Role of Relativistic Knowledge in Intelligent Tutoring

Joseph M. Scandura
University of Pennsylvania

Abstract — Understanding complex cognitive behavior necessarily involves interaction between two (or more) individuals. The teaching-learning process is just a special case. From the standpoint of an intelligent tutor (or human teacher) all student/learner knowledge is either assumed to be available or is judged relative to the teacher's model(s) of the content and/or skills to be acquired. In this regard, learner models play a central role in most contemporary approaches to intelligent tutoring. These models typically assign to the learner certain capabilities which are assumed by the intelligent tutor in administering instruction. If too much is assumed, however, it is impossible to determine where the student is going wrong, or how to correct the problem — unless the tutor model is enhanced to incorporate the assumed capabilities and these capabilities correspondingly are eliminated from the learner model. Research associated with the Structural Learning Theory shows that a common “goal switching” control mechanism is all that may safely be assumed with confidence. Maximum flexibility is achieved by assuming only this mechanism, along with basic encoding and decoding capabilities.

ESSENTIALS OF INSTRUCTIONAL THEORY

Any viable theory of teaching and learning must include, first of all, some way of specifying what must be learned, that is some way to represent knowledge. Second, any viable theory must elucidate the processes by which people use, acquire, and modify existing knowledge. Third, there must be some way to find out what individuals know at any given stage of learning, including a way to determine their initial knowledge. In addition, a fully adequate theory of teaching and learning must allow for the growth of knowledge over time as learners interact dynamically with a changing teaching environment.

During the past two decades considerable progress has been made in the aforementioned direction. Thus, considerable progress relevant to various aspects of the above has been made in such fields as artificial intelligence (Bobrow & Collins, 1975; Minsky & Papert, 1972), individual differences measurement (Cronbach & Snow, 1977; Glaser, 1963; Hively, Patterson & Page, 1968; Scandura, 1964), and cognitive psychology (Anderson, 1976; Kintsch, 1974), as well as educational psychology per se (e.g., Gagne, 1962; Glaser & Resnick, 1972; Merrill, M. D., & Boutwell, 1973; Merrill, P. F., 1978; Rothkopf, 1972; Tennyson & Rothen, 1977).

Requests for reprints should be addressed to Professor Joseph M. Scandura, 3700 Walnut St., University of Pennsylvania, Philadelphia, PA 19104.
There also have been important developments in dealing with the instructional process as a whole (Landa, 1976; Pask, 1975), and with relationships to general systems theory. Specifically, significant progress has been made in understanding the interrelationships among content, cognition, and individual differences and in the way they interact over time as a result of instruction.

Global considerations, of course, necessarily play some role, even in the most prescribed research, as does actual human behavior in global systems-oriented theories. Nonetheless, the extent to which "top-down" considerations have influenced the former and the extent to which "hard data" have influenced theorizing about instructional systems have generally been quite limited.

Somewhat orthogonal to the above dichotomy has been the widely sensed gap between theory and practice (Scandura, Frase, Gagne, Groen & Stolurow, 1978). Typically, theories associated with the various academic disciplines have been perceived as having, at best, peripheral relevance to instruction. On the other hand, pragmatically generated teaching techniques and/or design principles have been largely devoid of theory.

To date, I am aware of only one theory which has seriously probed the "no- man's land" between these alternative views and concerns, that is Structural Learning Theory. This theory and a rather large body of supportive empirical research have been well documented in the literature, most recently and comprehensively in my books on Structural Learning and Problem Solving. I shall not attempt to survey this literature here. Rather, to provide a basis for comparison with contemporary theorizing in cognitive psychology, I shall emphasize essential historical, global, and methodological considerations with special attention to relationships to the instructional process.

Historically, development of the Structural Learning Theory was motivated by instructional considerations. Specifically, the goal of my very first piece of serious research (Scandura, 1962) was to help clarify the roles of expository and discovery modes of problem-solving instruction. What I found was that it is essentially impossible to obtain reliable results no matter how precisely one attempts to specify instructional treatments. More critical than how that information was imparted was when that information was given in relationship to what learners knew at the time. If presented too early, pupils not only were unable to use the information but also they gradually learned not to attend when presented with subsequent information. This research, incidentally, employed a combination of informal observation and analysis supported by precise operational definitions and experimental methodology not unlike that which characterizes much contemporary cognitive psychology.

