AD-A099 209   NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN D--ETC F/G 5/9
                CONFERENCE PROCEEDINGS: APTITUDE, LEARNING, AND INSTRUCTION, VO--ETC(U)
                JAN 81   R E SNOW, P FEDERICO, W E MONTAGUE
UNCLASSIFIED   NPRDC/ONR-TR-81-5-VOL-2   NL

1 OF 4
AD
A099209

npdc

NAVAL PERSONNEL RESEARCH AND DEVELOPMENT CENTER

AD A099209

NPRDC/ONR TR 81-5   JANUARY 1981

LEVEL

CONFERENCE PROCEEDINGS:
APTITUDE, LEARNING, AND INSTRUCTION

VOLUME 2

COGNITIVE PROCESS ANALYSES OF LEARNING
AND PROBLEM SOLVING

DTIC FILE COPY

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

onr

81 5 21 002

CONFERENCE PROCEEDINGS:
APTITUDE, LEARNING, AND INSTRUCTION.
VOLUME 2. COGNITIVE PROCESS
ANALYSES OF LEARNING AND PROBLEM SOLVING

Richard E. Snow
Pat-Anthony Federico
William E. Montague
(Editors)

Reviewed by
John D. Ford, Jr.
and
Marshall J. Farr

Released by
James F. Kelly, Jr.
and
Albert J. Baciocco

Navy Personnel Research and Development Center
San Diego, California 92152

and

Office of Naval Research
Arlington, Virginia 22217

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
NPRDC/ONR-TR 81-5 —	AD-AC99209	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
CONFERENCE PROCEEDINGS: APTITUDE, LEARNING, AND INSTRUCTION. VOLUME 2. COGNITIVE PROCESS ANALYSES OF LEARNING AND PROBLEM SOLVING	Conference Proceedings
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Richard E. Snow Pat-Anthony Federico William E. Montague (Editors)	N00014-78-C-0022 NR-154-419

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Navy Personnel Research and Development Center San Diego, California 92152 Office of Naval Research Arlington, Virginia 22217	61152N ZR000-01-042-06.01

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Navy Personnel Research and Development Center San Diego, California 92152 Office of Naval Research Arlington, Virginia 22217	January 1981
	13. NUMBER OF PAGES 336

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Proceedings of the conference were previously published as two volumes by Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Aptitude	Problem Solving
Learning	Information Processing
Instruction	Cognitive Representation
Cognitive Process Analyses	Adaptive Instruction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Summary of activities at a conference on Aptitude, Learning, and Instruction held March 1978 in San Diego, California. This volume consists of 13 formal papers, including formal comments by discussants.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

FOREWORD

This technical report is the outgrowth of the proceedings at a conference on Aptitude, Learning, and Instruction jointly sponsored by the Navy Personnel Research and Development Center (NAVPERSRANDCEN) and the Office of Naval Research (ONR). The conference was organized by the editors and held in San Diego in March 1978. It was funded out of NPRDC program element 61152N, task area ZR000-01, work unit 06.01 (Instructional Psychology), and ONR grant N00014-78-G-0022, work unit NR-154-419, to Stanford University.

The intent of this conference was to bring together outstanding individuals whose research reflects the latest theoretical thinking about cognitive processes in aptitude, learning, and instruction. Presentations by participants, combined with formal comments, provided a "state-of-the-art" summary of the field and identified directions for further research and development in and implementation of Navy instruction and training.

JAMES F. KELLY, JR.
Commanding Officer
NAVPERSRANDCEN

JAMES J. REGAN
Technical Director
NAVPERSRANDCEN

ALBERT J. BACIOCCO
Chief of Naval Research

GLENN L. BRYAN
Director, Psychological Sciences
Office of Naval Research

DISTRIBUTION LIST

These volumes are dedicated to the memory of

JOSEPH W. RIGNEY
Professor of Psychology
University of Southern California

whose career contribution to Navy personnel and training research cannot be overestimated. Still extending that contribution, he lost his life *en route to San Diego on* September 25, 1978, leaving the chapter found in volume 1 as one of his last works.

