This retrospective is based my long history of research on the teaching learning process. It was motivated initially by communications with the TICL publisher, with whom I have worked for many of the 50 years in question. This effort was reinforced by comments and suggestions from colleagues some of whom played a significant role in my research.

We begin with my long history of basic research on teaching and learning and the structural learning that evolved therefrom. Our success in building on this theory and research in building our AuthorIT and TutorIT systems has played a significant role in motivating what follows. I begin with some initial observations.

My current AuthorIT and TutorIT systems are designed to model human tutors (see www.TutorITweb.com). Unlike other contemporary adaptive learning systems, TutorIT tutorials are designed to interact with students as would a good human tutor or teacher. Other contemporary adaptive learning systems are eclectic in nature, build on assumed learning theories (e.g., Carnegie Learning, ALEKS) or use Big Data systems that make pedagogical decisions automatically (Knewton) based increasingly large collections of data about student accomplishments and preferences.

AuthorIT and TutorIT build on decades of basic research and theory development in the Structural Theory (SLT). Designed from inceptions as a theory of teaching and learning, SLT is fundamentally different from other “learning theories”.

There are two major differences in the way AuthorIT and TutorIT work. AuthorIT makes it possible to specify what needs to be learned for success -- with arbitrary degrees of precision. TutorIT takes this information as input and makes ALL pedagogical decisions automatically. This approach makes it possible to pre-specify the degree of mastery while completely eliminating the need to program pedagogical decision making.

**Introduction**: My professional life has always been goal driven. During the early years, my biggest challenge was in sports. Tipping the scales at 150 pounds (still close) my future in football was strictly limited. I competed in high school, college and Olympic style wrestling, winning a national title and OSW award in 1955, didn’t quite win a berth on the ’56 Olympic Team and coached the Syracuse championship team while working on my PhD in mathematics.

Academics came easily during this period and began as a sideline. Soon, however, I became absorbed in research challenges. A fellow post-doc in experimental psychology commented that his professional goal was to “add a few grains of sand on the academic beach”. I remember thinking at the time -- if sprinkling sand was all I could do, I rather spend my time sunning on the beach.

I hope this retrospective will give some sense of my goals, opportunities, challenges, successes and failures over the years. Looking back, my research and professional life took place in two
This retrospect is a work in progress – pp. 1-10 draft plus List with Notes

phases. Phase one focused on defining a field of basic research that did not exist at the time. In retrospect, this work focused on finding answers to four basic questions:

1. What does it mean to know something, initially focused on mathematics?
2. How do people acquire new knowledge – e.g., solve problems in mathematics for which they have not been taught solutions?
3. How can one (e.g., a teacher) determine what mathematics a student does and does not know at any given point in time?
4. How can this all be put together in a theory of teaching and learning, a theory that is at once comprehensive, rigorous, precise and operationally defined?

I didn’t want to just improve mathematics and science education, my initial areas of training. I wanted to understand the process of teaching and learning at a deeper level. During my high school and college years, I had developed a deep appreciation for both formal systems in mathematics and classical theory in physics. The former focused on the importance of identifying fundamental assumptions underlying formal systems generally, and deriving implications therefrom. Classical physics covered physical phenomena that could be observed more or less directly (albeit often requiring a telescope). Building on the work of scientists like Copernicus, Galileo and Kepler, Newton pulled it all together in grand theory that was at once rigorous, comprehensive, explanatory and predictive.

Ignoring gaps later exposed by the likes of Maxwell and Einstein together with Heisenberg’s Uncertainty Principle, Newtonian physics offers a rigorous and comprehensive framework that integrates a broad range of physical phenomena. It offers a rigorous, comprehensive and testable theoretical framework – most importantly, a deterministic theoretical framework that has and continues after centuries to serve a useful purpose.

When I first entered the field of educational research in the 1960s, there was nothing comparable in either education or psychology. Looking back, my major goal both then and now has been to create, test and refine the equivalent of Newtonian theory in behavioral science. I wanted to understand, be able to predict and even control the behavior of students in specific (initially math learning). I felt first intuitively and later more formally that the study of teaching and learning might and should be placed on an equally rigorous foundation.

Looking back, my research has taken place in two major phases. In large measure, this was a result of both progress and opportunity. The history that follows draws directly on my publications and the software systems we have built over the years. My early work took place in a world where the very notion of deterministic theory in behavioral science appeared unthinkable. I first proposed the idea at one of our Structural Learning conferences at Penn (in 1970). This was followed by the publication of my “Deterministic Theorizing in Structural Learning: Three Levels of Theorizing” (#55 in my list of publications) in 1971. Some seemed fascinated with the idea, and the paper became one of ISI’s most widely cited papers. Most, however, trained in standard experimental methods found it hard to conceive of predicting the behavior of individuals in specific situations.

To counter this near universal belief among psychologists and educational theorists, I took great pleasure in talks on the subject. Relaying a hypothetical event in physics, I challenged
listeners to imagine Galileo at the Leaning Tower of Pisa. Instead of dropping a large stone and small stone, imagine if Galileo had instead (as I did when doing so was still allowed at the tower) dropped a small stone and a feather. How differently physics might have developed. Instead of focusing on what happens in a vacuum under idealized conditions, one can only imagine instead “an alternative science of droppings” – calculating and documenting the average rates of fall of different kinds of objects. Read on: Our research demonstrates that deterministic results are not only feasible in research on human behavior, but in many cases preferable.

**Part I:** Part one in this report focuses on basic research and theory. My goal during this period was to help shape development of the field while devising a rigorous, deterministic theory reminiscent of classical physics – a theory that would make it possible to understand, predict and, yes, even control the learning process. It quickly became clear that achieving this goal would require far more than mathematics itself, or the action research characteristic of the so-called “new math” of the early-mid 1960s. It also would require fundamental revision of what mathematical, experimental, developmental psychologists had to offer.

Among the investigators and works that most influenced my thinking during this period (with apologies for inevitable omissions) included Polya’s “Mathematical Discovery”, Dienes’ work helping young children learn mathematics, mathematical foundations (e.g., Gödel’s incompleteness theorem), Bruner, Goodnow & Austin’s “A Study of Thinking”, research on concept and rule learning (e.g., Bourne, Kersh, Wittrock), the rigorous methods used in experimental psychology (Atkinson, Melton, Postman et al), Suppes, Estes and colleagues work in mathematical psychology, Gagne’s conditions of learning and to a lesser extent Piaget’s stages of cognitive development.

Dissatisfied with the informal, incomplete and/or statistical nature in this work, I sought to emulate goals adopted by renaissance physicists. By analogy, my goal was to understand, predict and even control human learning in specific situations (under idealized conditions). I sought to develop and test a simple, cohesive, yet deterministic and testable theory making it possible to explain predict and even control the behavior of individual students in specific problem situations.

**Part II:** Part two is planned to focus on our work in automation. Finding both hardware and then current software technologies inadequate for full application of the SLT, we first turned our attention to software engineering -- building needed technical foundations. Initially, this required extending, applying and refining the process of Structural Analysis (SA) in software engineering. This work in turn enabled us to formalize key essentials of the SLT and to implement those essentials in automating the tutoring process.

Our initial goal in this area was to create software that is correct by design. Given any content domain, this work brought home the critical need to systematize our previously informal processes of Structural Analysis (SA). Initially used to identify what needs to be learned for success in any given content domain, we found that SA also offered a rigorous foundation for designing and implementing complex software systems.
As we shall see, solutions to two key problems in the initial formulation of the SLT were a welcomed side effect of applying Structural Analysis in software engineering. This work led to highly efficient automated technologies for both creating and delivering systems that model the human tutoring process. There is more to do, but key essentials have already been realized in our AuthorIT authoring and TutorIT delivery systems. These systems are now fully operational. They offer unprecedented control of pedagogical decision making while dramatically reducing the cost and effort of development.

But let’s not get ahead of ourselves. In Part 1, we focus on the foundational theory and research on which Part 2 rests.

**PART 1 – Theoretical and Empirical Research Proving the Structural Learning Theory**

Given my background in mathematics and education, my initial research took place in the context of the so-called new math. A common theme in this work was that students learned best if they discovered mathematics on their own. In retrospect, this was not a surprising hypothesis because inventors of the so-called “new math” were almost uniformly mathematicians -- for whom mathematics came easily. Unfortunately, converting students into little mathematicians went only so far.

The goal of my dissertation research was to understand the teaching-learning process at a deeper level. If mathematical discovery is truly better, why is it better? After a couple years of intensive work, I discovered that what a student knows when information is given is far more important than how it is learned – by discovery or by expository. Moreover, the ideas involved appeared to transcend mathematics per se.

My work took place in a context largely devoid of serious research. The new breed of math educators called it “action research”. A good deal of my professional energies during this period went into identifying the need for and motivating serious foundational research in math education. In parallel, I got deeply involved in experimental and mathematical psychology. Although disagreeing with the dominant focus of this research, the discipline in S-R research in those days was far more advanced (and replicable) than that in educational psychology.

As many will recall, Gagne in his influential conditions of learning proposed a category of types of learning. Each required a different way to learn. Most relevant for education were: S-R learning, chains of S-R associations (verbal or otherwise), concept learning, rule learning and problem solving. In short, S-R associations were considered basic. The others essentially were represented as various combinations of S-R associations.

After various explorations, a few of us helped move the focus toward rule learning. I ultimately found that knowledge could better be represented in terms of rules. Rules initially consisted of D, O, R triples, operations (O) acting on task domains (D) and generating solutions/ranges (R). In this framework, concepts and associations were viewed as special cases. Problem solving involved (higher order) rules operating on other rules and generating new (solution) rules.

In my research, rules rather than S-R associations became the basic unit of behavior. Concepts and associations became special cases. Problem solving was required when no learned knowledge was sufficient. Problem solving in my view required higher order rules operating on and generating other rules.
During this period, with help from a large contingent of graduate students, I published a monograph and several books (M1, B1-B8) along with a large number of studies in experimental, educational and developmental psychology (Nos. 2- to about 131). For their help I would like to particularly single out Merlyn Behr, William Roughhead, Jay Norman Wells, George Lowerre, John H. Durnin, Don Voorhies, Walter Ehrenpreis, Judy Anderson, Joan Barksdale, Bob McGee, Francine Endicott and Jaqueline Veneski and Wally Wulfeck (forgive me for those overlooked). John and Wally contributed over a long period of time to some of our most definitive basic research and deserve special credit. Other colleagues contributed directly relevant research in mathematical psychology (e.g., Suppes, Greeno), mathematics and science education (Z.P. Dienes, Paul Rosenbloom, Jack Nelson, David Johnson, G. M. Thomason) and educational psychology. Single out for her help over the longest period of time in so many ways academic and otherwise is my wife of over 50 years (Alice B. Scandura). After bearing and raising our four children, she returned to graduate school returned to earn her PhD at Penn. Her work applying SLT principles in Piagetian research remains definitive to this day (B8).

An increasingly common theme in this research was that the more precisely we could identify what needed to be learned for success, the less necessary the associated empirical research became.

The focus of my research, initially, was in mathematics education. It soon broadened to include parallel research in experimental, educational and developmental psychology and artificial intelligence. This ultimately led to the first iteration of the Structural Learning Theory (SLT) and a broad range of basic research reinforcing basic assumptions in the theory. From there, the focus of my research during Phase II shifted first to software engineering and ultimately to the AuthorIT authoring and TutorIT delivery systems that dominate current work and enable us to model the human tutoring process.

Part I focuses on foundational research leading to and supporting the Structural Learning Theory (SLT). The following commentary on my long list of publications, many with the grateful support of my former students (and now colleagues), gives some sense of the broad range of challenges we faced, the research we conducted, along with our accomplishments and disappointments.

My first goal was to understand mathematical (and other) problem solving. In parallel, considerable effort was made to motivate broader interest in research within the math and science education communities. I also worked to extract relevant information from basic research in experimental psychology and artificial intelligence. All played a role in the creation, testing and refinement of the Structural Learning Theory (SLT).

The following commentary closely parallels my publications. After considering various alternatives, it became increasingly clear that following the printed record offers as good and perhaps the best and most accurate sense of the challenges I faced, along with my solutions other accomplishments and disappointments. In short, what follows is contemporary commentary on my publications (more accurately comments on those of my publications for which electronic versions are available). Others are listed with little or no commentary.

Each publication is listed in turn followed by a link to an electronic version of the article along with a summary and/or contemporary commentary consistent with its importance. The nature
and depth of these comments depends primarily on my sense of historical significance and/or on-going importance.

   --*done at Syracuse University*
   Oddly enough, my first published article had nothing to do with my academic research. I include mention because it documents my earliest goal directed activities – in this case winning the EIWA team championship and coaching the first (and third) black national wrestling champions.

   --*Syracuse University and SUNY Buffalo*
   My dissertation research at Syracuse University¹ (numbers 2 and 3) was motivated by an attempt to understand the problem solving process and its implications for education. Research in education at the time was at best ambiguous and heavily dependent on student ages and abilities, materials used in individual studies and educational settings. During this time period, the focus in math education was on discovery learning which played a central role in the so-called “new math”.

   Materials used to study problem solving at the time varied so greatly that it was almost impossible to discover generalities. It was especially difficult to bridge the vast chasm between problem solving research in the classroom and in the laboratory. Much of this difficulty stemmed from the wide variety of research materials used.

   To make sense of conflicting often contradictory research, I conducted a series of experiments, in which we obtained detailed tape recordings. To minimize individual differences in prior knowledge, we constructed abstract card material that could be used in both classroom and laboratory settings. This material was a significant extension of the card tasks used in Bruner, Goodnow & Austin’s “A Study of Thinking” published in 1956. These card tasks were equally novel for students at a wide range of educational levels. They enabled us to: a) minimize the effects of prior knowledge and problem solving sophistication and b) make it possible to construct a wide variety of related multi-stage problems.

   Specifically, this abstract (card) material had both semantic and syntactic characteristics similar to school mathematics. These abstract card tasks were designed to minimize the effects of prior knowledge. They were largely independent of prior learning, mathematical or otherwise, and were in used in a series of experiments comparing different modes of problem solving research -- in both laboratory and classroom settings.

   Following is a small sample of the problems used in this research. For more detail, please see linked publication 2.

---

¹ *This work was published while I was an assistant professor in mathematics education in both the School of Education and the Mathematics Department at SUNY Buffalo. I also was invited to join the Departments of Psychology and Statistics but decided to accept a Research Professorship at Florida State University beginning in 1964.*
Side Note 1: My dissertation advisor, Robert B. Davis, was a mathematician from MIT, one of the originators of the new math with his “Madison Project”. Other members of my dissertation committee were Frank di Vesta, a respected experimental psychologist, and two statisticians, Lew Cote in the math department and another in education. As the reader might imagine, each dissertation committee member had a different perspective on my research. I could have said “like blind men examining an elephant” but that would be unfair.

These were all good people and even I certainly wasn’t exactly sure how best to attack the problem. Lew Cote put it best. At one of many meetings of my dissertation committee, he said something to the effect: “... when you talk with (Prof.) di Vesta, I suspect you emphasize the mathematics but when you talk with me, you emphasize psychology ...”. This was probably a true statement. Coming from a home where my father (a skilled barber) dropped out of school
in the fourth grade and a mother who while first in her class stopped after 8th grade\textsuperscript{2}, I had what most today would consider an unrealistic view of what a PhD dissertation was supposed to be. I likened it to a major contribution to knowledge, something along the lines of what Galileo or Newton did – in any case something that no one had done or was able to do before.


Led primarily by mathematicians and supported heavily by the NSF, discovery methods of instruction gained increasingly widespread support in the 1960s. Direct presentation purportedly led to verbal glibness by the learner without true understanding. Many teachers, however, still used some form of exposition. They believed that this made the educational process more efficient. Among others, Gagne supported this point of view by asking, "Is it [verbalization] bad for problem solving because it is expressed in words, or is it bad because it does not have enough words?"