Nonetheless, although certain analytical tools were used (e.g., the use of algorithms to represent what was to be learned), a major problem with this research was the inability to operationalize individual knowledge. Specifically, it was difficult to tie the phenomena being studied in with S-R and concept learning studies, or with the computer simulation studies of the day. Given what seemed to me to be an inadequate S-R language and unnecessarily cumbersome computer programs, I turned my attention in the early and mid 1960s toward the development of a simple but suitably general scientific language for theorizing about such phenomena.

Others during that period also were concerned with clarifying relationships between simple S-R and more complex kinds of learning. Rather than attempting to represent rule, problem solving, and other complex forms of learning as
complications of S-R learning, however, it seemed to me both more parsimonious and more useful to take the rule as basic and to explain simpler types as special cases (Scandura, 1967). (Equally important, this type of formulation appeared to be considerably more precise, thereby making it possible to avoid certain problems which arise, for example, in attempting to represent rules or principles in terms of concepts or associations.)

The Set-Function Language developed as a result of this work, the emphasis was on sets of observable input-output (stimulus-response) pairs and on rule (function) constructs needed to explain how outputs were to be generated from the inputs. Specifically, rules were characterized as triples, each rule having: (a) a domain, or set of conditions to be satisfied by inputs, (b) a range, or set of anticipatory conditions characteristic of the outputs the knower expects the rule to produce, and (c) an operation or procedure (algorithm) which, when applied to inputs in the domain, generates a unique output (Scandura, 1970).

My students and I used rules, so defined, during the 1960s in the analysis and empirical study of a wide variety of rule-based phenomena ranging from simple to complex. (Many of these studies are summarized in Scandura, 1969.) This characterization was subsequently adopted in research by a number of influential educational psychologists (Merrill, M. D. & Boutwell, 1973; Merrill, P. F., 1978) and apparently is now widely accepted.

**OVERVIEW OF THE STRUCTURAL LEARNING THEORY**

The Structural Learning Theory is a natural extension of this early work and provides a unifying theoretical framework within which to view the teaching-learning process. In fact, the theory is not really a (specific) theory at all but rather the Structural Learning Theory defines a class of theories.

The Structural Learning Theory, however, is not simply a scientific language. As I shall show below, very definitive assumptions are made about how and why people behave as they do. Furthermore, numerous specific realizations of the theory have been detailed and empirically tested to good effect (Scandura, 1977; Scandura & Scandura, 1980).

**Universal Characteristics of Human Cognition**

**Control mechanism.** Control mechanisms are about the most important universal characteristics. They are essential in all information processing systems, whether man or machine, and serve to tell the learner which rules to use and when to use them. Whereas all complete information processing theories make a distinction between processes (rules) and control, control in most cases either plays a subordinate role (Newell & Simon, 1972), or is distributed among a variety of different control mechanisms whose coordination, in turn, is often left unspecified (Pascual-Leone, 1970).

In contrast, in structural learning theories control mechanisms have been subjected to direct empirical study (Scandura, 1973, 1974, 1977). It is well known, for example, that people are not always able to solve problems, even when they know all the necessary components. What has not been so clear, however, is why this is so. Are successful persons somehow more capable than the others? Or, do they simply know something that unsuccessful people do not?
The explicit introduction of higher-order rules helps provide answers to these and a wide variety of other questions. As observed above, for example, higher-order rules provide an explicit basis for explaining, predicting, and/or controlling behavior involving analogy, generalization, problem definition, rule selection, etc. (Scandura, 1973, 1977, 1981).

However, even introducing higher-order rules does not provide a sufficient basis for explaining individual behavior. A complete theory must include (control) mechanisms which explain how and why various rules are used in particular situations. In this regard, the Structural Learning Theory postulates a simple goal switching control mechanism that makes minimal assumptions about the processor, assumptions that appear to be generalizable to all people. This mechanism simply makes precise what has been implicitly assumed for many years. If a person does not know how to solve a given problem, but still wants to solve it, then he will automatically turn his attention to finding some way to do it.