Contents

**Volume 1:
Cognitive Process Analyses
of Aptitude**

v

Volume 2:
Cognitive Process Analyses
of Learning and Problem Solving

Cont'd

x CONTENTS

Foreword

Marshall J. Farr
Office of Naval Research

This conference takes on a formidable task, that of trying to relate in a meaningful way the processes underlying human aptitude and intelligence to the cognitive aspects of learning and the real world of instructional practices. Trying to link aptitude in a systematic way to learning and instruction means a number of different things. It means confronting a Pandora's box of individual differences, as one tries to make sense out of human variability. It means having to bring together, as Cronbach pointed out in his 1957 APA Presidential Address, the psychometric approach of correlational psychology with the methodology of experimental psychology. It means a focus not only on both *organismic* and *treatment* variables but an equal concern with their *interaction*.

Aptitude, or even *ability*, is not a typical experimental psychology construct. I looked under the subject index of my 1954 Woodworth and Schlosberg *Experimental Psychology*, the edition to which many of the current crop of cognitive psychologists were exposed, and was not surprised to find no index entry for either *aptitude, ability*, or even *intelligence*. (In all fairness, the authors do acknowledge that organismic variables are of some consequence, with a listing of *individual differences* and a subheading *ability-performance* listed under *learning*.)

Although mainstream experimental psychology in about 1954 was relatively insensitive to the approach of correlational psychology, Kohler, one of the fathers of Gestalt psychology, recognized the issue in his 1947 classic, *Gestalt Psychology*. In discussing Fechner and his psychophysics work, he states:

> Today we can no longer doubt that thousands of quantitative psychophysical experiments were made almost in vain. No one knew precisely what he was measuring.

Nobody had studied the *mental processes* upon which the whole procedure was built. . . . When observing the energy with which able psychologists measure individual intelligences, one is almost reminded of Fechner's time. From a practical point of view, it is true, their work is obviously not without merits. It seems that a crude total ability for certain performances is actually measured by such tests. For, on the whole, the test scores show a satisfactory correlation with achievements both in school and in subsequent life. This very success, however, contains a grave danger. *The tests do not show what specific processes actually participate in the test achievements.* The scores are mere numbers which allow of many different interpretations [pp. 44–45, italics mine].

It is instructive to note how this quote by Kohler foreshadows the following notion expressed by Cronbach and Snow (1977) in the preface to their *Aptitudes and Instructional Methods:*

This state-of-the-art report has been more difficult to assemble than anticipated when we began in 1965. One reason is the breadth of the topic. To study scores on conventional ability tests is not sufficient, for the student's response to instruction is, in principle, conditioned by *all* his characteristics, including personality traits. It is necessary also to consider what Glaser calls "the new aptitudes," the specific intellectual-processing skills that are lost from sight in an aggregate mental measure [p. viii].

The Office of Naval Research (ONR) has long had an abiding interest in trying to link individual ability and aptitude differences with learning. As Federico discusses in some detail in Chapter 1 in this book, ONR sponsored a 1965 symposium at the University of Pittsburgh that focused on the ways in which people differed in their learning and how these ways might be measured as individual differences. (The proceedings were edited by Gagné (1967) and published as *Learning and Individual Differences.*) In this Pittsburgh conference, Melton concludes that there is an impressive consensus to the effect that we must consider individual-differences variables in terms of the *process* constructs of contemporary theories of learning and performance. And Melton concisely pinpoints the then-emerging zeitgeist when he states:

The most significant development in theoretical and experimental psychology in recent years is acceptance of the need for theoretical statements about processes or mechanisms that intervene between stimuli and responses. The argument is no longer about whether such intervening processes occur and have controlling effects on behavior, but about their defining properties, their sequencing, and their interactions [p. 240].