The literature at the time was at best confusing and often contradictory. Studies conducted under sound experimental conditions supported both points of view and were at best contradictory. Studies conducted in classroom settings were even more confusing. In short, too much precision, too soon, is at best unrealistic and at worst misleading.

The initial goal of my research was to identify important variables and interrelationships, to provide a framework for future more precise experimentation. A series of experiments were conducted, each comparing one form of expository and discovery learning. Each experiment included both statistical analysis of results and intensive post-analysis based on verbatim recordings.

The results of these studies made it difficult to conceive of any explanation of the teaching-learning process that did not consider S-feedback and the nature of the materials used. Perhaps the most striking conjecture to be drawn from this research was the use of "timing" as a coordinating variable. Information given to soon or late served no useful purpose. Indeed, given to soon, it could interfere with deeper learning.

Although evaluated subjectively in the present context, by analyzing tape recordings, informal assessment of S learning, subsequently led to more objective measures. It led to methodologies that made it possible to infer the current cognitive state of learners based on S-feedback.

While “timing” may be characterized in terms of variables such as rate of presentation, complexity of material, S-ability and/or prior knowledge, such reduction is not necessarily desirable. To the contrary, we found that timing made it possible to deal, at once, with the cumulative effects of commonly manipulated variables. Information given indirectly often acts as a catalyst. Instances of the desired concepts are presented and attention is directed so that the learner may abstract for himself. Although the discovery methods employed in this

\[\text{\textsuperscript{2} She didn’t go on because she “didn’t have clothes like the other girls going to high school”. After I received my PhD, she did later return to earn her high school equivalency.}\]
research generally were designed to exhibit the basic structure of the material, the Ss, on occasion, inadvertently were led off the main stream. Apparently, when prerequisite learning is inadequate, indirect information is of little value. Indeed, too early presentation may actually inhibit later discovery when the appropriate learning sets have been developed.

Relatively few of the Discovery(D)-Ss in this research had time or were able to discover how to apply given algorithms. However, those that did were generally successful not only on the (R)outine-problems but on the (N)ovel-problems as well. Relatively speaking, those D-Ss who used systematic, yet heuristic, modes of attack seemed to be hindered more by problem complexity than novelty. It is possible that complex problems may overtax the heuristic problem solver's information processing capacity so that he makes mistakes. Novelty may not have this effect when the same well-learned principles are involved.

Overall, methods research had never attained much esteem from behavioral scientists. Very few of them, however, concerned themselves with fundamental problems in education. On the other hand, we found that too little research had been devoted to prudent selection (or identification) of central and meaningful educational variables - variables serving to integrate rather than complicate. If this requires difficult to specify variables, I argued that we needed to learn how to specify them better - this is one of the difficulties besetting any young science. Among those most heavily referenced in my early research were Ausubel, Bruner, Corman, Craig, Dienes, Gagne, Haselrud, & Meyers, Kersh, and Stolurow.

In addition to publication 5. Fraction-names and numbers and 6. Probability Theory – Basic Concepts, my early publications included efforts to motivate support for more research in science and mathematics education: 4. The other role of schools of education, 7. The emerging research role of the subject matter educator. 8. Educational research and the mathematics educator. 9. Federal funds for the improvement of mathematics education. 4. Scandura, J. M. The other role of schools of education. Western New York Study Council Journal, 1964, 15, 11-13 was another attempt to promote the need for more research in schools of education.

Distinguishes notation used to represent fractions and fractions as numbers.

   - SUNY Buffalo
Probability is widely recommended for inclusion in the secondary curriculum. Nonetheless, there was considerable disagreement as to just what probability should be taught. While anything like a full treatment of measure theory is much too difficult for high school students, this can be done for the finite (discrete) case. This paper introduces fundamental notions of probability theory in a way designed to provide motivation for measure theory without giving so much detail as to confuse or so little as to be misleading.

- SUNY Buffalo

As new faculty in a sleepy School of Education at SUNY Buffalo, my colleague Jack Nelson and I argue for the inclusion of research specialists in departments of science and mathematics education. This paper identifies strengths and weaknesses of various approaches to educational research, and describes what we saw as an emerging research role of STEM educators.


- Florida State University

This paper is another arguing for a new approach to research in mathematics education. In retrospect, battling a reluctant bureaucracy was a major factor in motivating my increasing attention to psychological research. This review focused on a major difficulty in research on teaching and learning, wherein method and content are typically varied simultaneously, but not independently. I argued that this approach confounds experimental results. Thus, for example, if a modern mathematics program is taught by the discovery method and a traditional program by traditional methods, then the results on any criterion variable such as problem-solving ability, attitude and the like may be due to methods differences, content differences, or both in some unknown combination.


- SUNY Buffalo and Florida State University

In retrospect, increasing emphasis on research in math education was growing significantly, but not fast enough for my taste at the time.


- Florida State University

In the mid-60s, I proposed that National Council of Mathematics Education (NCTM) publish a Journal of Research in Mathematics Education. To prove the idea, NCTM commissioned me to solicit and edit the monograph *Research in Mathematics Education.* With a range of cutting edge chapters, including both “young upstarts” like myself and senior investigators, like Bob Gagne and Pat Suppes, the monograph was highly successful. Given its success, NCTM formally introduced the Journal of Research in Mathematics Education. JRME was an almost immediate success and continues to this day.

Switching gears, and more or less simultaneously, I received a couple postdoctoral research grants for research in first in mathematical psychology at Indiana University with Bill Estes, founder of mathematical psychology, and next in experimental psychology at Michigan with Art Melton, editor of the influential APA Journal of Experimental Psychology. While retaining my
During this period, I conducted a number of research studies in this area. These studies involved all-or-none learning and transfer, conceptual organizers and short-term memory, all topics in vogue in experimental psychology during the mid-60s. Incidentally, there were many fewer journals and even fewer influential journals at the time. Publication was a far more arduous affair than it is today, with the appearance it seems of a new journal every other day. A quick summary of each follows:

Florida State University
266 Ss were given seven different kinds of familiarization training with complex stimuli, each having only one discriminating attribute. Those familiarized with the stimuli used during Paired Associate learning had fewer errors than controls. Although a significant over-all effect was noted, sub-comparisons revealed that differential reinforcement of the discriminating stimulus cue, during familiarization, was superior to reinforcement of non-discriminating cues only with the control stimuli. The corresponding result with the learning stimuli was in the same direction, however, and no interaction was noted. The former result supported other pre-differentiation studies; the latter findings were harder to explain, but tentatively attributed to asymptotic conditions.

32 Ss in two experiments were familiarized with a set of stimulus and response nonsense syllables and then learned a list of eight pairs. Each of familiarity condition (neither stimuli or responses, stimuli only, responses only, and both) was represented by two pairs, one learned tinder a repetition condition and the other under a non-repetition condition. In Exp. 1 the pairs were presented for 2 sec. each; in Exp. 2 presentation was self-paced. The difference between learning under the repetition conditions was shown to be a function of stimulus and response familiarization. It was difficult to criticize these findings on the basis of the item-selection artifact usually referred to in studies of this type. Results favor an all-or-none learning when stimuli and responses act as single elements. Contrary to the widely accepted notion that paired-associate (PA) learning takes place in an incremental fashion, Rock (1957) presented evidence favoring an all-or-none conception. This conclusion was based on a comparison of two methods of PA learning. One method was the usual repetition procedure in which the same items appeared on each trial. In the other method, those items missed on one trial were replaced by new items on the next. Rock (1957) reasoned that, if associative strength increases incrementally, then performance should be better under the former condition. No differences were obtained.
Parallel work in experimental psychology, further work following up on initial studies, and further experiments crossing the lines between math educ., experimental, educational and mathematical psychology and beginning of research in structural learning focusing on rule based (structural learning) knowledge rather than what were then contemporary explanations in terms of S-R theories.

-- Syracuse University and SUNY Buffalo

This research addressed generalization and transfer in the context of all-or-none learning. A theory was proposed combining features of strategy election and pattern-conditioning theory. A question is raised whether transfer of training involves recognition when a new item is first presented, or continued opportunities for transfer as long as the new item remains unlearned. This study compares different hypotheses about the initial and transition parameters describing the learning of items during a transfer task. The data are consistent with the hypothesis that transfer occurred on the first presentation of a new item or not at all. The data are adequately fit by a theory expressed as a Markov chain in which learning is an all-or-none event.


-- University of Pennsylvania & Florida State University

Evidence suggests that short-term memory (STM) is a function of the number of units into which S organizes the to-be-recalled message. Scandura (1967) found that STM is also a function of the frequency with which nouns elicit a common descriptive adjective-associate. This study was to determine whether the advance introduction of a conceptual recoding cue (i.e., a common descriptive adjective) improves STM. Whereas the previous study (Scandura, 1967) demonstrated a relationship between adjective identification and retention, this research sought to manipulate STM. In parallel with the above research, I conducted several studies extending my earlier work on problem solving.

This study determined the effects on the STM of five noun lists of: (a) conceptual receding cues (adjectives) presented prior to the lists, (b) list dominance (high, low, and no dom), and (c) retention interval (1, 5, 15, and 30 sec). Number of nouns correctly recalled was a function of recoding cue, dominance level (high > low > no dom), and time. The recoding cues facilitated retention of the high and low dom lists, but depressed 1-sec recall of the no dom lists. There were relatively more within concept-category noun intrusions during high dom and recoding-cue recall than with their controls. Mean latencies between nouns correctly recalled were shorter at 1-sec recall than thereafter, but there were no other differences due to time, recoding cue, or dominance. These results, involving the reorganization of input, were interpreted in terms of a simple model postulating familiar and unfamiliar storage locations.


-- University of Pennsylvania

This research dealt with how learning to solve problems depends on prior learning. Two variables were considered: Practice on subordinate material (P) and Criterion practice (C). Ss were tested on criterion (R), generalization (RG), and non-specific transfer (N) tasks immediately after learning and after five weeks. P affected R and, to a lesser extent, RG performance on immediate tests. There were no significant effects due to C.

In exploring a related problem, I had found evidence suggesting that mere presentation of subordinate material is not always sufficient to insure subsequent learning. Rapidly covering preliminary material, so as to get to the main topic more quickly, may in the long run defeat its
own purpose. Good teachers have long recognized that when material is introduced determines to a large extent how well it is learned. At the time this study was conducted however, there are no published experimental studies bearing directly on the present problem.


University of Pennsylvania

Four groups of 21 Ss were presented with an algorithm (A) and tested on routine (R), generalization (RG), and novel (N) problems. The A Ss had no prior training; the SA Ss were presented with information (level S) deemed necessary for learning the R algorithm; the SA Ss also were presented with information (level C) deemed useful in modifying the R algorithm (so as to solve the RG and N problems); the SA Ss, in addition, had the problems defined and were presented with relationships between an illustrative R problem and its solution (level P). The SA Ss performed better than the A Ss on the R and RG problems (p < .001); there were no other significant increases due to amount of information presented. Only the SA-SCA RG difference resulted in a t greater than 1. On the N problems, only the SA-SCA difference was significant (p < .05), the A-SA difference resulted in a t of 1.44. Successful problem solving did not depend on an understanding of the problem involved. Transfer was attributed to learning syntactic constraints (principles) relating the algorithm and the problem characteristics. The results also demonstrated the feasibility of predicting problem solving performance by subjectively analyzing structural relationships between the criterion and the information presented. Implications were discussed.


University of Pennsylvania

This was the third in a series of experiments dealing with problem solving and prior learning. In each of these experiments, the Ss were presented with two (or more) hierarchically related aspects, prerequisite (P) and criterion (C), of the same abstract card material, and were tested on (training) problems defined at the C level. Both the P and C materials were presented via ordinary discourse. The words, symbols, and/or graphic materials used to denote P notions were also used in presenting the C material. The C material consisted of a definition of the training problems, an illustrative problem, its solution, and pointing out relationships between the problem and its solution. Performance was tested on both the training problems, and new, though related, transfer problems.

In experiment one, I found that practice, with feedback, at the P level facilitated problem solving performance whereas practice in solving the problems themselves did not. These effects were attributed to increasing familiarity with P denotants thereby facilitating C learning. In experiment two exposure to information about the problems, coupled with related practice, reliably improved problem solving performance, via an algorithm, but only when the information was either specifically used to describe the algorithm or clearly provided a basis for modifying the algorithm taught so as to successfully attack transfer problems. Other information about the problems, and particularly definition of the problems, in terms of component elements, did not facilitate problem solving in any case.
The present study was concerned with three variables: 1) amount of relevant prior P knowledge, 2) order of presenting the P and C material, and 3) prerequisite practice. Previous studies showed that there is a strong relationship between P achievement and C learning. Successful performance on C tasks almost never obtained without the prior acquisition of all prerequisites. Nonetheless, in these studies, degree of P learning was not experimentally manipulated; as such, it was not demonstrated to be a causal factor in improving C test performance. Some unknown factor may have concomitantly improved performance on both the P and C tests. In the present study, P learning was an independent, rather than dependent, variable. This study was not, however, simply concerned with comparing the effects of pre-learning versus no pre-learning. The question asked was whether mere exposure to abstract P material is sufficient for successful C test performance. This question is an apparent assumption in many college mathematics classrooms.

The second variable of concern was the order of presenting the P and C material. Experimental Ss were presented first with Task A (with practice), then Task B, and finally were tested on Task B. Control Ss were either given Task A without practice or were not shown Task A. Although it was postulated that the relationship between Tasks A and B may have been a critical boundary condition for the experimental findings, the design did not allow definitive statements in this regard. The P material may have improved problem solving directly and not by mediating C learning. Comparing the two possible orders of presenting the P and C material would settle the issue. If P learning facilitates C learning then the presentation order PC should result in better problem solving performance than the order CP.

The effects on problem solving of P practice were also determined. In accordance with the preceding discussion, P practice was hypothesized to affect problem solving by increasing P learning only in the presentation order PC. In effect, this aspect of the study represented an attempt to both replicate and extend my earlier finding. Replication seemed particularly desirable since other investigators have not found a reliable practice effect. There were, however, some potentially, important differences between these studies. In the Reynolds and Glaser experiment, the materials were not hierarchically ordered. This may be one critical boundary condition. The materials used in the Gagne et al experiment were programed in small steps; the presentation did not involve extended discourse as in my study. The use of P terms to describe C material may be another important boundary condition.

Problem solving performance was determined immediately after the instruction and repeatedly thereafter with feedback, reintroduction of the C material, and hints intervening. This procedure made it possible to determine the effects of the independent variables under different conditions of familiarity with the C material and the problems. Transfer was also assessed.

SUMMARY

Symbols used to denote prerequisite notions were used in defining the problems. Amount of pre-learned prerequisite material, presentation order (prerequisite-criterion or criterion-prerequisite), and prerequisite practice; were manipulated in a 2x2x2 factorial design. Ss were tested on the prerequisite material as well as on training and transfer problems. The major
results were that problem solving performance was improved: 1) by the pre-experimental availability of prerequisite material, even after repeated re-introduction of criterion level materials, hints, and practice in problem solving, 2) when the prerequisite material was presented prior to the criterion material, and 3) by prerequisite practice only when the prerequisite material came first.


This article reports results similar to the above using actual mathematical materials. Among other things, we found that mere exposure to new terminology may provide an insufficient base for describing higher order material. Insufficient familiarity with prerequisite terminology is a real problem in many mathematics classrooms, particularly at the college level, where the material presented during single lessons is typically hierarchical in nature. Instructors should present criterion material in a way that depends only on requisites already available to the Ss.

The next study aimed to communicate with practitioners.


Good teachers have long observed that when information is given is just as important as what information is given. The implications of this observation are of real concern to all educators, particularly teachers of mathematics and science. Nonetheless, most specialists in education would be hard pressed to explain the mechanisms involved. Although there were a variety of peripheral results, four fundamental conclusions may be drawn from this project.