More specifically, given a problem the human information processor is assumed to first check to see if a solution is directly available. If not, the processor is assumed to search through (the ranges of) his available rules to see which, if any, might solve the problem. (A rule is a potential solution rule if its range “matches” the problem goal and its domain includes the problem given.) If a unique rule is found, then the rule is applied and the output is tested to see if it solves the problem. If there are no potential solution rules, then the search takes place at a still deeper level. In this case, control directs a search for (higher-order) rules that generate potential solution rules.

If such a higher order rule is found, then it is applied; the newly generated rule is added to the set of available rules and the search reverts to the next lower level. The augmented rule set is then checked as before; only this time the newly derived (potential) solution rule is available. In general, whenever there are no rules (or more than one rule) which apply at a given level of search, control moves to a still deeper level. Conversely, whenever a match is achieved (so that a rule is applied), control reverts to the preceding level.

The simple “goal switching” mechanism is hypothesized to be common to all humans and to govern all cognition irrespective of the specific knowledge involved. Consider, for example, the problem of rule derivation, of how individuals derive new solution rules for solving new problems they have never seen before. According to the Structural Learning Theory, rule derivation takes place as a result of applying various higher-order rules to other rules. These higher-order rules may serve to combine component rules, to generate analogous rules, to generalize given rules, etc.

To make things concrete, suppose a child knows rules for converting yards into feet (multiply by three) and for converting feet into inches (multiply by twelve) and that he is asked, “How many inches are there in two yards?” Clearly, this problem can be solved by combining the two available rules, the rule for converting yards into feet and the rule for converting feet into inches. But how does the child know how to combine the given rules? Knowing component rules is surely not logically equivalent to knowing when and how to use them.

A basic assumption in the Structural Learning Theory is that new rules are derived by application of certain rules to other rules. In the present case, a child might be expected to succeed if he knows a higher-order rule that operates on pairs of rules of the form, A → B, B → C (i.e., like but not limited to those above), and
combines them to form composite rules of the form \( A \rightarrow B \rightarrow C \), in which the components are performed in sequence.

While knowing both requisite higher-order rules and requisite lower-order rules is a necessary condition for solving problems this is not sufficient. In order to effectively use available rules to derive solution rules and to solve problems, some type of control mechanism is needed to determine when each rule is to be used and how. The question here is whether the above control mechanism is sufficient for this purpose and, if so, whether this mechanism is available to all human beings as postulated. Although the former question is strictly analytic in nature (Scandura, 1977), it seems best here to consider both questions in the context of empirical evidence.

I must caution that this simple control mechanism is an idealization and applies only in situations where processing capacity is not a factor, and specifically where all of the requisite higher- and lower-order rules are learned perfectly and are active in “working” memory (Scandura, 1971, 1973, 1977). Perhaps surprisingly, however, this limitation has not proved to be as critical as one might expect. Empirical support has been strong, although not deterministic, even under “real-world” conditions. Ehrenpreis and Scandura (1974) for example, found that higher-order (as well as lower-order) rules underlying a college course for teachers could be identified in a systematic manner and that instruction on such rules had a highly positive effect on prespecified kinds of “far transfer.” Furthermore, the degree of transfer was directly related to the degree to which the test conditions approximated the ideal (Scandura, 1977).

When used in conjunction with appropriate kinds of higher- (and lower-) order rules, the goal switching control mechanism provides an adequate basis for explaining, predicting, and controlling a wide variety of behavior. These include: (a) solving analogy problems, (b) generalizing given rules, (c) motivation (rule selection), (d) problem definition (sub-goal formation), (e) automatization, and (f) rule retrieval. For details and related empirical support the interested reader is referred to Scandura (1973, 1977).

Assessing Individual Knowledge

In contrast to universal cognitive constraints, specific knowledge is assumed to vary over individuals. The theory shows how prototypic competence (i.e., competence prototypic of given populations) may be used to operationally define the knowledge available to actual individual members of such populations. Specifically, the theory tells how, through a finite testing procedure, one can identify which parts of to-be-taught rules individual subjects know. The rules in a very real sense serve as rulers of measurement and provide a sufficient basis for the operational definition of human knowledge (Scandura, 1977).

