper presented at the Mathematics Education Research Group meeting, Philadelphia, April 4, 1970.

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, 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: N.C.T.M., 1967, 35-43.

Tolman, E. C. There is 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.

FOOTNOTES

[1]Portions of this article were presented at the APA convention in Washington, D. C., September 1, 1967. The author would like to thank Charles N. Cofer for his most valuable suggestions for the improvement of the manuscript and John H. Durnin for his general assistance in the preparation of this article.

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

[2]In this regard, Shaw and Jenkins (1970) have recently presented cogent arguments as 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.

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

[4]By an equivalence class of overt stimuli (responses) or an <u>effective</u> 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, I (1966, 1967a) have used the term _functionally distinct_.

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

[5]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(overt) - r_m and encoding, to s_m - R(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.

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

[7]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. 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.

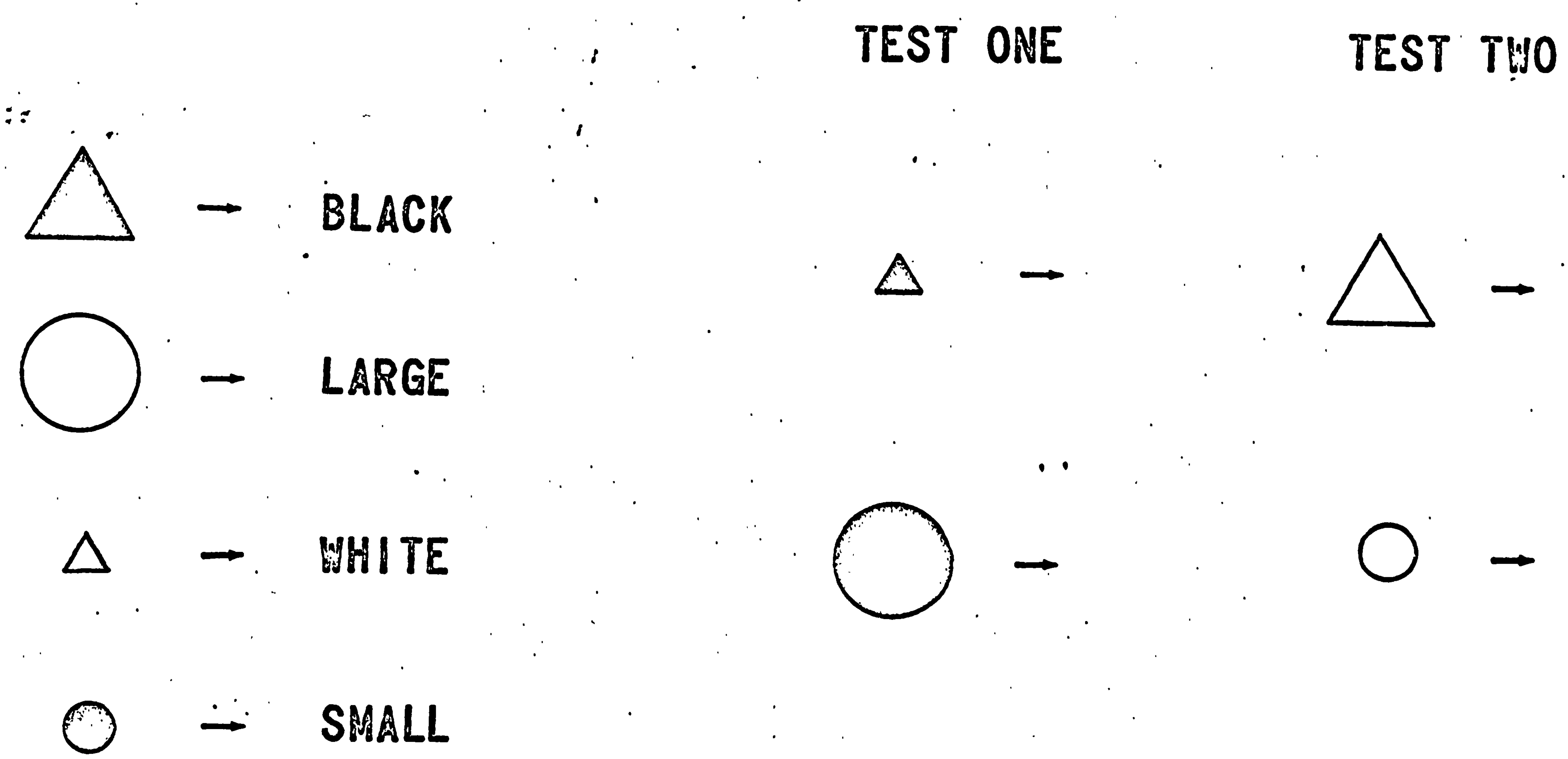
[8]I originally felt that a stronger assumption of stronger assumption of this sort was needed--in particular, that S will continue using the same rule as long as his goal remains unchanged and feedback otherwise indicates that he is responding in an appropriate manner (Scandura, 1969b). While this Einstellung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

[10]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 a 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 [illegible]dura, 1970.)

FIGURE CAPTION

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


TEST ONE
TEST TWO
BLACK
LARGE
WHITE
SMALL

Sandra

Fig 2

Top

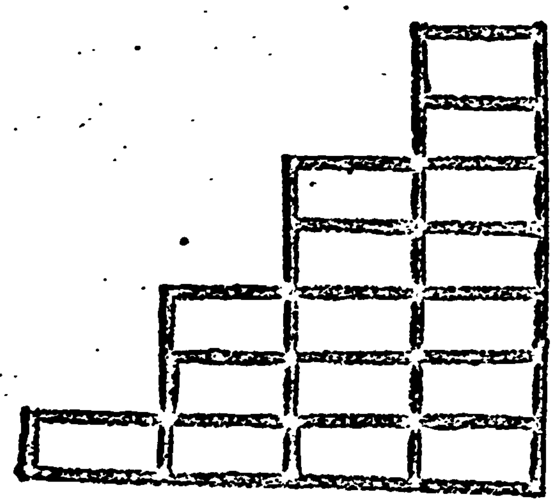


Fig 3