1. Practice in applying prerequisite terminology significantly improves the learning of higher order material, as judged by criterion test performance, when the description of the criterion material involves prerequisite terminology (Experiments One and Four). This conclusion implies that teachers, particularly of mathematics and probably of science, should be careful to make provision for practice with newly introduced prerequisite terms before using these terms to describe higher order materials. This is not always done in high school and college classrooms.

2. The effects of prerequisite practice obtain only when the criterion description involves use of prerequisite terminology (Experiments Two and Four). In effect, there are often alternative means available for achieving a desired objective and these should be considered in those cases where the teacher is primarily interested in promoting acceptable criterion performance with the least expenditure of time. This possibility was demonstrated in Experiment Two by the use of an efficient algorithm and in Experiment Four by the use of alternative criterion descriptions in familiar, as opposed to prerequisite, terms. An incidental implication of this conclusion, rather glaringly evident in Experiment Two, is that it is possible to solve problems and even generalize solution procedures to new problems without even knowing what the problem is. This finding suggests that understanding is an imprecise term. The educationally relevant question is not understanding versus no understanding, but simply what is learned.

3. Learning prerequisite material, via practice or otherwise, does not affect criterion test performance directly, but indirectly by increasing criterion learning. This conclusion may be drawn from the Experiment Three result where the order prerequisite-criterion was superior to the order criterion-prerequisite. The interaction between prerequisite practice and presentation order (Experiment Three) provides further support. This conclusion makes it possible to discard the possibility that prerequisite
learning simply amounts to learning part of the task or that a substantial amount of criterion material can be retained and then correctly interpreted when the prerequisite meanings later become available. Insofar as possible, prerequisites should be taught first and not after the criterion to insure efficient learning.

4. The most far-reaching conclusion to be drawn from this project is that learning ability is far from innate; it depends fundamentally on the amount of prerequisite learning already available. Furthermore, it may be concluded that spending more time on criteria and having an opportunity to practice on related problems, cannot be expected to diminish the advantage of prior prerequisite knowledge. Presumably, the only way this can be done is by going back and teaching the pre-requisites to the lacking students.

The practical implications of this conclusion seem clear. Just because a student from one of our top high schools, for example, does a better job in college calculus than a student from an underprivileged background, is no reason to believe that this same outcome would obtain if the underprivileged student had had the opportunity of equivalent prior training. Since the effects of such pre-training are probably cumulative, we need to devise ways of salvaging those students who would normally fall by the wayside, probably by individualizing instruction. Although developmental activities are needed now, it should be emphasized that we know precious little about the fundamental bases on which individualized instruction rests. Clearly, more basic research on this problem is needed.


The choice suggested in the title, "Teaching—Technology or Theory," in actuality, is non-existent. The act of teaching, being a practical endeavor with practical aims, is clearly not theory. Although there may be some question at present, as to whether teaching is best viewed as an art or as a technology, this controversy also is more imagined than real. Since relatively little is formally known about classroom teaching, it must of necessity be considered an art — subject to the ill-specified whims of the master practitioner (Davis, 1962). Yet, there currently are underway a number of significant attempts to develop teaching as a technology, based on underlying science and theory (Glaser, 1984). If unforeseen improvements are to be made in the instructional process at some point in the future, it must be by this means. Art is ageless; only science and related and technologies are cumulative with time. A predominance of the artistic view may well insure that instruction during the 21st century will be little different from what it is today. Science and technology, on the other hand, may drastically alter present day conceptions.

The purpose of this paper is to contrast teaching as a technology, based on a science of learning, and as a technology, based on a yet-to-be developed theory of teaching. In the process, I will attempt to establish the need for one type of teaching theory, some of its characteristics, and the feasibility of its development independent of learning theory. The paper went on to compare theories ranging for Skinner’s model of operant conditioning to experimental psychology. Task analysis was coming into play during this period. The article concluded that a theory of teaching would of necessity consider teaching as well as learning variables, and that teaching theories may evolve that are largely independent of traditional learning theories.

By way of summary the paper concludes: “Learning theories have been concerned with how learning occurs and say nothing about the categorization and sequencing of information and little about
relationships between past behavior in one situation and present behavior in another. Information about these problem areas would be invaluable to the teachers. It is doubtful that molecular relationships between stimuli and responses would be very helpful. Any complete theory of teaching and learning would of necessity consider teaching as well as learning variables—and that teaching theories may evolve that will be largely independent of traditional learning theories.

**NOTE:** An integrated theory of teaching and learning took a long time to develop, and even longer to realize in technology. Number 18 was a first step toward the Structural Learning Theory (SLT). But read on. First, a couple more basic research studies in mathematical and experimental psychology (19, 20).


-- Florida State University.


The problem of generalization and transfer is discussed relative to all-or-none learning, and a theory is proposed which combines features of the selection-of-strategies theory and the pattern-conditioning theory. A question is raised whether transfer of training involves recognition that must occur when a new item is first presented, or continued opportunities for transfer as long as the new item remains unlearned. This question is seen to involve different hypotheses about the initial and transition parameters of the Markov chain which describes the learning of items during a transfer task. Data are presented from an experiment in which positive transfer of training occurred based on generalization among members of a verbal concept-category. The data are consistent with the hypothesis that transfer occurred on the first presentation of a new item or not at all. The data are adequately fit by a theory expressed as a Markov chain in which learning is an all-or-none event. Statistical tests indicate that the effect of transfer can be described as a change in the initial vector of the chain, but that the learning parameters in the transition matrix varied little or not at all among conditions which differed widely in amounts of transfer. Finally, the amounts of transfer in 12 conditions are analyzed in relation to assumptions about the acquisition of verbal concepts and the recognition of new instances after a concept has been acquired.


The recent revolution in mathematics education has raised many fundamental questions and given rebirth to others. What are the objectives of mathematics education—and, how can they be used as a basis for curriculum development? What is really learned when mathematics is discovered? What does the master teacher do which makes him a master teacher—can others be taught to do these things? How does one measure mathematical knowledge and ability? To what extent can mathematical creativity be taught? These are just some of the questions about which we have only partial answers and about which more complete answers must be attained if mathematical education is to reach its potential ...


THERE are few examples of commutative non-associative systems that are readily comprehended and meaningful to junior high school students. Some examples of such systems are outlined in this short article.


As a result of a decade of “action research” characterizing the so-called “New Math” many mathematics teachers were unclear as to the nature of scientific research. This article aimed to inform mathematics teachers about the interrelationships between: (1) basic research, (2) product-oriented research, and (3) development.


A RESULT of the past decade of "action research" and curriculum innovation there is a confusion in the minds of many mathematics teachers as to just what is and is not scientific research. To be an intelligent consumer of research, the mathematics teacher must become more aware of the interrelationships between: (1) basic research, (2) product-oriented research, and (3) development.


The following excerpts are based on a late draft as the original publication has been lost:

In reacting to Professor Suppes’ paper, I find myself in the position of a sophomore wrestler in the company of a champion. It would be presumptuous of me to seriously attempt a detailed critique of his excellent and scholarly paper and, yet, I feel that I would not be doing what is expected of me if I did not attempt something along these lines. My position is made still more difficult because I am in almost complete agreement with what Suppes is trying to do. The importance of the problems he has identified and the sophistication of the approach he is taking surely cannot be denied. Even more relevant, his implicit plea for more basic research on mathematics learning could hardly fall on more ready ears. My only hope is that his advice is followed by increased and more diverse support for such research from the various funding agencies.

Note: As above, I outlined my response in rather lengthy draft of my paper is available in Dropbox upon request. The essence of my proposal was an alternative to the detailed, narrowly scoped studies reported by Suppes. Suppes research strategy was one of developing precise mathematical models for specific mathematical tasks (e.g., the time it takes to perform addition by counting from versus counting up from the first number). The hope was that miniature theories of this type would ultimately lead to grander and more encompassing theories. After describing limitations of the linear regression methods used in his research, I proposed an alternative of focusing on more complex behavior.

My justification was that, in many of the most interesting kinds of mathematical learning and behavior, time is of secondary concern. It seemed to me undesirable at this stage of the research to unnecessarily complicate an already complicated area. In mathematical problem-solving, for example, we are typically more interested in whether the learner can solve a problem than in how long it takes. Similarly, mathematics teachers rarely present (non-drill type) information repeatedly for short periods of time until it is learned as is done in most experiments. For the most part, the learner is allowed all the time he needs.


The advent of modern mathematics, with its emphasis on the more abstract aspects of the subject, has forced the mathematics teacher to re-evaluate his methods of presentation. In teaching abstract mathematical material, one important variable is how the new material is introduced. Frequently, in the mathematics textbook or lesson, abstract material is introduced with a few paragraphs of history about the subject or its inventor, presumably to increase motivation. A well known algebra textbook, for example, introduces a section on group theory by relating the fact that Evariste Galois, an early mathematician instrumental in the invention of group theory, was killed in a duel in 1832. Another way of introducing such material is with concrete examples, models (e.g., Scandura, 1966c) or embodiments (e.g., Dienes, 1964) of the abstraction to be learned. The use of such illustrative materials is commonplace in mathematics instruction, but at the college level its introduction typically follows, rather than precedes, more formal presentation in terms of definitions and theorems.

In studying a related problem, Ausubel (1960) found that learning meaningful verbal material could be enhanced by using, what he termed, an advance organizer. Organizers were defined as introductory material at a high level of abstraction, generality, and inclusiveness. In this (Ausubel, 1960) study, the organizer provided an overview, in familiar terms, of a topic on metallurgy. The criterion material consisted of more detailed information about the topic presented in more technical language. By advance organizer, then, what Ausubel seems to be referring to is a general non-technical overview or outline in which the nonessentials of the to-be-learned material are ignored. Such introductory material has typically proved easier to learn than the more technical material that follows.

Presenting mathematical abstractions, on the other hand, normally involves the use of words or symbols having no referential meaning for the naive student (Scandura, 1964, 1966a, 1966b, in press). For this reason, descriptions of concrete models of abstract mathematical ideas, although the models themselves involve extraneous features, may be more readily interpretable than formal presentations of the corresponding abstract ideas in terms of the underlying definitions and axioms.

In this study, historical and model introductions to formally described abstract mathematical content were compared for their effects on learning efficiency with pre-service elementary school teachers. It was hypothesized that learning would be enhanced more by the model introductions which would, in effect, serve as advance organizers.


Two experiments were conducted. In experiment one, 51 college Ss were taught one of three rules, of varying generality, for winning the game of NIM. Two additional groups of 17 Ss each served as controls. In experiment two, the variables were rule generality (3 levels) and example (given-'not given). The materials, based on arithmetic series, were presented to 114 junior high school Ss. All ^s were tested on three problems, the first within the scope of each rule, the second within the scope of the tv70 more general rules, and the third only within the scope of the most general rule.

The results generally justify the categorization of verbally presented rules as to generality. There was positive transfer to an outside scope problem in only one case and each group's performance was at essentially the same level on the within scope problems. In experiment one, the most specific rule was better learned than the others; a similar, but weaker, effect was noted in experiment two.

A third facet of the study dealt with response consistency. Except for one case where the effect was rather directly attributable to prior learning, those Ss who used the rule taught on one problem tended also to use it on succeeding problems whether or not the rule was appropriate. Both practical implications for testing and theoretical questions were discussed.

Implications and Theoretical Comment: The results of these experiments demonstrate, in a rather conclusive fashion, the behavioral relevance of principle generality. For the most part, successful performance was noted only on tasks...
within the scope of verbally stated principles. When principles are presented in an expository fashion, it is normally too much to expect generalization to problems to which the principle does not immediately apply.

Of perhaps even greater practical significance were the lack (there was one exception) of performance differences on within scope problems and the consistency results.

It should be emphasized that Rule S, in experiment two, was conceptually different from the others used in experiments one and two. Rule S applied to only one stimulus (series) whereas each of the others applied to a set of stimuli. The fundamental nature of this difference has been discussed in detail elsewhere (Scandura, 1966a, 1966d). The former result demonstrates that (almost) any stimulus within the scope of a principle is equally as difficult to respond to correctly as any other. Furthermore, coupled with the consistency data cited in the introduction, that obtained consistency results suggest that only one (new) test stimulus is needed to determine whether, in fact, a stated principle has been learned (i.e., correctly interpreted). More information is gained by using additional test instances. These results could have far-reaching implications for the development of highly efficient measuring instruments.

In addition, the pronounced tendency of the Ss to attack all of the test problems in the same way, irrespective of whether the procedure used was appropriate, suggests that the ability (i.e., knowing how) to solve problems and knowing when to solve them are quite distinct. Testing for the latter ability necessarily must involve the presentation of extra-scope problems. More important than the results of these exploratory experiments were the post hoc analyses they made both necessary and possible. In particular, the preceding discussion strongly suggests that the roles played by various aspects of a principle statement need to be more clearly specified. The form, "If A, then B" does not detail all that appears relevant. For one thing, it was not possible in this study to distinguish between the roles played by A, L, and N (the stimulus variables entering into the rule (A + L)/2\[H and the algebraic expression, [(X + Y)/2\]Z (the form of the combining operation of the rule by which the appropriate response sums are determined). The variables relate to properties of arithmetic series stimuli, while the algebraic expression represents a ternary operation by which another property (e.g., siams) may be derived. A, of course, while it played no role in this study is also critical. It tells when a rule can and cannot be applied.

Thus, the rule, is appropriate whenever an arithmetic series consists of the odd integers beginning with 1 while [(A + L)/2\]w works whenever there is a common difference between adjacent terms. These observations suggest that a principle statement may be represented more appropriately by the form, "If I', then O' (D') = R'," where I' refers to the set of stimulus properties which indicate when the rule, denoted O' (D'), should be applied, D' refers to the set of those properties which determine the responses, and O', to the operation from which the responses, denoted by R', may be derived from the properties referred to by D'.^ That part of a principle statement represented by 0* (D') corresponds to what is typically called a rule.

Similarly, a principle, that internalized representation which determines a learner’s responses to stimuli, may be characterized by an ordered four-tuple (I, D, O, R). Primes, of course, have been used to distinguish between the referents (e.g., O) and the symbols used to represent them (e.g., O', O", etc.). These definitions, along with that given in the introduction, form the basic elements of the Set-Function Language (Scandura, 1966a, 1966d)

Although the actual symbols used in a statement may be an important factor, as suggested above, the hypothesis advanced in this study to the effect that rule generality and interpretability are inversely related finds a formal rationale in the nature of the characterizing elements. Making operational use, for example, of the arithmetic series property (i.e., dimension), "the difference between adjacent terms is some common value," necessarily presumes that, "the difference between adjacent terms is "etc.," can all be correctly interpreted. The converse does not necessarily follow. A similar relationship exists with respect to the rules, 50 x 50 and N x N. To correctly apply the latter, more general, rule to any particular
series requires the ability to determine any value of the dimension M, including 50. Being able to apply
50 X 50 does not.
It would appear that the more general the principle the more is expected of the learner. Whether such
differences v/ill be reflected in behavior, however, may depend on not only rule generality but the
population involved, particularly on whether the Ss have the necessary requisite abilities (Gagne, 1962;
Scandura, 1966b).
In effect, differences in generality appear, on analysis, to be equivalent to a collection of sets only one of
which has the latter property. For the same reason., the property represented by the placeholder X is
more abstract than the number since it refers to a still higher order collection. Unfortunately, we have
not yet conducted a study designed to provide definitive information on these points.
For the present, this analysis remains hypothetical.
COMMENT 30. My initial challenge to the status quo in experimental psychology – introducing “rules” as the basic
unit of knowledge versus “associations” as commonly assumed at the time in experimental (often called S-R)
psychology.