Clearly, this is not the place to detail how this may be accomplished in general. For one thing, discussing the way process atomic rules are operationalized gets one deeper into issues of representation than is necessary here (for details, see Scandura, 1977).

For present purposes, it will be sufficient to consider the process of assessing individual behavior potential with respect to rules of competence represented in terms of behaviorally atomic components. Even here I consider operationalization only with respect to single rules rather than with respect to sets of competence rules con-
sidered collectively. Consider a rule (procedure/algorithm) for subtracting numbers. This rule is broken down into atomic components (i.e., steps that are so simple that each individual in the target population may be assumed able to perform each step either perfectly or not at all). In line with my previous discussion, it is worth emphasizing that atomic units relative to one population may not be atomic units with respect to another (e.g., less sophisticated) population.

Because success on any path of a rule depends on success on all atomic components, each path through the rule also acts in atomic fashion. Furthermore, there are only a finite number of behaviorally distinct paths; I do not distinguish paths according to the number of repetitions of loops, because the same cognitive operations and decisions are required regardless of how many times a given loop is traversed in carrying out a given “cognitive computation.”

Collectively, the paths of the subtraction rule impose a partition on the domain of column subtraction problems: That is, they define a set of distinct, exhaustive, and homogeneous equivalence classes of subtraction problems such that each problem in a given equivalence class can be solved via exactly one of the paths.

The fact that each path is associated with a unique subclass of column subtraction problems makes it possible to pinpoint through a finite testing procedure exactly what it is that each subject knows relative to the initial rule. It is sufficient for this purpose to test each subject on one randomly selected item from each equivalence class. Success on each item, according to my atomicity assumptions, implies potential success on all other items from the same equivalence class, and conversely for failure.

Individual knowledge (or behavior potential), in effect, may also be represented in terms of rules, specifically in terms of subrules of given rules of competence. Notice in this regard that the knowledge attributed to different individuals may vary even where only one rule of competence is used to assess behavior potential. For example, if a person succeeds on only one path and fails on the others, then his knowledge would be represented by that path (which is a subrule of the initial one). Where two or more paths are involved, a combination of the paths would be used to represent individual knowledge.

Fortunately, the above discussion is not just a theoretical exercise. A significant amount of supporting data has been collected over the past several years on a variety of problem domains, with subjects ranging from preschool children to Ph.D. candidates. Given a class of tasks, the general form of each study went as follows: (a) one or more prototypic rules were identified which were both adequate for generating solutions to each of the tasks and compatible with the way a knowledgeable or idealized (prototypic) member of the target population might be expected to solve them; (b) these rules singly and/or collectively were used to partition the class of tasks into equivalence classes; (c) subjects in the target population were tested on two items (tasks) from each equivalence class; and (d) performance on one item from each equivalence class was used as a basis for predicting success or failure on the (second) item.

With highly structured tasks run under carefully prescribed laboratory conditions, it has been possible, given performance on initial items, to predict performance on new (second) items with over 96% accuracy (Scandura, 1973; Scandura & Durnin, 1978). When testing took place under ordinary classroom conditions, with the subjects run as a group, the predications were accurate in about 84% of the cases (Durnin & Scandura, 1973).
There is one further major advantage of the structural approach to assessing behavior potential: The approach makes it possible to identify precisely not only what individuals can and cannot do, but also what the learner does and does not know relative to the particular rules involved. A simple basis for instructional decision making follows directly: Assume the paths the learner already knows and gradually "build in" those that he or she does not.

In summary, it would appear that any viable theory of performance testing must take into account underlying competence. Not only do rules of competence (associated with populations of subjects) provide a basis for measuring individual knowledge and for providing remedial instruction but they also provide an explicit basis for selecting appropriate test items.

Furthermore, because the appropriate level of rule representation varies directly with population sophistication (and desired level of behavioral detail), it is often practicable to analyze even complex task domains (at a level of analysis that is sufficient for assessing the behavior potential of individuals in the population).

The interested reader is referred to the literature for information regarding hierarchical relationships among paths and the conditional testing this makes possible (Durnin & Scandura, 1973; Scandura, 1973; Scandura & Durnin, 1978), the consolidation of knowledge (Scandura, 1977), a possible basis for assessing sentence production capabilities (Carroll, 1975), the use of sets of rules for assessment purposes (Scandura, 1977), and the assessment of skilled performance where response measures (e.g., latencies) more refined than success/failure are required (Scandura, 1973).