For about the last 6 years, ONR has been conducting a thematically oriented contract research program aimed, in large part, at developing the kind of broad theoretical framework necessary for a workable *process* interpretation of ap-

titude, learning, and performance. The papers in this collection are generally addressed to three broad areas that are central to these interests of the ONR Personnel and Training Research Programs. One area is concerned with individual differences in information processing, as revealed in simple laboratory or psychometric tasks. Whereas conventional measurement of abilities and aptitudes relies on the actuarial criterion of their success in distinguishing between high- and low-level individuals, the emphasis here is on the direct measurement of the component, basic information-processing operations that undergird the target abilities.

The second area focuses on the structural aspects of learning and performance, using tools and concepts from semantic memory theory to describe what is learned and how it is learned. And the third area is aimed at the management of instruction: It addresses itself to the kinds of research and instructional designs required for effective implementation of adaptive instruction.

ONR primarily supports mission-oriented basic research. The cosponsor of this conference, the Navy Personnel Research and Development Center (NPRDC), generally supports more applied research. That organization's support in this case demonstrates the strong practical implications it sees in this research. ONR and NPRDC are proud to have joined forces in what we believe will become a landmark work in the field.

<div style="text-align: right">

MARSHALL J. FARR
Director, Personnel and Training
Research Programs
Office of Naval Research

</div>

REFERENCES

Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist,* 1957, *12,* 671–684.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods.* New York: Irvington, 1977.

Gagné, R. M. (Ed.). *Learning and individual differences.* Columbus, Ohio: Merrill, 1967.

Kohler, W. *Gestalt psychology.* New York: Liveright, 1947.

Melton, A. W. Individual differences and theoretical process variables: General comments on the conference. In R. M. Gagné (Ed.), *Learning and individual differences.* Columbus, Ohio: Merrill, 1967.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology* (rev. ed.). New York: Holt, 1954.

Acknowledgments

14 Some Examples of Cognitive Task Analysis with Instructional Implications

James G. Greeno
University of Pittsburgh

As concepts and methods for the analysis of complex cognitive performance have developed, it has been increasingly attractive to think about their potential use in analyzing tasks that are used in instruction. The idea of applying concepts and methods of cognitive psychology to the analysis of instructional tasks is certainly not new. However, recent developments seem to have added a new dimension to the potential use of ideas from psychology and other cognitive sciences in the analysis and design of instruction. At least that seemed the case to me when I wrote a chapter entitled "Cognitive Objectives of Instruction" in 1974 (Greeno, 1976a). The organizers of this conference requested that I prepare a chapter on that same topic. Perhaps it will be useful in this context if I present a brief progress report of the work in which I have been engaged in the meantime. Much of this work is still in very early stages, and I apologize that this presentation is still more a research program than a set of results. However, some of the potential research that I sketched in 1974 has become actual research, and it may be useful to report the directions in which those ideas have developed during the short period since publication of that earlier article.

In my earlier paper, I discussed three kinds of instructional tasks: performing calculations in arithmetic, proving theorems and solving other problems in geometry, and understanding concepts in science. I did not intend to suggest then, nor do I now, that these topics exhaust the instructional domains in which cognitive science will contribute to instructional practice. For example, my short list did not include the analysis of reading skill, which probably is the domain in which the most has been accomplished in relating cognitive science and instruction. However, the three tasks that I discussed represent three important theoretical foci, and my research has progressed in ways that are relevant to those three

1

kinds of tasks. The analysis of calculating skills uses concepts in the theory of cognitive procedures. The analysis of knowledge acquired in geometry uses concepts in the theory of problem solving. And the analysis of understanding scientific concepts uses concepts in the theory of semantic schemata used in the process of understanding language. The work that I discuss in this chapter has involved analyses of geometry problem solving and arithmetic. Thus far, our studies of geometry have fit rather well into the research domain of problem solving. However, in our study of arithmetic we have become concerned with processes of understanding and semantic schemata, as well as with the procedural knowledge involved in computational skill.