(Also in The School Review, 1967, 75, 329-341.) University of Pennsylvania
To provide a substantive base for their research, educational psychologists have frequently resorted to
the languages, paradigms, and theories of the mother science of psychology. Mediational elaborations
and operant conditioning paradigms of the S-R language and more general, but less well specified,
cognitive theories have been popular. Each approach has important limitations. From one point of view,
parsimony suggests that the properties of overt S-R associations should also be attributed to
mediational links. Yet, practice has shown that mediational interpretations become increasingly
cumbersome and less precise as situations become more complex. Similar difficulties have plagued
researchers who have used operant techniques to study meaningful verbal learning. A general limitation
of cognitive theories is their relative imprecision. Typically, "cognitions" are either not clearly specified
in observable terms or are only partially defined.
In short, the choice to date has been between a precise, but seemingly inappropriate S-R language, and
presumably more relevant cognitive formulations which leave much to be desired insofar as scientific
cohesiveness and rigor are concerned. The purpose of this paper is to introduce what I feel are the basic
ingredients for a new scientific language for formulating research questions on meaningful learning. This
so-called Set-Function Language (SFL) is precise and seems particularly well suited for dealing with
mathematics, my own area of concern, and science, but it undoubtedly can be used with other subject
matters as well. Rather than try to detail the SFL or to summarize the related research that we have
completed or have under way, let me simply try to convey the general idea. In the process, SFL and S-R
formulations of several meaningful learning tasks will be contrasted. The SFL is behavioristic, as is the S-
R Language, but, unlike the S-R Language the SFL denies the primacy of the S-R association.
To illustrate some of the advantages of the SFL, consider an experimental situation where 5 is required
to respond appropriately to learning stimuli (objects) which are large or small, black or white, circles or
triangles. Suppose further that S is required to learn the particular S-R pairs: (a) A large black triangle >
"black" (Si>Ri); (b) a large white circle > "large" (S2 > R2); (c) a small white triangle > "white" (S3 > R3); (j)
a small black circle > "small" (S4 > R4). After the four S-R pairs are mastered so that 5 can reliably give
the correct response to each stimulus, the question still remains as to just what was learned. Did the
subject learn four distinct pairs—four discrete associations—and notice no relationships between them?
Or, did he learn the two principles, "If triangle, then color," and "If circle, then size"?
This question first began to bother me during the summer of 1963. Greeno and I found, in a verbal
concept learning situation, that essentially an S either gives the correct response the first time he sees a
transfer stimulus or the transfer item is learned as its control. The thought later occurred to me that if
transfer obtains on the first trial, if at all, then responses to additional transfer items should be contingent on the response given to the first transfer stimulus. In effect, a first transfer stimulus could serve as a test to determine what had been learned during the original learning, and to make it possible to predict what response a subject would give to a second transfer stimulus. To obtain evidence on this point, I had 15 pre- and postdoctoral Ss overlearn the four S-R pairs listed above. 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 transfer stimuli. After learning, two Test 1 stimuli (a small black triangle and a large black circle) were presented and the Ss were instructed to respond on the basis of what they had just learned. All answers were reinforced as correct. The Test 2 stimulus (a large white triangle and a small white circle) were presented in the same manner.

The results were clear-cut. All but three Ss gave the correct responses to Test 1 and Test 2 stimuli. It would appear that when a S thinks he is right and the new situation remains relevant, he will continue to respond in a similar manner.

On what basis could this happen? It was surely not a simple case of stimulus generalization; the responses were distinct and, in any case, did not depend solely on common stimulus properties. The first Test 1 stimulus (a small, black triangle), for example, is as much like the fourth learning stimulus (a small, black circle) as the first (a large, black triangle)—and, yet, "black" was invariably given as the response rather than "small." Perhaps the simplest interpretation of the obtained results is that most of the subjects discovered the two underlying principles while learning the original list and later applied them to the test stimuli. In effect, the relationships between the S-R pairs, themselves, Combined With A Response Consistency Hypothesis, Provided A Basis For Assessing What Was Learned.

... Comment 31, 32 And 33. Following are a few more studies bridging the gap between experimental psychology and rule learning.


24 college Ss learned 4 rule statements and were then tested on 2 problems to see if they could apply the rules involved. A 2 X 2 factorial design, with repeated measures, was used. 1 factor was the form in which the rules were stated, either succinctly worded English or initially unfamiliar mathematical symbolism. The other factor involved the presence or absence of symbol pre-training on the meaning of the constituent symbols. The results were: (a) symbolic statements were applied successfully if and only if the Ss were taught the symbol meanings (and the underlying grammar); (b) symbolic statements were learned more rapidly with or without symbol pre-training; (c) English statements were applied equally as well as those symbolic statements which followed symbol pre-training; (d) rate of learning the English statements was unaffected by symbol pre-training; (e) success on 1 application problem implied success on the other.


Learning rate and transfer to new stimuli requiring new responses, were found to vary directly with the number of instances (1, 2, 3, 6) of principles in a 12-pair list of paired associates. Also, a positive relationship was found between learning rate and transfer within the 6-instance condition.


Four groups of 16 Ss were presented with four kinds of list (one, high, low, and no dom) which differed as to the frequency with which the five nouns in a given list elicited a common descriptive adjective. There were two retention intervals (5 and 30 sec). The adjective-recall (AR) group was instructed to state the common adjective immediately after each list was presented. The recall-adjective (RA) group was similarly instructed to do so after attempted recall. The categorization groups (CR and RC) were
instructed to classify new nouns. Number of nouns recalled was a function of time and dom level, one > high > low > no dom. The One vs. High and High vs. Low Dom by Time interactions were significant. Retention was poorest in group CR. Adjective identification and retention were correlated at the high and low dom levels after 30 sec. The lack of adjective forgetting suggested that whereas content may be forgotten over time, the storage location itself may not. Mediation and search model interpretations were offered.


35. Scandura, J.M. Pro...Con Ebel. PHI DELTA KAPPAN. P. 255. University of Pennsylvania. Comment: Again, challenging the status quo, this time challenging the reigning AERA president. (NOTE; I was not shy as a young upstart, perhaps a side effect of my competitive background in wrestling. 😊 ) The first time I glanced at Robert Ebel's "Some Limitations of Basic Research in Education" [October KAPPAN] I passed it off as another editorial mistake; the author was probably just trying to court the establishment by expressing a point of view, presently popular with the governmental funding agencies. For some reason I went through the article again—and this time I got mad. Ebel has set up and attacked a straw man and should not be allowed to go unchallenged. . . .

I particularly disagree with three reasons Ebel selects to show why "basic research in education can promise very little improvement in the process of education." These reasons* are precisely those which are most appropriate to the very sort of research he is proposing—the collection of raw empirical or _survey data of the type that has characterized much of educational research for the past 30 years. 1) Its record of past performance is very poor. 2) . . . explanations of that poor performance call attention to serious basic difficulties that are unlikely to be overcome in the foreseeable future. 3) . . . education is not a natural phenomenon of the kind that sometimes rewarded scientific study....

, . . I propose, for example, that one of the major reasons why the expected advantages of individualized instruction and related technologies in education have failed to be as dramatic as systems enthusiasts would have us believe is precisely because they lack the basic information necessary to devise such systems. The fact of the matter is that we have, until very recently, had very little research in _education_ that would classify as basic. One indication of this is the increasingly acceptable conclusion that educational psychology is _not_ applied psychology but is an area of study greatly in need of its own constructs and conceptualizations.

Rather than further increase the current emphasis on development, I would suggest that what is presently most needed is renewed vigor in and support for basic research in education aimed at clarifying some of the problems involved in such important areas as individualized instruction. Precious little is known about crucial problems like 1) how to conceptualize relationships between apparently discrete but nonetheless related subject matter competencies, 2) how to assess what a student knows at each stage of learning, and 3) how "what is learned" and what is presented interact to produce new knowledge. There seems to me to be a strong parallel between recent failures in devising truly individualized learning and the inability of applied physicists to harness the H-bomb on the basis of the presently very incomplete knowledge about the particles and forces comprising atomic nuclei. The failure of development and engineering to achieve its set objectives is usually good reason to return to the laboratory to find out what is going on. In further support of this argument, I would simply call attention to the fact that the really difficult problems in computer-assisted instruction are not due to inadequate hardware but to grossly incomplete information about the instructional process.


The purpose of this paper is not to review all or even much of the past or current research in science education; this has been done elsewhere. ...
We would like here to put such research into a perspective and to indicate some of the strengths and weaknesses of various approaches to educational research generally. Some of the examples given will pertain specifically to our area of mathematics education, but their general relevance can also be shown in science education.


Comment 21-26: Announcing a new doctoral program in the American Mathematical Monthly after my move to Penn and another mathematical observation. Similarities between MATH AND SCIENCE EDUCATION moved me increasingly toward the study of what the mathematician Z.P. Dienes called Structural Learning. In the same time frame, and by analogy to psycho-linguistics, I introduced the name psycho-mathematics to help identify what I was doing – an integration of psychology and mathematics.

Ironically, Dienes adopted the name psycho-mathematics for an institute he created in Sherbrook, Canada. I came to prefer “structural learning”, and adopted the name in introducing my Structural Learning Theory in Reference 55. This was name switch we both endorsed.

During this period, I was still struggling to find a way to formalize my original approach theorizing about teaching and learning. This involved going back and forth between psychology, mathematical formulations and subject matter education, critiquing while trying to find the theoretical key to bringing it all together. Among other things during this period, I had the opportunity to react to a major address to NCTM by Pat Suppes (a luminary at the time). Suppes was also working in mathematical psychology at the time, and was applying those technologies in mathematics education. This work focused on stochastic models to model the behavior of students as they solved simple addition problems – e.g., to add 4+2, the student count up from 0 or from the first number and increment by one (either 6 or 2 (from 4) times). In short, Suppes was creating precise mathematical models to explain simple arithmetic in the hope that the approach would scale to more complex mathematics. My goal was the opposite, working from the top-down. It was this top-down approach that finally led to Structural Analysis in the SLT. Other mathematics educators also felt simple stochastic models would be too limiting, but I was the only one to express his concerns public forum – perhaps because I was the only one specifically asked to prepare a reaction. In fact, I year or so later, I was invited to spend most of a year in Suppes Institute at Stanford. This was the year I did most of the work on my book Structural Learning 1: theory and research. I believe my paper 55 had come out previously, but let me not get ahead of myself.

COMMENT: In this paper, I extended my discussion of the Set-Function Language to enable describing complex (focused on mathematics) learning.

DOING basic research on mathematics learning is a risky business. In the first place, it is hard to know whether one is asking the right questions, and, in the second place, it has been difficult to formulate significant questions involving mathematics learning in a researchable form. One of the major reasons for this state of affairs has been the lack of any suitable theoretical superstructure from which to work. The purpose of this paper is to outline some of the theoretical work under way at the University of Pennsylvania and to show how this work has helped to improve our understanding of how people learn mathematics.


NOTE: this paper was a final blow to the assumption that learning by discovery is somehow better being told. When one discovers something on one’s own one may ALSO learn how to discover in similar situations. This may be useful.
However, it also is useful (more so in some situations) to learn how to learn from verbal (or other) forms of instruction. The point is that any given skill or bit of knowledge may be learned either by discovery or by expository. Again, the question is what is to be learned. The method of instruction is a separable matter. Following is an abstract of the research.

2 questions were asked: (a) can "what is learned" in mathematical discovery be identified and taught by exposition with equivalent results, (b) how does "what is learned" depend on prior learning and on the nature of discovery? The major hypothesis was that discovery Ss may discover derivation rules for deriving classes of solutions but only when the solutions are not initially known. 4 programs, (specific) rule-given (R), discovery (D), guided discovery (G), and exposition of derivation rule (E) were administered to 7 groups. 1 group received program R alone; the others received R with 1 of the other programs. Both orders of presentation were represented: RD, DR; RG, GR; RE, ER. All Ss were required to derive new solutions within the scope of the derivation rule. As hypothesized, Groups R and RD performed at 1 level which was reliably (p < .001) below the common level of the other 5 groups.

Theoretical and practical implications were discussed.


Experiments in which two or more teaching methods are compared as to effectiveness have been notorious for their inconclusiveness. In one study, method A might result in superior learning whereas, in another study, method B might prove to be best. More frequently, there are no significant differences at all.

In conducting a series of exploratory studies in the classroom, each comparing one expository method with one discovery method, Scandura (1964b) was able to identify one of the major reasons why such comparisons have not been definitive. Relatively minor within-method differences seemed to have a greater effect on the experimental outcomes than differences in the methods (i.e., expository and discovery) themselves. Nonetheless, the situation did not appear hopeless. In each of the experiments conducted, the major determining factor appeared to be the timing of the information given. Good teachers, of course, have long observed that when information is given is just as important as what information is given. Nevertheless, most teachers and even college teachers of teachers, would be hard pressed to explain the mechanism involved.

A better understanding of the underlying causes and effects of timing might well help to improve existing methods courses for teachers of mathematics and other hierarchical subject matters and, in turn, teaching in the schools. Still, the number of controlled studies which have dealt with this problem is small indeed. Those which involve reading technical materials have been almost non-existent; yet, this is precisely the way in which the learning of most hierarchical subject matter takes place.²

Although timing may be crucial, it is not always possible to tap and make effective use of learner feedback when teaching by expository methods or in preparing expository material. Not only must the prerequisites to learning be identified, but techniques are needed to help insure the attainment of these prerequisite abilities before higher order information is presented. In effect, timing must be characterized in terms of variables under the direct control of the teacher or writer.

The general purpose of this research was to identify factors affecting the learning of hierarchically arranged expository material and to help determine the relative strengths of these factors. Familiarity with the ideas denoted by the signs (i.e., symbols and icons) used to describe higher order material was given the major emphasis. Towards these ends, four experiments were conducted. All of the instruction took place via carefully prepared text-like passages and practice exercises.

... (we skip details on the methods and results)
CONCLUSIONS AND IMPLICATIONS: Four major conclusions may be drawn from this research. Summaries in italics.

1. Practice on prerequisite tasks significantly improves the learning of higher order material, as judged by criterion test performance, when the prerequisite terminology is used to describe this higher order material (Experiments One and Four). While this conclusion is hardly earth-shattering, one has only to witness what goes on in many classrooms to see that many teachers fail to take this fact into account. Fortunately, elementary school teachers are typically far more inclined to provide ample practice at prerequisite levels than their counterparts at the high school and college levels.

2. Knowledge had by a learner affects future learning ONLY when this knowledge is prerequisite to the to-be-learned material (Experiments Two and Four). In Experiment Two, for example, the learners were able to solve problems and even generalize the algorithmic solution procedure to new problems without even knowing what the problem was. This particular finding suggests that understanding is an imprecise term. The educationally relevant question is not understanding versus no understanding, but simply what is learned (see Footnote 4).

3. The possibility that prerequisite learning simply amounts to learning part of the criterion task or that a substantial amount of criterion material can be retained and then correctly interpreted when the prerequisite meanings later become available is untenable (Experiment Three). This conclusion may be drawn from the Experiment Three result in which the order prerequisite-criterion was superior to the order criterion-prerequisite. The interaction between prerequisite practice and presentation order provides further support. On this basis, it would appear to be poor practice to present higher order concepts first and then shore up deficiencies; rather, the reverse order of presentation seems indicated.

4. Simple exposure to prerequisite information is often not sufficient to insure later learning. Furthermore, spending more time on higher order material and having an opportunity to practice on related criterion problems, cannot normally be expected to overcome prerequisite inadequacies (Experiments One and Three). Since the effects of inadequate prior knowledge are undoubtedly cumulative and since (by conclusion 3) it is inefficient to make up prerequisite deficiencies after students have wasted time trying to learn higher order concepts, the task of salvaging students who would normally fall by the wayside is made doubly difficult. It would appear that there is little room for error. It is urgent that we learn more about both the proper sequencing of materials and how to determine when students have the necessary prerequisites. Clearly, more basic research on these problems is needed.


NOTE: I include this article in full, not because it is my best theoretical article, but because it is the first to detail an alternative to then prevailing wisdom in S-R theory that dominated thinking and research in experimental psychology at the time, and by deference by educational psychology as well. As will become increasingly apparent, what I called rules replaced S-R associations as a better way understand school learning.