**The Learner**

Prototypic competence is not the same as rules of knowledge which characterize individual behavior potential. It is assumed in the Structural Learning Theory that what an individual does and can learn depends directly and inextricably on what he already knows. More particularly, as shown in Figure 1, it is assumed that human cognition may be adequately characterized in terms of: (a) universal characteristics of the human information processor and (b) individual knowledge. As described above, the latter is operationally defined in terms of rules (of both high- and lower-order) relative to prototypic competence (prototypic competence being associated with given problem domains and learner populations).

Clearly, instruction is concerned primarily with individual knowledge. From an instructional point of view, universal characteristics are best thought of as those aspects of human cognitive functioning that are inherent to man generally, they need not, and indeed, in some cases cannot be taught.

**The Observer**

Any attempt to understand human behavior necessarily depends on the observer. The observer determines what in the behavior is important (or even noticed), and consequently what is to be explained, predicted, or controlled. Pask (1975) was one of the first to emphasize the importance of this reality in instructional theory. The dependance of knowledge representation on the perspective of some observer also has played a long and perhaps more operational role in my own Structural Learning Theory (SLT) (Scandura 1970, 1971, 1973, 1977, 1980, 1985): This view is crucial in assessing individual knowledge (Durnin & Scandura, 1973; Scandura,
Overview of Structural Learning Theory

Figure 1. Overview of structural learning theory.

1971; Scandura & Durnin, 1978). The theories themselves, however, as well as the underlying cognitive constructs and methodologies differ greatly. Pask's Conversation Theory, for example, is inherently stochastic whereas the SLT has a deterministic foundation.

The Relativistic View of human knowledge, which this approach requires, has important implications for the design of intelligent tutoring systems (as well as for understanding the teaching/learning process in general). In this summary, we direct the reader to relevant research in structural learning, and make some important implications of the relativistic view more explicit.

Understanding complex cognitive behavior involves an interaction between two (or more) individuals. One can only understand the other's behavior in terms of the observer's perspective(s). Put somewhat differently, how one individual assesses and/or reacts to another depends on the observer's perspective (against which the other individual's behavior is judged). (Mutual understanding, therefore, can be achieved only to the extent that the two participants share a common perspective through which they may communicate.)

As shown in Figure 1, the teaching/learning process is just a special case of the above in which the teacher is assumed to fully understand the content to be taught (or, in less than ideal cases, to determine the learning agenda).

Having adopted this point of view, the teacher theorist must decide how to represent student knowledge. Consider the possibilities. From the perspective of an observer, the learner is essentially a black box as shown in Figure 2.

The learner is observed to respond in various ways in given environmental situations. To explain such behavior from an information processing perspective, the
1. COGNITIVE THEORY

- Problem: Devise a cognitive theory which is also compatible with requirements of instructional systems.

- Basic Question: How to explain cognitive behavior/learning with respect to a given task domain

"Black Box" Alternatives

<table>
<thead>
<tr>
<th>INPUTS</th>
<th>OUTPUTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. encoding/decoding capabilities</td>
<td></td>
</tr>
<tr>
<td>b. structures/procedures</td>
<td></td>
</tr>
<tr>
<td>c. universals (e.g., control mechanisms)</td>
<td></td>
</tr>
</tbody>
</table>

- "Black Box" alternatives are formally interchangeable

Figure 2.

observer can call on encoding/decoding mechanisms, cognitive structures and processes which operate on such structures, and/or control mechanisms which determine which processes are to be used and when.

From a mathematical, or purely computational point of view, explanation can be assigned to any one or more of these classes of constructs. Furthermore, within any one class of constructs, the theorist can choose from a variety of possibilities. Frames (Minsky, 1975); production systems (Newell & Simon, 1972), or SLT rules and higher-order rules (Scandura, 1971, 1987), for example, are all computationally equivalent. As noted below, however, these constructs are not interchangeable insofar as human behavior is concerned.