PROBLEM SOLVING IN GEOMETRY

When the cognitive processes involved in an instructional task have been analyzed, the results can be viewed as a hypothesis about the knowledge that students acquire when they successfully learn the material given in instruction. The knowledge required for problem solving in geometry has been represented in a computer simulation model that I have given the name *Perdix*. The major source of empirical data used in developing Perdix was a set of thinking-aloud protocols that I obtained from a group of six ninth-grade students during a year in which they were studying geometry in a course. I interviewed the students individually approximately once each week throughout the year. At each session, the student solved a few problems, thinking aloud during the process. The protocols were recorded on audiotape, and the transcriptions are accompanied by diagrams that the students drew during problem solving. In developing Perdix, I have included procedures and structures of knowledge that enable the model to solve the problems that these students were able to solve, in the same general ways that the students solved the problems.

The form of Perdix is a production system, which means that each component of its knowledge is a pair consisting of a condition and an action that is performed if the condition is tested and it is found to be true. The productions that constitute Perdix's knowledge about geometry are in three groups, and these three groups of productions can be considered as three domains of knowledge required for students to solve the problems they are given in their study of geometry.

The three domains of knowledge required for geometry problem solving are the following:

1. Propositions used in making inferences.
2. Perceptual concepts used in recognizing patterns.
3. Strategic principles used in setting goals and planning.

The propositions needed in geometry problem solving are the familiar statements about geometric relations, such as "Corresponding angles formed by

parallel lines and a transversal are congruent"; or "If a triangle has two sides and the included angle congruent to two sides and the included angle of another triangle, the triangles are congruent"; or "If two angles are congruent, they have equal measure." Inferences based on this kind of proposition constitute the main steps in geometry problem solving. Geometry problems require students to show relationships between objects (e.g., "Prove that angle A and angle B are congruent") or to find the measure of an object, such as the size of an angle or the length of a line segment. Information is given in the problem in the form of segments or angles that are congruent, lines that are parallel, the measures of some angles or segments, and so on. Each step in solving the problem consists of an inference in which some new relation or the measure of some additional component is deduced from information that was given or that has previously been inferred. The problem is solved when this chain of inferences reaches the relation or measure that is the goal of the problem. Each of the inferential steps is based on one of the if-then propositions that the student knows. The antecedent condition of the proposition is found in the given information or the diagram, and the consequent relation is added to the problem situation.

The perceptual concepts needed for geometry problem solving include the patterns that are mentioned in the antecedents of many propositions. For example, the proposition "Corresponding angles formed by parallel lines and a transversal are congruent" mentions a pattern—corresponding angles. To use this proposition as a basis for inferring that angles are congruent, a student must look at a diagram and determine that the angles are in the correct positions relative to a pair of parallel lines and a transversal to be called corresponding angles.

The strategic knowledge that is needed in geometry includes knowledge of general plans that lead to the various kinds of goals that occur in geometry problems. For example, when solution of a problem requires showing that two angles are congruent, three alternative approaches are available. One approach is to prove that triangles containing the angles are congruent. A second approach is to use relations between angles that are based on parallel lines, such as corresponding angles or alternate interior angles. A third approach is to use relationships between angles whose vertices are at the same point, such as vertical angles, or angles that are formed by the bisection of another angle.

The design of the planning process in Perdix is similar to the one developed by Sacerdoti (1975) in his program NOAH (Nets of Action Hierarchies). As with NOAH, Perdix has knowledge of some general actions that it can perform. Knowledge about each general action includes the consequences of the action and prerequisite conditions that are required for the action to be performed. Perdix selects a plan for its current goal by checking the general actions that have consequences that achieve the goal. If the prerequisite conditions for one of the actions are present in the situation, Perdix adopts the plan of achieving the goal using that action. Then Perdix proceeds to try to execute the plan, using procedures that are also stored as part of the knowledge that Perdix has about the general action. These procedures can include the setting of further goals, which

may require selection of plans for their achievement, leading to a hierarchy of plans and goals for the solution of the problem.