Theoretical development in educational psychology has been extremely slow. One major reason has been the typically imprecise definition of independent and dependent variables in research on meaningful learning and teaching. Such research can bear only an ambiguous relationship to theory. Similarities and essential differences often go undetected. Stating research objectives and defining variables in unambiguous terms, however, is not sufficient. The teaching-learning process has all too frequently been studied in terms of such "administrative" variables as class size, grade level, IQ, and amount of teaching experience. To be theoretically relevant, the variables chosen must have broad
This retrospect is a work in progress – pp. 1-10 draft plus List with Notes

explanatory potential. They should not merely be symptomatic of and inextricably related to the question at hand. Theory development depends on much more than mere fact finding.

To provide a substantive base for their research, educational psychologists have frequently resorted to the languages, paradigms, and theories of the mother science of psychology. Mediational elaborations and operant conditioning paradigms of the S-R language and more general, but less well specified, cognitive theories have been popular.

Each approach has important limitations (Scandura, 1966a; 1967a, b). From one point of view, parsimony suggests that the properties of overt S-R associations should also be attributed to mediational links. Yet, practice has shown that mediational interpretations become increasingly cumbersome and less precise as situations become more complex. Similar difficulties have plagued researchers who have used operant techniques to study meaningful verbal learning. The results simply are nowhere near as clear in complex human learning as they are in the "Skinner Box." It is increasingly recognized, for example, that knowledge of results is not directly analogous to feeding a pigeon and that, in any case, other factors, such as subject matter structure, are probably of greater importance in promoting efficient learning. A general limitation of cognitive theories is their relative imprecision. Typically, "cognitions" are either not clearly specified in observable terms or are only partially defined. Under these conditions it has been impossible to construct a predictive theory—the sort of theory needed if practical implications are to be obtained.

In short, the choice to date has been between a precise, but seemingly inappropriate S-R language, and presumably more relevant cognitive formulations which leave much to be desired insofar as scientific cohesiveness and rigor are concerned.

In an attempt to overcome these difficulties, I have proposed a new scientific language for formulating research questions on meaningful learning. Because the language is framed in terms of the mathematical notions of sets and functions, the name Set-Function Language (SFL) was adopted.²

²See Scandura, 1966a; 1967a,b. The descriptor "Set-Function" should not be confused with set functions.

The studies reviewed and data cited in this paper are concerned with mathematics learning, but the general procedures and results are also applicable to other subject matter areas. This paper is based on a series of papers presented at the A. E. R. A. meeting in New York City on February 18, 1967. The author would like to thank Joan Bracker for her assistance in the preparation of this article. Her participation was made possible by a Research Training Grant from the U. S. O. E.

I won't go into detail here, but before outlining the SFL it may be instructive to briefly consider the S-R mediation language. In S-R psychology, the basic building block is the association, a construct which was abstracted directly from observed connections between overt stimuli and overt responses.

Learning a concept, presumably a more complex form of learning, involves the ability to give a common response to any one of a set of stimuli. To say that a subject has acquired the concept of "red," for example, implies that he is able to give some common response, when shown any red stimulus object, but will know not to give this response to any non-red stimulus. Similarly, a child may be said to have acquired the concept of "four" if he can say "four" when presented with any conglomeration of four objects but will not say four to any conglomeration not containing four objects—i.e., assuming, of course, that the child is operating under the same set of instructions in each case. In short, whereas an association pairs one stimulus with one response, a concept is a many-to-one relationship.³

S-R theorists have felt obliged to represent the many-to-one concept relationship as a composite of one-to-one associations,
Notice that the stimuli $S_1$, $S_2$, and $S_3$ are connected to the mediating response $M$ whose stimulus properties, in turn, elicit the observable response $R$.

**The Basic Unit in Meaningful Learning.** Most subject matter learning involves neither associations nor concepts but, as they have been variously called by different investigators, rules, principles, schemas, heuristics, and TOTE units. This is true even more so of mathematics learning. To be more specific, meaningful learning implies the ability to give the appropriate response in a class of responses to any stimulus in a class of stimuli. Unfortunately, this fact has often been overlooked because the term "concept" has been used so widely in discussing subject matters. When we say that a child "has the concept of addition," for example, what we probably mean is that he can give the appropriate sum when presented with any pair of numbers. Put another way, the learning involved connects a large class of stimuli with a large class of responses. By definition, a concept connects a class of stimuli to exactly one response.

To see what is involved in meaningful learning, learn the following S-R pairs (i.e., overt inputs and outputs): $(4 \ 3 \ 1) \rightarrow 3$, $(8 \ 1 \ 6) \rightarrow 2$, $(7 \ 9 \ 2) \rightarrow 5$, and $(9 \ 5 \ 1) \rightarrow 8$. Now, on the basis of what you have just learned, give what you think should be the responses to the stimuli $(7 \ 2 \ 1)$ and $(4 \ 7 \ 2)$. Did you give the responses 6 and 2? If so, you probably acquired a unit of knowledge (i.e., rule) which might be stated, "Subtract the number in the third position from that in the first." If you did not give these responses, you presumably learned the pairs on the left as discrete entities—i.e., as distinct associations without noticing any relationships between them. (Of course, I didn’t indicate that there was any such relationship so you may not have been looking for one.) In this case, the rule governed responses, 6 and 2, would be expected on the basis of chance alone.

I should like to point out that in talking about single stimuli (responses), I am actually referring to equivalence classes of overt stimuli (responses) in which the members (i.e., the stimuli) of the classes are either indistinguishable or otherwise play exactly the same role. For example, the stimuli "5," "five," "Five," "fIVE" are all equivalent insofar as the number five (as opposed to the numeral "five") is concerned.

I should like to emphasize that this situation was not picked arbitrarily simply to embarrass S-R psychologists. Whereas a "patch job" can be done with certain special cases, I believe that I am safe in saying that to date no satisfactory way of representing rules *et al.* solely in terms of associations has been found (See Scandura 1966a; 1967a, b). Stimulus dimensions which uniquely determine the responses (e.g., the first and third positions) and the combining operation or transformation by which the responses are derived from the determining stimulus properties (e.g., subtraction) appear to be crucial aspects of all rules. While stimulus properties and derived responses play a central role in S-R mediation theory, there is no counterpart for the transform or combining operation.

**The Set-Function Language (SFL)**
Fortunately, all of these characteristics play a central role in the SFL. In fact, I (Scandura, 1966a; 1967a, b) have proposed that four characteristics are needed to specify a principle. Three of these characteristics specify a rule and a fourth determines when the rule is to be applied. Those stimulus properties which determine (D) the corresponding responses constitute one such characteristic, the covert responses or derived stimulus properties (R) are another, and the transform or combining operation (O) by which these covert responses are derived from the determining properties is the third. The fourth consists of those, usually higher order, stimulus properties which identify (I) the rule to be applied. For example, the rule, \( N^2 \), for summing number series, where \( N \) is the number of terms, applies only to those series which consist of the consecutive odd integers beginning with one (e.g., \( 1+3+5+7 = 4^2 = 16 \)).

Since principles, as well as associations and concepts, can be represented in the S-R mediation language, it is appropriate to ask whether these notions can also be represented in the SFL, in which rules and principles are taken to be basic. This is indeed the case. Furthermore, unlike the S-R language, explicit distinctions are made between: (a) the observable S-R instances of a rule—the denotation, (b) the rule or principle itself, that which underlies the behavior and whose presence can be inferred only indirectly, and (c) statements of the rule or principle in symbolic form.

The denotation is simply a junction, a (mathematical) notion which may be defined as a set of ordered stimulus-response pairs such that to each stimulus there is one corresponding response. The denotation of a concept is simply represented as a constant function in which there is one response common to all stimuli. To represent an ordinary association, the set is further restricted so as to include only one S-R pair.

The rule construct is characterized as an ordered triple (D, O, R), where D, O, and R are as defined above. Principles, of course, are ordered four-tuples (I, D, O, R). In the case of concepts and associations, there are certain relationships among these characteristics (Scandura, 1967b) but they need not concern us here. In stating rules or principles, I have used primes to distinguish the signs, used to represent the constructs I, D, O, and R, from the constructs themselves.

The one point I would like to emphasize is that even S-R theorists have been forced to adopt the idea of a transformation or combining operation in order to represent meaningful learning and thinking. Berlyne (1965) has done this, for example, in his fine book *Structure and Direction in Thinking*. In his latest article on the topic of principle learning, Gagne (1966) has adopted this point of view as well. Even many long-time mediation enthusiasts have recently become convinced that mediation theory is inappropriate for dealing with verbal learning.

In so doing, these theorists are in fact denying the primacy of the association as a construct. The idea of a transformation or combining operation takes over this primary role. I might note parenthetically that what here have been called response determining properties of stimuli, S-R theorists have called mediating responses. Similarly, our covert responses (or, derived stimulus properties) correspond to stimulus properties of mediating responses. What the combining operation does is to make explicit that the mediating stimuli are determined from the antecedent mediating responses. While this process may not be important in many forms of simple learning, we have just seen here how it becomes crucial in rule learning. The transformation, mapping, combining operation, or whatever term is used, becomes of central concern, and to still call any suitable representation associationistic or even neo-associationistic is stretching the term beyond its reasonable limits.
In the final analysis, of course, the choice between scientific languages (and theories) involves efficiency and cohesiveness as well as the sheer ability to represent or account for observable phenomena. It is in this sense, that the much heralded adaptive quality of the S-R language too often has led psychologists to overlook the fact that it is \textit{always} possible to "patch-up" an existing formulation to meet new situations. Parsimony does not simply refer to the maintenance of existing concepts but to the formulation of emerging structures in the simplest possible way. I might add that what is presently being done with the S-R language (e.g., the inclusion of associative structures, reference mechanisms, etc.) is quite analogous to what Copernican astronomers were doing when they invented epicycles to represent planetary motions in an attempt to salvage geocentric theory.

**EMPIRICAL RESEARCH BASED ON THE SFL** Let me turn now to a consideration of two problems that have long plagued mathematics educators, the problems of \textit{rule-generality} and \textit{discovery learning}. In the space remaining, I shall attempt at least partial resolutions of these two problems and, in the process, will indicate how the SFL helped to achieve these resolutions by making it possible to formulate the underlying questions in a precise way. Ernest Woodward and Frank Lee assisted me with the Rule-Generality Study and the Discovery Learning Study was conducted with Dr. William Roughead.

\textit{Rule Generality} (Based on Scandura, Woodward, & Lee, 1967)

In instructional situations, the question often arises as to how general the presentation of material ought to be. Subject matter specialists and most educators tend to emphasize that the more general the presentation, the more useful it will be.

Learning oriented experimental psychologists, on the other hand, are often inclined to point out that the more specific the presentation, the better the learning.

The question of generality arises particularly often in mathematics instruction. Should addition and subtraction be taught as two distinct, although related, operations, as has been done traditionally, or as one operation as is done in more modern treatments? Should the three cases of percentage be taught separately or as variants of the rule, "base X rate = percentage?" Should pupils be taught the method of "casting out nines" or be taught the more general principles of modular arithmetic? How generally should theorems be stated? Proofs? The answer to these and related questions hinges, in part, on the learnability and utility or scope of the rules involved.

Mathematics educators have concerned themselves with such questions, but they have had to make judgments on largely intuitive grounds. There is a real need to better understand the psychological principles involved. Unfortunately, however, previous studies involving rule learning have dealt only incidentally with this problem. Perhaps more important, even the best designed studies in this area are subject to criticism for failure to distinguish between structure and behavior variables. The variables chosen (e.g., rule and example given, discovery, answer given) are often merely symptomatic of, rather than basic to, what is involved.

The fundamental assumption underlying our approach to the problem was that more rapid progress can be made by distinguishing clearly between structure and behavior variables and by identifying the important parameters of each. I would go even farther and say that we can never hope to understand mathematics and other subject matter learning without making such a distinction.

At the time the generality study was designed, the SFL had only developed to the point where rules were defined in terms of their denotative sets of ordered stimulus-response pairs. No consideration was given to the nature of the underlying rule construct. Even so, the fact that sets can be ordered as to their inclusive-ness led naturally to the question of rule generality. Not only did this question have practical relevance but, more important from the standpoint of theory, the SFL provided a basis for rigorously
defining just what is meant by generality. One rule is said to be more general than another if the denotative set of the former includes that of the latter as a \textit{proper} subset (i.e., the former set includes all of the instances of the latter plus some of its own).

Our primary motivation for the rule generality study, then, was to "try out" this definition to see if it had the sort of straightforward behavioral implications we had hoped. In particular, notice that all S-R instances of a principle are treated equally. \textit{Any} stimulus within the scope of a rule should provide an adequate test of its acquisition. Similarly, performance on extra-scope test stimuli should be uniformly nil. Once learned, a highly general rule would, of course, be expected to induce appropriate performance on a wide variety of tasks. At the same time, however, it is quite possible that the ease of learning a rule statement, as judged by the ability to use it, may vary directly with its specificity.

Another facet of our rule generality research concerned the consistency with which a learned rule is applied. In my earlier research (e.g., Scandura, 1966b) it was found that, under certain conditions, experimental subjects respond consistently in accordance with a derived rule. When told that their first response was correct, those subjects who used a rule as the basis for responding to a first test item also used the rule on a second test item. These pilot results were obtained in a discovery-learning situation with simple materials. There was a need to extend this finding to more complex subject matters which are presented by exposition. In general, we found strong support for this contention. Apparently, people tend to respond in a consistent manner unless the context is changed or feedback otherwise indicates that the rule has changed.

For the purposes of this discussion, we shall be primarily concerned with only two hypotheses. First, performance on problems within the scope of a rule does not differ appreciably and successful problem solving is limited almost exclusively to within-scope problems. Second, the ease of learning a rule statement so that it can be applied to within-scope problems varies directly with the rule's specificity. Thus, the more general a rule statement, the poorer the learning.

To test these hypotheses we conducted two experiments, only one of which I shall outline here. The crux of the experimental design and the results can be seen in Table 1. The three groups of experimental subjects (*Ss) were undergraduates enrolled in a mathematics education course for elementary teachers. Each group was presented with one of three rule statements of varying generality. All of the Ss were then tested on the same three problems. \textit{Rule S} was the most specific and was appropriate for solving any problem in a certain class—problem one was one such problem. \textit{Rule SG} was more general and was potentially applicable to a wider range of problems. In particular, it was logically sufficient to solve both problems one and two, but not problem three. \textit{Rule G} was the most general rule; it was applicable to all three test problems.

\textbf{Table 1. Number of Correct Solutions}

<table>
<thead>
<tr>
<th>N</th>
<th>Problem One</th>
<th>Problem Two</th>
<th>Problem Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>17</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>SG</td>
<td>17</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>G</td>
<td>17</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

In agreement with the first hypothesis, there was almost no extra-scope transfer, and performance on within-scope problems was essentially the same. Notice, in particular, that the performance obtained could be predicted on the basis of a strictly logical argument. If people learn an unfamiliar rule \textit{at all} then they \textit{should} be able to apply it to any problem or stimulus within the scope of the rule but not to similar problems beyond the scope of the rule.\textsuperscript{6}
Defining rule generality in terms of the scope of the respective denotative sets of S-R pairs, however, does not provide a sufficient basis for explaining the results pertaining to the second hypothesis which was concerned with the learnability of rule statements of different generality. In the beginning, this hypothesis was based simply on intuition.

To check this hypothesis, consider the performance of the three groups on problem one. Since this was the only problem within the scope of all three rules, we expected on the basis of our hypothesis that group S would do better than group SG which, in turn, would do better than group G. As you can see, group S did do much better than the others, but the performance of the subjects in groups SG and G did not differ appreciably.

Actually, the facts are not quite this simple, but the statement appears to be a good first approximation. John Durnin and I have a study underway in the Penn Laboratory which we hope will add further clarification.