SUMMARY AND DESIGNING INTELLIGENT TUTORS

In designing an intelligent tutor the learner may be characterized in two basically different ways. First, encoding/decoding, structures/processes, and/or control mechanisms may simply be assigned to the learner. These assumptions constitute the learner model. This model is used heavily by the tutor to interpret the learner’s observed behavior; the validity of the student model, however, may not be determined by the tutor.

Alternatively, the tutor may have in mind (i.e., be ascribed) some characterization of the content and/or skills to be acquired. In this case, learner behavior is judged in relation to this characterization. Indeed the behavior of all learners in the targeted population is judged relative to the same standard.

Most contemporary approaches to intelligent tutoring fall in the first category (e.g., Anderson, 1984, unpublished proposal). These models typically assign to the learner certain capabilities which are assumed by the intelligent tutor in administer-
ing instruction. Where knowledge is viewed as relative, the learner’s capabilities are inferred relative to the teacher’s idealized perspectives (e.g., Brown & Burton, 1978; Scandura et al., 1973, 1977, 1986, 1988a, 1988b) on what is to be acquired.

The representation of knowledge in terms of SLT rules and higher order rules is especially useful in this regard. Such rules are more general than individual production rules, falling somewhere between them and production systems in comprehensiveness. SLT rules play a central role in the Structural Learning Theory, and involve a domain, range, and restricted type of procedure. (For a historical view of their evolutions see Scandura, 1970, 1973, 1977, 1987.)

A large body of research and theory demonstrates that SLT rules provide an explicit basis for identifying efficient sets of test items for assessing knowledge relative to given rules as well as a precise and natural means of representing the knowledge available to individual learners. For details see Scandura (Durnin & Scandura, 1973; Scandura & Durnin, 1978). It is important in this regard to emphasize that the relativistic view provides an explicit means of assessing individual knowledge and for guiding student learning.

Adaptive instruction is not possible where the learner is simply assumed to have certain capabilities. In particular, if too much is assumed, it is impossible to determine where the student is going wrong or how to correct the problem. The only way this can be accomplished is by enhancing the tutor’s model of what is to be acquired, eliminating corresponding capabilities from the student model, and using the enhanced tutor model to assess student capability.

A key question then is what, if anything, may safely be assumed of the learner? Of what should the learner model consist? The Structural Learning Theory provides a provisional answer which is highly consistent with a body of deterministic research (Scandura, 1974, 1977, 1981). This research shows that a common “goal switching” control mechanism is all that may safely be assumed with confidence. Human beings, from preschool children through adults, all act as if they come “wired in” with some such mechanism. It need not be learned or taught. The same cannot be said for cognitive structures/processes, and explicitly not for the kinds of inferencing capabilities that are commonly assumed in contemporary expert system models (e.g., Scandura, 1971, 1973, 1974).

Where such structures and/or processes are simply assumed, the intelligent tutor loses the ability to assess the knowledge of individual students, or to assist in overcoming misunderstandings. The tutor loses flexibility.

In general, structures and processes correspond to that which is to be taught and/or which may be learned. They are properly viewed from a relativistic perspective. “Goal switching” control, on the other hand, refers to a universal which may safely be assumed available to all human beings. It properly belongs in the intelligent tutor’s “learner model.”

Finally, observe that the inclusiveness of encoding and decoding capabilities varies inversely with the explicitness of the associated structure/process representations. This follows directly from my work on structural (cognitive task) analysis (Scandura, 1977, 1982, 1984). Given a class of tasks to be mastered, structural analysis (SA) is used to represent the rules (and higher order rules) which collectively constitute a model of what is to be learned. In this regard, structural analysis works in a top down fashion. At each stage of such an analysis, certain assumptions must be made about corresponding encoding and decoding capabilities. The more detailed the analysis, the less about encoding/decoding that must
be assumed. Consider the arithmetic algorithms. In this case, the ability to encode (i.e., to perceive) written numerals would ordinarily be assumed. In principle, however, further analysis might show how the numerals are constructed from more basic elements, such as straight and curved lines. The point is that encoding and decoding abilities of some sort are necessarily assumed in any teaching/learning situation. Such abilities correspond to what in educational jargon has traditionally been referred to as "readiness."

REFERENCES