Most of the features of the model for geometry problem solving have been developed by applying standard concepts in the recent literature on problem solving in psychology and artificial intelligence. There have been some interesting new developments required to simulate problem solving in this domain, which are discussed in other places (Greeno, 1976b, 1977, 1978). However, the main results have been obtained by examining the nature of the geometry task environment in some detail, studying the performance of subjects who are successful in performing the tasks that are used as a criterion of learning in that domain, and using concepts and methods that have been worked out in the general theory of problem solving to develop a theory about the knowledge structures and cognitive processes required for successful performance in the domain.

The result of this theoretical analysis can be considered as a model of the outcome of successful instruction for those aspects of the course that have been included in the analysis thus far. It has the advantage over purely rational task analysis in that it is generally consistent with the performance of human learners who did succeed in learning how to accomplish the criterion tasks. On the other hand, it does not characterize all the students who were in the course; some of them did not succeed in acquiring the necessary knowledge, and I do not have a model for their unsuccessful performance. Furthermore, to provide a really strong guide for instructional practice, we need to develop models of the process of acquisition in addition to models of the knowledge that is acquired.

On the other hand, a clear representation of the outcome of successful instruction probably can be useful. In the case of this geometry model, some interesting issues appear when the characteristics of the model are considered in relation to the content of the geometry curriculum as it is represented in texts for the course.

The theoretical analysis of geometry problem solving led to the conclusion that three main components of knowledge are required for a student to accomplish successfully the criterion tasks used in the domain. These are propositions for inference, perceptual concepts for pattern recognition, and strategic knowledge for planning and setting goals. Of these three, the first two are included explicitly in the instructional materials used in teaching. There is explicit presentation of the propositions that are used as the bases of inferences. When a new proposition is introduced, it is always explained carefully, and often a proof of the proposition is given. There is also explicit presentation of the perceptual concepts that are needed for pattern recognition. These are usually presented in diagrams, with exercises that emphasize the relevant features needed to identify instances of the concepts.

However, the components that I have been calling strategic knowledge are not represented explicitly in the instructional materials of geometry. The knowledge that is needed for planning and setting goals can be given an explicit characterization; indeed, it has such a characterization in the model I have been describing.

However, most references to that knowledge in texts that I have examined are relatively indirect, and my impression is that most teachers do not explicitly identify principles of strategy when they teach their students.

One interesting question is the following: If the instructional materials of a course do not include an important part of the knowledge needed to perform on criterion tasks, how do students acquire that knowledge? We know that many students must acquire strategic knowledge in some form, because they are able to solve problems that we are confident require strategic knowledge. It seems a reasonable conjecture that this knowledge is often acquired by induction. Texts include sample problems that present the steps of solutions in sequence, and teachers solve these examples during class, both before and after students have attempted to solve problems as exercises. The principles of strategic knowledge that must be applied in solving problems probably can be induced as general properties of the sequence of steps that students observe in example solutions. Knowledge that is induced in this way probably is implicit in nature. As with many intellectual skills, when we ask subjects to explain how they decided to perform in the way they did, the answers are not very coherent. Thus, the induced strategic principles appear to be in the form of tacit procedural knowledge involving things the learner is able to do, but not things the learner can describe or analyze.

It is not surprising that students' knowledge of strategic principles is implicit; it has only been in recent years that our scientific theories have included concepts that make it possible to describe strategic knowledge in explicit ways. In our general wisdom about problem solving, we attribute the skill some students show in problem solving either to their intelligence, to their motivation in the form of persistence, or, at most, to their ability to use very general, heuristic problem-solving methods. However, when current theoretical concepts and methods are used to analyze problem-solving tasks in a domain, the analysis indicates important strategic principles involving planning knowledge that is quite specific to the domain of problems that are analyzed.