To explain these results, it was found necessary to define an underlying construct, a cognitive competency underlying the behavior we observed. This led me to the four-tuple characterization previously outlined. When the competencies necessary for interpreting the rule statements of varying generality were analyzed it became apparent that the more general the rule, the more is required of the learner. Thus, to apply a highly general rule statement, once memorized, requires that the learner be able to apply any rule of lesser generality but not conversely. To subtract any two numbers, for example, implies that 2 can be subtracted from 6, but being able to subtract 2 from 6 does not imply that the learner can subtract with any two numbers.

The fact that performance of the groups given the most general (G) and intermediate (SG) rule statements did not differ can be attributed to prior learning. In effect, those additional abilities needed to interpret and, hence, to apply rule G were not likely to have caused the college students involved any difficulty.

Since post-diction is held in generally low regard in the psychological (but not scientific) community, let me add a brief word of defense. In this particular study, we began with a precise definition of rule generality based on a preliminary version of the SFL; we conducted a study, and, finally, we used the results to extend and otherwise improve the very foundations on which our a priori analysis had been based.

What educational implications can be drawn from this study? To begin with, the results of these experiments demonstrate, in a rather conclusive fashion, the behavioral relevance of rule generality. For the most part, successful performance was noted only on tasks within the scope of verbally stated rules. When rules are presented in an expository fashion, it is normally too much to expect generalization to problems to which the principle does not immediately apply.

Potentially of even greater practical significance was the lack (there was one exception) of performance differences on the within-scope problems and the consistency results cited earlier. The former result demonstrates that, under certain specifiable conditions, any stimulus within the scope of a rule is equally as difficult to respond to correctly as any other. Furthermore, the obtained consistency results suggest that only one (new) test stimulus is needed to determine whether, in fact, a given rule has been learned. Under certain specifiable conditions, no more information is gained by using additional test instances. These results could have far-reaching implications for the development of highly efficient measuring instruments.

In addition, the pronounced tendency of the $s to attack all of the test problems in the same way, irrespective of whether the procedure used was appropriate, suggests that knowing how to solve...
problems and knowing when to use this knowledge are quite distinct. Testing for the latter ability necessarily must involve the presentation of extra-scope problems.

Most important, in view of the problem I posed originally, this study helps to reconcile the views of subject matter and learning specialists as to how general the presentation of material ought to be. While one may expect maximum transfer potential by introducing rules of the greatest possible generality, this transfer potential is bought at a price the teacher may or may not be willing to pay. The greater its generality, the harder a rule statement is to interpret. Hence, before determining how general a rule statement should be, it is essential to first consider whether the students involved have the necessary requisite interpretative abilities (Scandura, 1967c).

See footnote 5. I might add that some transfer does take place and we are now attempting to pin down the source of this transfer.

Actually many teachers already do this, at least intuitively. All that the present study does is to make these intuitions more explicit. Frankly, I am always pleased when our results conform to what might be called "common sense." Too often psychological results, while perhaps relevant to animal learning or the memorization of nonsense-syllable lists, have very little to say about the learning of mathematics and other subject matters.

*Discovery Learning* (Roughead & Scandura, 1967)

One of the fundamental assumptions underlying several of the new mathematics programs is that discovery methods of teaching and learning increase the student's ability to learn new mathematics. Indeed, this assumption has guided the development of many new curricula in all of the subject matter fields. Attempts to demonstrate advantages or disadvantages of self-discovery, however, have either failed, been open to criticism on scientific grounds, or are seemingly inconsistent even when apparently well-controlled.

Research on discovery learning has been confounded by differences in terminology, the frequent use of multiple dependent measures, and vagueness as to what is being taught and discovered. While the difficulties due to the use of inconsistent terminology can often be minimized by a careful reading of research reports, the use of multiple dependent measures often makes it impossible to unambiguously interpret experimental results. Several investigators, for example, have found that groups which are given an expository statement of a rule perform better on transfer tests than groups which are required to discover this rule for themselves from instances of the rule. The obtained differences in transfer ability, however, may well have been because the discovery groups simply did not discover the rule.

Gagne and Brown (1961) overcame the dependent measure problem by equating original learning and investigating only transfer differences on new problems. On the basis of an analysis of the learning programs used in the Gagne and Brown study, Eldredge (1965) hypothesized that the obtained results could have been due to a number of flaws in the programs used. Eldredge proposed that exposition and discovery situations may be better characterized as differences in order of presentation. Exposition may be defined as giving rules and then examples of these rules, whereas discovery may be defined as giving the examples and then the rules. Contrary to his hypothesis, however, his discovery group did evidence more transfer than his exposition group. Unfortunately, there were a number of difficulties with the study that make the results difficult to interpret.

The Set-Function Language was used as an aid in removing these difficulties. The resulting analysis of what is involved in discovering rules indicates that discovery learners learn "something" by which they can derive solutions to an entire class of problems. Roughead and I called this "something" a *derivation rule*. Thus, discovery learners who actually succeed in making a discovery, should be expected to
perform better than expository learners on tasks which are within the scope of such a derivation rule. If the new problems presented have solutions beyond the scope of a discovered derivation rule, however, there would be no reason to expect discovery $s$ to have any special advantage.

This study was concerned with two basic questions. First, can "what is learned" by discovery be identified and if so, can that knowledge be taught by exposition with equivalent results? According to the SFL, all behavior is controlled by rules so that there might well be some identifiable rule which is equivalent to "what is learned" by discovery. Specifically, we hypothesized that "what is learned" by guided discovery in the Gagne and Brown study could be identified and, hence, could be presented by exposition. The second question was, how is "what is learned" by discovery dependent on what the learner already knows and/or the nature of the discovery treatment itself? More particularly, we hypothesized that the discovery of a derivation rule can actually be hindered by having too much prior information.

Assuming transfer depends only on whether or not the derivation rule is learned, sequence of presentation should have no effect on transfer so long as the subject is forced to learn the underlying derivation rule. That is, presenting the derivation rule by exposition or by guided discovery either before or after presenting the desired responses should have no effect on performance on transfer tasks. On the other hand, if a discovery program simply provides an opportunity to discover (with hints as to the solution) but does not guide the learner through the derivation procedure, sequence of presentation might well have a large effect on transfer. Assuming the learner is capable and motivated, he may well succeed in determining the appropriate responses and, in the process, discover a derivation rule. It is not likely, however, that a person would learn such a derivation rule if he already knew the correct responses.

We made three hypotheses: (a) what is learned by guided discovery can be presented by exposition with equivalent results; (b) presentation order is not critical when learners are effectively "forced" to learn derivation rules, either by exposition or by guided discovery; (c) presentation order is critical when the discovery guidance provided is specific to the respective responses sought rather than relevant to a general strategy or derivation rule.

The task we used was essentially identical to that used by Gagne and Brown, and Eldredge and involved finding formulas for summing the terms in number sequences. That is, the stimuli were number series, like 1 + 3 + 5 + 7, and the responses were formulas in $n$, the number of terms, for summing such series. For example, the appropriate formula for summing $1 + 3 + 5 + 7 + ... + 2n-1$ is $n^2$.

Using the SFL as a guide (i.e., by identifying, in turn, D, O, and R) we were able to identify that derivation rule taught in the guided discovery program used by Gagne and Brown. On the basis of this knowledge, four programs were constructed : (a) the formula-given program simply stated the correct summing formula for each problem series confronted in the learning program; (b) the guided discovery program remained essentially as it was in the earlier studies; (c) the expository program consisted of a precise expository description of that derivation rule which was presumably equivalent to that learned by guided discovery. It consisted of a general procedure by which the desired formulas could be derived; (d) in the opportunity-to-discover program, the problem sequences were presented along with encouragement and hints as to what the desired formulas were. These hints involved such statements as "the formula has a '2' in it." The same number sequences were used in each of these four programs.

Seven treatments were constructed by combining these four basic programs. After going through a common introductory program, one group of subjects simply went through the formula program. The other six groups received the formula program together with one of the other three programs. Two of these six groups received the guided discovery program together with the formula program; two
additional groups received the expository and formula programs; and the final two groups received the opportunity-to-discover and formula programs. One group, in each of the resulting three pairs, received the programs in one order; the other group received them in the reverse order. Only the order of presentation was varied. After finishing their respective programs, all of the students were tested on new series to see how well they could determine the appropriate summing formulas.

The results were rather clear cut. Essentially, the group given the formula program only and the group given the formula program followed by the opportunity to discover program performed at one level. The other five groups performed at a common and significantly higher level. Two points need to be emphasized. First, "what is learned" during guided discovery learning can at least sometimes be taught by exposition—with equivalent results. Of course, there are undoubtedly a large number of situations where because of the complexity of the situation, "what is learned" during discovery cannot be clearly identified. It is still an open question, for example, whether still higher order derivation rules, which have a more general effect on the ability to learn, may be learned by discovery. If we believe that the answer to this question is affirmative, there is no real alternative to learning by discovery unless or until we can identify just what is involved. Nonetheless, intuition-based claims that learning by self-discovery produces superior ability to solve new problems, as opposed to learning by exposition, has not withstood experimental test. The value of some forms of discovery to transfer ability does not appear to exceed the value of some forms of exposition. Apparently, the discovery myth has come into being not so much because teaching by exposition is a poor technique as such but because what has typically been taught by exposition leaves much to be desired. As we identify just what it is that is learned by discovery in more and more situations, we shall be in an increasingly better position to impart that same knowledge by exposition.

The second point to be emphasized concerns the sequence effect. While the group that was given an opportunity to discover and then the formula program performed as well on the transfer problems as those given the derivation principle in a more direct fashion, the group given these programs in the reverse order (i.e., the formula-opportunity group) did no better than those given the formula program alone. In effect, if a person already knows the desired responses, then he is not likely to discover a more general derivation rule.

An extrapolation of this result suggests that if <$ knows a specific rule, then he may not learn one which is more general even if he has all of the prerequisites and is given the opportunity to do so. The reverse order of presentation may enhance discovery without making it more difficult to learn more specific rules at a later time. In effect, prior knowledge may actually interfere in a very substantial way with later opportunities to discover.

This sequencing result may have important practical and theoretical implications. The practical implications will be attested to by any junior high school mathematics teacher who has attempted to teach the "meaning" underlying the various computational algorithms after the children have already learned to compute. The children, in effect, must say to themselves something like, "I already know how to get the answer. Why should I care why the procedure works?" Similarly, drilling students in their multiplication facts before they know what it means to multiply, may interfere with their later learning what multiplication is. Let me make this point clear, because it is an important one. I am not saying that we should teach meaning first simply out of some sort of dislike for rote learning—for certain purposes rote learning may be quite adequate and the most efficient procedure to follow. What I am saying is that learning such things as how to multiply without knowing what multiplication means, may actually make it more difficult to learn the underlying meaning later on. The theoretical implications are even more interesting for the researcher and, in fact, may be crucial to any theory based on the rule construct and framed in the SFL—but space limitations demand that I not go into that here.
In addition to the studies described above, we have conducted a number of other studies, which are based on the SFL. One of these studies is designed to help clarify the role of attribute and operation cueing in learning mathematical rules. Another deals with the role symbolism plays in mathematics learning. We are also involved in developing a completely new methods course in mathematics for elementary school teachers which is intimately tied to this point of view.

CONCLUDING REMARKS

In this paper, I have tried to share with you some of my thoughts on the psychology of mathematics learning—or what I like to think of as the emerging discipline of psycho-mathematics. As is obvious by now, I chose not to do this in a direct manner but by pointing out certain inadequacies in existing behavioristic theory as it relates to mathematics learning and, more important, by describing an alternative scientific language and showing how it can be used to formulate research questions involving mathematical learning and performance.

The mathematical notions of sets and functions were proposed as a basis for representing: (a) the denotative or observable aspect of rules and principles, (b) the underlying knowledge itself, and (c) meta-linguistic descriptions or representations of the underlying knowledge. Very recently, I have become intrigued with the idea of integrating these ideas by borrowing the very fundamental but more abstruse mathematical notion of a functor. The functor may also make it possible to distinguish in a very precise way between the sort of “ideal” competencies which have long been championed by linguists and competencies as they actually exist in human beings. In the present version of the SFL, it has only been possible to deal with idealized rules (i.e., competencies).

NOTE: Alas, “functors” were just another false alarm introduced as a result of discussions with a colleague in mathematics at Penn. It was just one of many mathematical concepts tried and tested as a foundation for what ultimately became the Structural Learning Theory.

To insure continued progress, it seems to me that a dual emphasis is needed in psycho-mathematics. On the one hand, there are a large number of unspecified, but crucial, “ideal” competencies which underlie mathematical behavior. These need to be identified. This is a problem area which in many ways is analogous to linguistics and whose solution will require a substantial knowledge of mathematics coupled with a behavioral point of view. There is also the urgent need to consider how the inherent capacities of learners and their previously acquired knowledge interact with new input to produce mathematical learning and performance. Again by analogy, we have a field much like psycho-linguistics, a field which will seek to integrate knowledge concerning psychology and mathematical structures and strategies. I feel that this kind of distinction will prove crucial to any deep understanding of how mathematics (and other subject matters) are learned.

REFERENCES


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. (a)


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


The notion of a restricted rule or strategy was introduced. It was hypothesized that extra-scope transfer depends on the extent to which a statement of strategy may be viewed as a restriction of a more general strategy. 66 high school Ss were taught a restricted statement (S', SG', or G') of 1 of 3 strategies of varying generality, S (= 8') < SG (SG') < G (GO. 22 Ss served as a control (C). All Ss were tested on 6 problems, the 1st 2 within the scope of the most specific strategy (S), the 2nd 2 within the scope of only the more general strategies (SG and G), and the last 2 only within the scope of strategy G. Statements S', SG', and G' were directly applicable only to the 1st 2 problems. Groups SG' and G' evidenced extra-scope transfer. Groups S' and C did not. In addition, performance on the 2nd problem of each pair was contingent on performance on the corresponding 1st problems indicating that "what is learned" may be determined by performance on single test items and used to predict performance on additional similar-scope problems. Suggestions were made for future research.

*COMMENT: Following was the rationale for the study:*

Scandura, Woodward, and Lee (1967) demonstrated that performance on transfer tasks is generally in accord with the logically determined scope of rule and strategy statements. 2 In each of two experiments, iSSs were presented with one of three statements of rules (or strategies) of varying generality and were tested on three problems. The first problem was within the scope of all three rules; the second, within the scope of only the two more general rules; and the third, within the scope of only the most general rule. In most instances, there was essentially no difference in the level of performance on the within-scope problems and no extra-scope transfer (to problems to which the rule did not directly apply).

There was one glaring exception involving extra-scope transfer. In Experiment II, Ss given the statement, "50 X 50," which was directly applicable only to Problem 1, performed equally as well on Extra-Scope Problem 2 as did those Ss given the statement, "n X n," where the dimension (i.e., variable) n was allowed to vary over the positive integers. This result obtained even though "n X n," was directly applicable to both Problems 1 and 2. While the study itself was inadequate to specify the source of this transfer, a post hoc analysis of the experimental treatments indicated that "50 X 50," was the only rule statement included in the study which was in some sense a restriction of a more general rule or strategy. The statement, "50 X 50," could be obtained from the more general statement "n X n," by replacing the response determining dimension, n, by the value 50. More generally, it would appear that a restricted statement may be viewed as one obtained by replacing response-determining dimensions (see Scandura, 1966, 1967a, 1968a) in the statement of a general rule or strategy with the specific
values of a particular instance. The authors, therefore, conjectured that a restricted rule statement might well provide a basis for generalization to all problems within the scope of the corresponding unrestricted rule. The primary purpose of this study was to test this hypothesis.

A secondary purpose was to obtain further information on the "consistency" hypothesis. Under certain conditions, it has been found that transfer to one instance of a rule almost invariably implies transfer to other instances of the rule (Scandura, 1966, 1967a, 1967b, 1968b; Scandura et al., 1967). As was the case with extra-scope transfer, however, one exception to the consistency hypothesis was obtained in the study by Scandura et al. (1967). The level of performance on one within-scope problem was considerably below that on the others. Whereas the response determining values of the homogeneous problems differed along a single dimension, the exceptional within-scope problem differed along a second dimension as well. Taking this observation into account, a modified form of the consistency hypothesis was advanced. It was hypothesized that if transfer to one problem indicates that a particular rule or strategy (e.g., "50 X 50") has been generalized along one or more familiar dimensions (e.g., to "n X n") then transfer to additional problems along the same dimensions (and within the scope of a less restrictive rule) should also be expected.

1 The participation of the second author was made possible by a United States Office of Education Graduate Training Grant to the University of Pennsylvania in Mathematics Education Research. The authors are indebted to I. R. Klingsburg, head of the mathematics department, and the participating students at the West Philadelphia High School for their generous cooperation in making this study possible.

2 The terms "rule" and "strategy" are used synonymously throughout this paper. While "rule" is the preferred technical term (e.g., Scandura, 1968a), "strategy" better connotes more complex multi-phased rules of the sort used in this study.


NOTE: This paper offers a perspective on research in mathematics and science education, including strengths and weaknesses of various approaches to educational research with a focus on teaching and learning.


NOTE: This paper is the first of three offering a more complete, better framed presentation of the Set-Function Language (an intermediary on the way to the Structural Learning Theory).

Quotation from p. 304: This paper is the first of three which together constitute a monograph on 'New Directions for Theory and Research on Rule Learning'. The purpose of this paper is: (I) to introduce a new scientific language, formulated in terms of the mathematical notions of sets and functions, and (2) to contrast the S — R language with this set-function language (SFL). In the SFL, the rule or principle 4 is taken as the basic unit of behavior. This adoption turns out to be highly parsimonious from a mathematical point of view. The second paper in the series reports some of the related empirical research by students and I have conducted. The emphasis is on the guidance provided by the SFL in formulating the research and on modifications and extensions of the SFL suggested by the obtained results. The last paper is concerned with SFL based analyses in such problem areas as reversal and non-reversal shifts, Piagetian conservation tasks, and symbolic and concrete learning. Theoretical direction is also given.


NOTE: The second paper summarizes and comments on a broad range of related empirical research.

NOTE: The third applies the SFL to a variety of problem areas. Like most of my articles and publications they can or soon will be available at [www.TutorITweb.com](http://www.TutorITweb.com).

COMMENT 46-49: Other assorted articles.


COMMENT 50: The following paper was sort of motivated by colleagues at the Penn psychology department who believed that to become official theoretical work had to be published in the Psychological Review. In fact, this was not true in my case as it was shortly followed by my more definitive papers 55 (and to a lesser extent in 53).


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 classes of entities denoted by those 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 (acting) on rules. This is a direct generalization of association 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.

NOTE: This article begins with observations that led to the Set Function Language (SFL) including the ability to reliably predict future behavior, defining encoding and decoding, symbols and icons, higher order rules acting on rules, and operational definition of what (rule) is learned. In a footnote (11), I proposed that research might better focus first on understanding what is learned in complex behavior rather than the precise values of boundary conditions imposed by short term memory and the like.

COMMENT 51-54: The next few papers gained further empirical evidence (e.g., 52), showed how rules can be used to account for creative behavior in mathematics (53) and with a more precise account of what exactly is a rule showing that associations and concepts are just special cases (of rules) (54). All this set the stage for my first paper detailing the essence of the Structural Learning Theory in 55.


Deterministic Theorizing in Structural Learning: Three Levels of Empiricism

JOSEPH M. SCANDURA
MERGE and University of Pennsylvania

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

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

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

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

There are three main ideas which my title conveys. The label "structural learning" sets the whole tone for the title, so we consider that first. Structural learning refers to the knowledge a person may have and the behavior (and learning) which this knowledge makes possible. More specifically, structural learning is concerned with complex human learning and behavior which cannot naturally be studied without giving explicit attention to what the subject knows before he enters the learning or behaving situation. Any attempt to study mathematics learning, for example, with reference only to the stimulus situation would be folly to the nth
degree. Individual differences in prior knowledge and other intellectual skills in mathematics may be very great indeed, and these differences must be taken explicitly into account in any theory that is to provide a viable account of complex mathematics learning. It should be noted parenthetically that one of the primary requisites for selecting tasks in most traditional studies has been that prior learning be of minimal importance. The reference here, of course, is to experiments on serial and paired-associate learning, classical conditioning, and the like. Dependence on prior knowledge, then, is important to my conception of structural learning. But this alone is not sufficient. The knowledge involved must also have a reasonably clear structure. In this sense, mathematics, for example, tends to have a clearer structure than, say, the social studies or the humanities. The fact that grammarians, like Harris and Chomsky, have been able to make as much progress as they have in linguistics attests to a good deal of structure in language as well.

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

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

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

In order to be completely honest, I must mention one further reason why deterministic theorizing appeals to me. I am basically lazy. I have done a good deal of traditional behavioral research, but I dislike with a passion poring over reams of raw data or computer printouts, especially when I know that, no matter what statistics are used to summarize the data, I am losing much, if not most, of what is important. It is perhaps this distaste as much as anything else which has moved me to search for a new and better way to do empirical research on complex human learning. How much nicer to have data which is clear-cut, 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 postdoctoral student being initiated into mathematical psychology at Indiana University. The time was the summer of 1962, and the field was bright and promising.

As part of my orientation, I was routed about to visit a number of the more prominent names on campus, including one very fine physiologist. Caught up by the emphasis on mathematics given by the psychologists, I asked him what kinds of mathematics he found most useful in his work, and how he used it. His answer was, "We count." After getting over my initial shock, I began to see the logic of his answer, and have been trying to meet his ideal ever since.

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

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

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

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

This approach may be a viable one in mathematics, where one aims for abstraction. One never knows where mathematical theories may ultimately prove useful (i.e., be applied), and
it would undoubtedly be a mistake to tie them in too closely to any particular model, by specifying aspects of these particular models with which the theory does not deal. This is not true in science, however, where the ultimate aim may be to devise theories that deal with more of the particular reality in question. A theorist may have many more kinds of phenomena in mind in attempting to construct a theory than he can possibly handle at one time. To get around this problem, he may purposefully ignore for a time certain of these phenomena to facilitate constructing what might be called a partial theory--a theory which deals with part of the reality but not all of it.

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

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

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 other behaviors one might wish to attribute to the knower (Chomsky, 1957; Scandura, 1970a, forthcoming). These additional behaviors constitute the predictions.

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

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

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

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

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

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

We can get a far more powerful and simple characterization by allowing rules to operate, not just on ordinary stimuli, but on other (lower order) rules as well. 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:

\[(1) \quad a, b \rightarrow a + b\]
\[(2) \quad x, y \rightarrow x \circ y \quad \Rightarrow \quad x, y \rightarrow x \circ' y\]

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 any binary operation, generates a rule for performing the corresponding inverse operation (denoted \(\circ'\)). 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,

\[(3) \quad a, b \rightarrow a - b,\]

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?

---

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

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

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

(Note: x, y, o are different from x, y, 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 (B) of rational numbers under addition. Rule (5) is extremely general and makes it possible to generate the composite (rule) of any pair of given rules such that the output of one of the rules serves as the input of the other.

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

(6) \[ [x, y \rightarrow x \circ y] \Rightarrow [x, y \rightarrow x \circ y] \]

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

(7) a, b \rightarrow a +’ b or a, b \rightarrow a - b where +’ = - .

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

In summary, the essentials of the theory of knowledge as outlined are just these. (1) The knowledge of any given individual at any given stage of learning can be characterized in terms of a finite set of rules. This implies among other things that there may be as many different theories of knowledge as there are individuals - or, equivalently, as many theories as there are conceivable curricula to be mastered.

COMMENT - 2006 Update: The distinction made here is more sharply formulated by representing competence (what needs to be learned in terms of structural & procedural Abstract Syntax Trees ASTs and representing knowledge attributed to individuals via assessment as programs (data & processes) corresponding to slices through ASTs.

(2) Rules may act on classes of rules as well as on simple stimuli. Allowing rules to act in this way amounts to a simple but conceptually major revision of existent competence theories.

(3) For purposes of the theory, it is assumed that the rules may be combined at will and without error as needed. Stated differently, the idealized knower is assumed to have mechanisms available for putting the rules attributed to him to use.4

---

4 There will always be behavior, of course, which cannot be generated by any given finite set of rules. Roughly speaking, when translated into behavioral terms, Godel's (1931) incompleteness Theorem suggests that no matter how bright an individual, there will always be certain behaviors he will not be capable of performing.
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 this task that we now turn—the task of introducing mechanisms of idealized performance, learning, and motivation into our formulation. The purpose of adding such mechanisms to the theory of knowledge is to obtain an extended theory that deals explicitly with the way in which available knowledge is put to use. This more encompassing theory is still a partial theory, however, one which applies only where subjects are unencumbered by either memory or their intrinsically limited capacity to process information. It should be emphasized, however, that it is a theory, which is assumed to apply no matter what knowledge an idealized subject has available. Thus, even though the knowledge had by different individuals may vary greatly, the same theory of idealized behavior is assumed to hold over all individuals.
The basic assumption on which this theory rests is that people are goal seeking information processors. In this case, much of what a subject knows becomes irrelevant once a goal situation is specified. Thus, at any given point in time, only a small fraction of the rules available to a knower may be applicable - namely, those rules which may be used directly or indirectly in satisfying the given goal.

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

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

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

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

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

Furthermore, the assumption has a number of important implications. One of these is that it provides an adequate basis for determining what might be called a subject's behavior potential, relative to a given class of rule-governed (RG) behaviors. It may be noted in this regard that it is one thing to devise a procedure (rule) which accounts for a given class of RG behaviors and quite another to identify that subclass of behaviors of which a given subject is capable. The first problem is an analytical one and involves inventing a procedure, which accounts for the given class of RG behaviors. No psychological assumptions are involved. Determining a subject's behavior potential, however, necessarily depends on what can be assumed about the mechanisms that govern human behavior. The basic idea goes like this: Given any familiar class of RG behaviors, like the class of addition tasks, we can usually identify those rules (algorithms) which the subjects in question are likely to use in solving the problems. We do not automatically know which aspects of these algorithms any given subject is capable of, however. To find out, we must test the subject. But on which instances is he to be tested -- how are they determined? The standard approach, of course, is just to select a random sample of test instances and then make probabilistic predictions about future performance on other instances in the class.

This approach is rejected in favor of systematic selection of test instances and deterministic prediction on individual items. To see how this can be accomplished, we first note that every algorithm for solving a given class of (RG) tasks can be represented by a directed graph (see Scandura, forthcoming). For example, the task of generating the next numeral in Base Three Arithmetic can be represented as follows.
Fig. 2. Sample stimuli and responses for the task of generating the next numeral in Base Three Arithmetic, together with the (total) graph of a procedure for generating the behavior, and four graphs representing the four behaviorally distinguishable paths through this procedure.

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

The subgraphs at the bottom of Fig. 2 correspond to the four possible paths through this procedure, which may be used in solving particular problems. Since the constituent rules are all atomic, it follows that each of these paths also acts in atomic fashion. Hence, to determine the behavior potential of a given subject with respect to this algorithm, we need only select one test instance for each path. In this case, the base-three stimulus (response) numerals, 101 (102), 2 (10), 1 12 (120), and 222 (1000), correspond respectively to the four possible paths. Accordingly, the behavior potential of a given subject on this class of tasks can be uniquely specified by his performance on just these four test instances -- as long as the atomic assumption is valid. (Hence, the assessment is conditional.) Any other set of four stimulus representatives of these paths, of course, would do equally well. Although its role was hidden in describing this method of assessing behavior potential, the methods' validity depends directly on the simple performance mechanism. According to this mechanism, if a subject has a particular path available for solving a given task, then he will use it and use it consistently on all instances to which it applies. That is, of course, assuming that the subject's goal remains the same.

None of this is idle theoretical speculation. Over the past several months one of my students, John Durnin, has collected a good deal of evidence which provides support which goes far beyond the bounds of what is normally considered sufficient evidence. In a total of 204 predictions utilizing a variety of tasks and subjects of greatly differing abilities and grade levels (from the preschool through graduate school), we have had a grand total of seven errors in prediction. A sample of this data is given in Table 1 for a procedure involving eight paths.

Let us next consider what happens when a subject has not explicitly learned a rule for achieving a given goal. In this case, the subject has a problem in the classical sense - a problem situation, a goal, and a barrier between them.
The major theoretical problem is to explain what happens when a subject is confronted with such a situation. If the problem can be formulated in a way that lends itself to prediction, so much the better. Why certain people are able to solve some problems for which they have never learned a specific rule, whereas others cannot, is a question of paramount interest. We want to know exactly what is involved, and why subjects perform as they do.

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

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

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

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

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

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

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

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

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

![Fig. 3. A schematic representation of the hypothesized mechanism for problem solving. R1 and R2 represent rules for converting yards into feet and feet into inches, respectively. HR refers to the higher order rule for generating composite rules.](image)

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 Show to use two simple rules, comparable to those described above (e.g., for converting yards into feet). These rules are denoted r1 and r2 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 r11 or r12. As shown, only one of the subjects was initially successful on this type of problem. Next, we taught the subjects with neutral materials how to combine pairs of simple rules such as the ones they had been taught. This time we were successful with all but one subject. (To accomplish this we also had to teach many of the subjects what it was they were trying to do—that is, find a rule that could be used to solve problems such as that requiring r1 o r12 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 r21 and r22) and then tested him to see if he could solve the corresponding composite problem, which required r2 o r2 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.

<table>
<thead>
<tr>
<th>TABLE 2 Summary of Experimental Procedure and Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of Subject</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>R_{11}</td>
</tr>
<tr>
<td>R_{12}</td>
</tr>
<tr>
<td>R_{11} \times R_{12}</td>
</tr>
<tr>
<td>HR</td>
</tr>
<tr>
<td>R_{21}</td>
</tr>
<tr>
<td>R_{22}</td>
</tr>
<tr>
<td>R_{21} \times R_{22}</td>
</tr>
<tr>
<td>R_{31}</td>
</tr>
<tr>
<td>R_{32}</td>
</tr>
<tr>
<td>R_{31} \times R_{32}</td>
</tr>
<tr>
<td>R_{41}</td>
</tr>
<tr>
<td>R_{42}</td>
</tr>
<tr>
<td>R_{41} \times R_{42}</td>
</tr>
</tbody>
</table>

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

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

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

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

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

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

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

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

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

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

TABLE 3
Results of Rule Selection Study

<table>
<thead>
<tr>
<th></th>
<th>Rule A</th>
<th>Rule B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>52</td>
<td>2</td>
</tr>
<tr>
<td>Test 2</td>
<td>13</td>
<td>64</td>
</tr>
</tbody>
</table>
The results are summarized in Table 3. This table shows that whenever a S selected a path of Rule A on Test 1, he almost invariably (52 times out of 54) selected a path of Rule A on Test 2. Rule B selections were consistent with the hypothesis 64 times out of 77.

To recapitulate, it should be re-emphasized that everything that has been said so far about learning, performance, and motivation only applies in situations where memory and the limited capacity of human subjects to process information do not enter. The proposed mechanisms have all assumed an information processor with an essentially unlimited ability to process information, and with perfect memory for previously acquired knowledge. This definitely clues not imply that the theorizing so far is of little value. That conclusion would be wrong on at least two counts. First, there are many practical situations in structural learning where memory is of minimal concern. In problem solving, for example, the S is almost always given all of the paper, pencils, and other memory aids that he needs. Typically, we also do our best to insure that the necessary lower-order rules are readily available, even to the extent of making textbooks available. The concern is generally with whether or not the individual can integrate available knowledge to solve problems. Considerations such as whether he can do it in his head or not, time to solution, and so on, are of secondary concern (cf. Scandura, 1967). Second, questions of memory can usually be eliminated in experimentation by insuring that relevant rules and memory aids are available to the subject. This can normally be accomplished by training.