A question about instruction arises in a rather obvious way. Now that we have discovered the nature of domain-specific strategic knowledge, should we include it explicitly in the materials of the geometry course? The argument for teaching strategies explicitly is quite straightforward. Strategic knowledge is part of the knowledge that students must acquire in order to solve problems in geometry. It is reasonable to try to teach that knowledge, like other knowledge of the course, in as effective a way as possible. Although it is possible that the unguided discovery method now used is more effective than a more explicit form of instruction would be, that seems unlikely in light of the research that has been done on discovery learning. The propositions for inference and concepts for pattern recognition in geometry are taught in the specific form in which they are required for geometry problem solving, and it seems reasonable to treat problem-solving strategies in the same way.

An argument against teaching specific problem-solving strategies explicitly

rests on the intuition that with the instructional methods we now use, students are required to generate the solutions of problems actively, and that this is a more valuable learning experience than would be provided if the instructional materials provided step-by-step guidance in methods of solution. The issue is an empirical one, albeit difficult to decide, and it would be desirable to have some empirical comparisons between instructional methods that are based on the two ideas. However, it seems certain that some methods of teaching strategic principles could be devised that would do more harm than good. It would probably not be helpful to most students to teach about strategies in an abstract way, with the strategic principles divorced from the context of problem solving in which they are used. Successful performance in solving problems probably should be considered as an intellectual skill, and it seems likely that successful instruction in problem-solving strategies will be based on principles of skill acquisition. Since we don't understand very much about the principles of skill acquisition, it is clear that we have a long way to go before we can make definite pronouncements about the relative merits of different forms of instruction in problem-solving strategies. It should be noted, though, that our present methods are quite analogous to the method of teaching swimming that consists of throwing a pupil into the water. That method is successful for some students, but it has obvious negative consequences for others.

Another possibility that I believe should be investigated is inclusion of explicit instruction about problem-solving strategies in the instruction that is given to mathematics teachers. I have not studied geometry teachers' understanding of problem solving in a systematic way, but the teachers with whom I have had conversations have quite an undifferentiated impression of the nature of skill in solving problems. In one meeting of teachers, when I described the strategic component of the problem-solving model Perdix, one teacher responded by asking whether what I was discussing wasn't just the students' intelligence. This teacher's view was that some students are better than others in applying mathematical ideas in problem situations, and that occurs because they are more intelligent. Another teacher proposed a motivational theory, in which failure in problem solving is caused by a lack of persistence. When difficulties are encountered, some students continue to work on the problem and may eventually find a way to make progress, whereas others give up as soon as the next move is not obvious. I am sure that both of these views have merit, but they are not the complete story. I am hopeful that teachers might be able to be considerably more helpful in facilitating their students' acquisition of problem-solving skills if their own understanding of the process became somewhat more sophisticated, with some concepts that refer to various components of the skill rather than being limited to very global concepts of intelligence and persistence.

I close this discussion of geometry by noting that the cognitive analysis of problem solving has not provided strong recommendations about how to teach the subject matter. It has provided a characterization of the knowledge that a

student should acquire, and some of the features of that knowledge raise issues about instruction that appear to be significant and interesting. It may be that specific recommendations about instruction would follow from a cognitive analysis of the learning process itself, but that is a point we will have to look into when we have some theoretical analysis of the learning process.

COMPUTATION AND UNDERSTANDING
IN ARITHMETIC

A second instructional task that we have been studying at Pittsburgh is elementary arithmetic. In this work we have begun with the basics—concepts of addition and subtraction that are taught in the first and second grades. As in the case of geometry, we are attempting to develop a model that represents the knowledge that students acquire if they are successful in mastering the material they encounter in arithmetic instruction.