TOWARD A THEORY OF MEMORY & INFORMATION PROCESSING

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

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

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

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

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

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

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

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

It should be emphasized, however, that the theory is essentially deterministic, and applies only where one is dealing with highly structured materials, where one can make reasonable assumptions about the kinds of rules used in storage and retrieval. The theory is not designed to handle data from typical short-term memory experiments. (Even here, however, it can be
suggestive (Scandura, forthcoming).) Rather, the theory calls for quite a different kind of memory experimentation -- experimentation with relatively complex and more highly structured materials, where explicit attention is given to the goal conditions imposed on the S and the kinds of storage and retrieval capabilities with which he enters the situation. The only fundamentally new hypothesis involves information, processing. The basic assumption is that each individual subject has a fixed finite capacity for processing information. While this capacity may vary somewhat over individuals, it is assumed to be of the order, 7 it 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 where the "dropoff" would occur. What was needed was a general scheme for calculating memory load for any given rule - but developing one did not prove to be a simple task. We have recently come up with something that seems promising, however, and about a week ago, our data reached about 80% level of predictability, which may be about as good as can reasonably be expected with this type of task. 5

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

5. 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.
FIG. 4. The performance of four subjects on the indicated tasks with percentage of perfect responses plotted against memory load. For comparative purposes, repeating n digits had a calculated memory load of n, repeating n digits and then saying "1" had a memory load of n + 1, repeating n digits after saying "1" had a memory load of n + 2, addition of two 2, 3, and 4 digit numbers without carrying had memory loads of 7, 8, and 9, respectively, with carrying, the memory loads were 8, 9, and 10.

CONCLUDING COMMENTS AND IMPLICATIONS

The foundations of three partial theories of structural learning have been described and some relevant data have been reported. First, a partial theory of structured knowledge was proposed, in which it was argued that the knowledge had by any given S may be characterized in terms of a finite set of rules. By allowing rules to operate on other rules (in the set), it was shown how new rules could be generated. Examples were also given to show how these new rules, in turn, could account for creative behavior. With the addition of several performance assumptions, this theory was extended so as to account for learning, performance and motivation under idealized conditions where behavior is unencumbered by memory. Finally, we outlined how memory and information processing might be dealt with, and reported some preliminary data in favor of our main hypothesis. Even the most encompassing theory, however, does not deal with a number of behavioral phenomena, specifically the ultra-short term after images reported by Sperling (1960), Averbach and Coriell (1961), and others. Whether the theory might be extended further to account for these phenomena is difficult to say. But, in any case, this might well be left until later given the large number of questions raised by the theory as it presently exists.
The theory itself represents a sharp departure from existing theories of cognitive behavior, although it does have some things in common with existent competence and information-processing theories. The differences even here, however, are not minor, but have a fundamental effect, both on theoretical adequacy and on the very kinds of empirical questions one asks. Probably the most basic departure is the idea of introducing different levels of empiricism, and the possibility of deterministic theorizing at each of these levels. According to this view, it is possible to do behaviorally relevant empirical research at at least three quite distinct levels. Although all competence models, such as those proposed by Chomsky in linguistics, purport to deal with knowledge, concern traditionally has been limited primarily to the so-called mature speaker or hearer who effectively knows all there is to know about the language. In the present formulation, it is just as reasonable to talk about the knowledge had by different individuals, naive ones as well as mature. This is an extremely important characteristic in dealing with subject matters like mathematics, science, or even language, where knowledge is not a static thing, but grows with experience.

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

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

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

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

REFERENCES

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

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

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

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

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

3 Basically, the technique involves calculating for each step of the given algorithm: (1) the number of states needed to determine future states, (2) the number of operators needed to determine future operators, and (3) the number of subsets of the needed states and operators which must be distinguished in completing the procedure. Details will be published in a future article.
COMMENT 56-61: The first two articles are commentaries on chapters in my book Structural Learning II: Issues and Approaches. Chapters in this book were written by a number of leaders in associated fields, each describing their research in what was broadly conceived as structural learning. Among the authors:

56. Scandura, J.M. A reply to Wittrock. In Structural Learning II: Issues and Approaches; see B-4.
57. Scandura, J.M. Theoretical note: S-R theory or automata? September 10, 1970. (S-R Theory or Automata, A Final Word. In Structural Learning II: Issues and Approaches; see B-4.)
64. Ehrenpreis, W. & Scandura, J.M. Algorithmic approach to curriculum construction: A field test. Journal of Educational Psychology, 1974, 66, 491-498. (Also in Problem Solving; see B-6.)
70. Scandura, J.M., Durnin, J.H., & Wulfeck, W.H., II. Higher-order rule characterization of heuristics for compass and straight-edge constructions in geometry. Artificial Intelligence, 1974, 5, 149-183. (Also in Problem Solving; see B-6.)
80. Scandura, J.M., Wulfeck, W.H., II, Dumin, J. H., & Ehrenpreis, W. Diagnosis and instruction of higher-order rules for solving geometry construction problems. In 1974 Proceedings; see B-5. (Also in Problem Solving; see B-6.)


113. Scandura, J.M. Problem solving in schools and beyond: Transitions from the naive to the neophyte to the master. Educational Psychologist, 1981, 16, 139-150.


117. Scandura, J.M. Reaction to "When to use CAI in training." Training/HRD, 1981. (731 Hennepin Avenue, Minneapolis, MN 55403)


At the present time, the computer has three major, but quite distinctive, roles to play in education: (a) as an object to be understood both in relationship to the circumstances and society in which we live and as useful means (when combined with appropriate software) for getting things done more efficiently; (b) as an object of study in its own right, as knowledge and skills to be mastered, and (c) as a means of assisting the learning process.


To my mind, and undoubtedly to the minds of many cognitive theorists, Skinner is contemporary behaviorism personified. One might therefore wonder whether Skinner has anything significant to say about problem solving despite his long and illustrious career in psychology. Frankly, I am somewhat surprised at the extent to which Skinner’s ideas on problem solving find parallels in contemporary cognitive theory — especially since the essence of his paper — as originally published in 1966, when only a few of us were seriously pursuing the study of problem solving. A tremendous amount of work on problem solving has been done since that early period. Hence, it is of some interest to consider the extent to which Skinner anticipated more current developments, as well as where his ideas appear lacking. One of the first things I noticed was that Skinner’s description of the problem-solving situation is reminiscent of more recent information-processing representations. Specifically, Skinner refers to a problem as a situation in which we cannot emit any previously learned responses.
Compare this with “problem situations which cannot be solved via any previously learned rule” (Scandura 1971, p. 35).

Skinner’s distinction between first- and second-order rules is also more than just a bit reminiscent of my own distinction between higher- and lower-order rules (e.g., Roughhead & Scandura, 1968; Scandura 1970; 1971; 1973, pp. 205-13). It should be emphasized in this regard, however, that the order of a rule is a function of its level of use in solving particular problems, not a characteristic of the rules themselves (e.g., see Scandura 1977). Certainly, the elements on which higher-order rules operate must include rules, but whether elements are considered rules atomic entities is a function of the level of detail expressed in the cognitive representation. More refined analysis is always possible, and may indeed be necessary, depending on the experimental subjects whose behavior is to be explained.

Most contemporary problem-solving researchers would un-externally and self-evidently many other remarks made by Skinner; For example, Skinner’s remark to the effect that “second-order heuristic rules...can be followed as ‘mechanically’ as my first-order rule” has been demonstrated empirically in a range fine of studies on rule learning beginning in the early 1960s (e.g., Roughhead & Scandura 1968; Scandura 1964; 1969). Other similarities are less direct but, I think, equally real. (Inc, makes a major distinction between rule-governed behavior and contingency-shaped behavior, largely, it would ap-pear); to justify his continued commitment to behavioral con- tingencies. This distinction too is commonplace in modern theories of cognition but for quite different reasons. It corre- blons rather-directly to what is often referred to as the distinc- tion between nonexpert and expert knowledge and behavior. The non expert’s behavior is largely rule governed, (based on procedural knowledge) whereas the expert’s is more automatic (regulating on structural or more holistic knowledge).

Coven this degree of overlap, one might wonder what the last 1 &-20 years of work on cognitive processes has accomplished. [If Pe'knps the old behaviorist had it right from the beginning, if factoring the obvious benefits of talking about new problems in terms of behavior only familiar to Skinnerian behaviorists), ”fishman’s proposals have a fundamental irritatition. They gloss ! 'ver’or’ ignore important considerations and simply are r. ot; and given precise to allow prediction of behavior on nontrivial . P(b(cms, much less the control that Skinner himself would recta to demand. One of the major contributions of modern cognitive theory ”Over ttre gast 20 years is the development of precise languages f ”4 nalr it possible to represent cognitive processes in great )!( ’&tbk. ’Although there are differences of opinion as to what lndd I ”ff’-f”inage to use, and even whether it makes any differer. ce ! anguage is used or how to derive such representations in ~ ”’ & first place (Scalnruna 1982), most cognitive theorists are ’Wc)f on the need for such language. Most would also onc on the need for general purpose control mechanisms, ””Luutjr again there are major differences of opinion.

Skinner’s formulation says nothing ” th“ mar... . Eai- would not be a serious omission if S4nxer »xi p””die” n- control con)plex problem-solving behavi)err us f “2 con”’”ncn s and the like. But to my knowledge this Lrs ne e. le n crwe empirically. ’(It is also true that many em’ em” ovniil:w- theorists have not done so either, but ~ is miothe; nrat”-f. Ironically, just as Skinner criticizes Tr’Mm s e. ”mr;adam. of problem solving in terms of ‘tria ” err. m b&’. zrinrc “I cognitive theorists, can criticize Skinner Fiona ” Skirr- gets so involved in predicting specific re’” tin f r c arar r’l ever) gets around to explaining why IXirxns rrcrtirr r le”. simultaneously to. perform enfire cfas” of ”xr crue”’’ responses (e. g. Scandura 1970; see Scandu” let. Eforarerie. . if early papers on the relative merits of ” as”’cntril”i v”rb rules av the basic unit ofbehanoal arra”” bv. Sxppr, Arlbr Scandura, and others). For another’thin” its cr”-”cued Thordike’s trial and error explanations derive Gom ne pi” results based on multiple measures of d’ent srh).c. ” arri a: variety of problems. Yet, Skinner’s go+1 cf precrtirr pnroa- bilities of individual problem-solving resp)oses (”’ to mx knowledge has never been done successZv v W sicntriilM problems) is also limiting. Some confrontm:an czgkirtirv= tf’ried hists have shown how individual ins”ces of exrrpp?ex frurnan problem solvingcan be explained deter)rrmirtczlx (e ” Scvtvcj. ’ & Simon 1972; Scandura 1971;
19li) xfereov~ Sarrdirxra (1971; 1977) has demonstrated empirica]v that rirncfeckkafeae2 'conditions it is also possible to predict such grrcofermv-Mvtrtrg behavior (i. e. individual behavior of a specific pmibkmstar earrf ri time. There are many further instances of thns snrt, brnt frre p:esexrt purposes let me just mention two that strvkea per~"A mte In. attempting to reconcile the problem of moniviatia konn a strnc- tural learning'' perspective (i.e. what rajar) to arise wham mitre than one might do), I originally considered )m~ selection principles, including behavioral contrrr ~ tBrat would explain selection in aff cases. I ev~ ob~ erro~ evidence that strongly favored one such. Inv~ fneas (Scardurra 1971; 1973). The probability of sefectmg the moire specru= (fess general) of two applicable rules Nm ger exffy mrrrch hirsh+. . . than that of selecting the more general one (e. g.,Scaxduna 1963). Shortly thereafter, however, it becaire apparent that a moire deterministic alternative was ne'cessarv (Scar)(rd)na lPi~ 1``i ironically, the introduction of higher-orders l~ xxes fine this purpose not only increased the interreal coresicenev of tfire overall theory but provided a cognitive expo@~" fair hefrari= ioral contingencies (e. g. Scandura 197 3, pp. "o`` cl) Another parallel in my own work whrch caiore to xrrnd. nn reading Skinner's paper corresponds ~``to Skae= smajior . · distinction between rule-governed (pmceaural) ame conmorger~ cy-shaped (structural) behavior. Altbo~ I c4 nor4 eann toharuve convinced other cognitive theorists of tfius, recext rr`` strongly suggests that the very same mn9uan~ tbt urrrcexies problem solving is also responsiblefor rdie proccess es erf autra- matrzation (e. g. Scandura 1981r Scandura * Sc-ra]una HS. Specificaily, procedures of the rules used W non`` to be gradually transformed via '``or~ rauh inrtn tfire structurally more complex rules (v)&cfr are ~(ed) arrrrC practice) characteristic of experts. r In eff~ prexduxa3 eorm- plexity, corresponding to Skinner's rule-grvvexne)d lefnzviir, rrs ~ transformed into structural complexity, cocre go``io Srfr~ ner's con tingency-shaped behavior Given 54ne= s ffrLfong emphasis on analyzing behavior qua behaviior(or)- is rrx supc``mt . that he has overlooked this potentially furcLoxnmx* retainedhip between the two types. NOTE 1. Scundura and Snndura (le&) result1``a ~ sviu~ ~ r)fear ' must be au tomatved before they cur serve to deEine ~``rde= srofi- lems, whi cb explains why it may be dioicuh tri trtue ~`` o`` of development in tasks at a higher level


This letter was excerpted from a report commissioned by the NIE (Scandura, 1984) on Independent research vs. Labs and Centers. I strongly proposed the former, but obviously did not succeed.


155. Scandura, J.M.  Comments on rules and higher order knowledge as they pertain to Stahl and his information-constructivist perspective.  *Journal of Structural Learning and Intelligent Systems*, 1992, 11, 247-54.


168B. Scandura, J.M. Role of instructional theory in authoring effective and efficient Learning technologies. Computers in human behavior, 12, 2, pp. 313-328


177. Scandura, J.M. Introduction to Compendium I. Technology, Instruction, cognition & Learning (TICL), 2003, 1, 2, 115-117.


190A. Clean galley copy before refining galley


0. .PDF Summary of SLT in Wikipedia .org

BOOKS AND MONOGRAPHS


MAJOR COMPUTER-BASED SYSTEMS

_*MicroSystem80 Tutorial Series (published by Borg-Warner Educational Systems):*

S-1. Critical Reading, Fall 1980 (8 disk system)
S-2. College Board Preparation, Winter 1980 (12 disk system)
S-3. Language Arts - Word Structure, Fall 1982 (8 disk system)

_*MicroTutor II Tutorial Series (published by Queue, Inc.):*

S-4. Tutorial Arithmetic, Fall 1982 (12 disk system for addition, subtraction, multiplication, division).
S-5. EZ Author & EZ Author Plus Authoring systems, Spring 1983.
S-6. SURPAS Authoring system with "intelligent" interface, Spring 1983.
S-7. Mosaic Tutorials (over 200 titles covering areas from pre-school perceptual skills to college preparatory science and mathematics).

72
Software Engineering

S-10. SoftBuilder, 1999-present
S-11. AutoBuilder, 1995-present

AuthorIT and TutorIT: Highly Adaptive and Configurable (Intelligent) Tutoring Systems
S-12. AuthorIT, An authoring system for building highly adaptive and configurable tutoring systems, 2003-.
S-13. TutorIT, an general purpose tutoring system, which takes any AST knowledge representation and provides customizable tutoring, 2003-.
S-14. A rapidly expanding series of dynamically adaptive (intelligent) and configurable tutoring systems covering a broad range of content in mathematics at the elementary, middle and secondary school levels along with a critical reading workbook series and isolated samples. These tutoring systems are ready for field testing, refinement and pilot testing in authentic educational environments.

PATENTS