Instructional objectives for primary arithmetic have two aspects: skill and understanding. In the domain of skill, students are expected to learn the basic addition and subtraction facts, so they can answer questions such as "What is 8 − 3 ?" or "What is 3 + 5 ?" or perhaps " 3 + ? = 8." In the domain of understanding, a variety of tasks are included in the curriculum, and they probably relate to rather different ideas about the nature of understanding. We have focused on the kind of understanding needed for children to be able to apply their knowledge of arithmetic in concrete situations, or in the semiconcrete situations that are presented in the form of word problems.

A substantial number of studies have analyzed processes for answering questions involving basic arithmetic facts. A considerable body of evidence now supports the idea that children use methods based on counting when they answer simple questions such as " 3 + 5 = ?" The method used by practiced subjects for addition is shown in Fig. 14.1. Evidence supporting this model has been obtained in studies by Groen and Parkman (1972) and by Groen and Resnick (1977). The evidence supports a model of subtraction that is similar in character. If the gap between two numbers in a subtraction problem is small—as in "8 − 6 = ?"—the child finds the answer by counting the size of the gap. If the number to be subtracted is small—as in "8 − 2 = ?"—the child uses a procedure that requires only a couple of counts; it might involve counting backward, but more likely, it involves some process of generating a small sequence of numbers near the larger term and then identifying the appropriate member of that sequence (Groen & Poll, 1973; Woods, Resnick, & Groen, 1975).

The main feature of these models is their procedural character. We should conclude from these analyses that the knowledge acquired by students when they learn the basic facts of addition and subtraction is a set of procedures that are based on their knowledge of counting. This implies that to understand the learn-
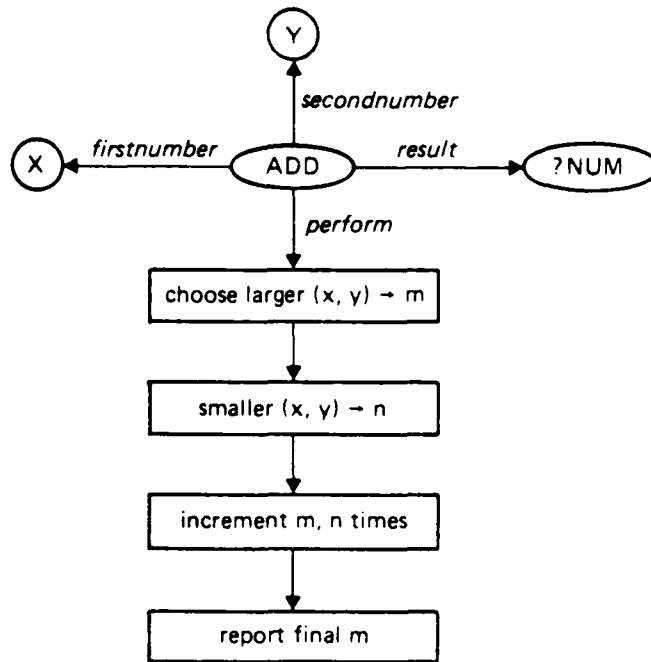
FIG. 14.1.   Procedure for answering simple addition questions.

ing of these procedures, we need to understand the nature of children's knowledge structures that are involved in counting. We have been fortunate to be able to collaborate with Rochel Gelman, who has conducted several studies of children's counting, focused on analyzing general principles that children understand and that affect their performance in counting tasks. This collaborative project, in which Mary Riley is also participating, has the goal of representing children's counting knowledge in a simulation model that we test by comparing its performance on various tasks with the performance that Gelman (1978) has reported. A long-term goal is the development of a simulation of learning, in which the knowledge structures that we identify for the counting tasks are transformed into knowledge structures that are capable of performing addition and subtraction.

The second aspect of knowledge about arithmetic involves children's understanding of concepts and procedures. In one test of understanding, children are asked to solve problems consisting of brief stories involving quantitative information such as the following: "Jill had three apples. Betty gave her some more apples. Now Jill has eight apples. How many did Betty give her?"

One project that we have begun is a simulation model of the process of solving arithmetic word problems (Heller & Greeno, 1978). A model of solving word problems has been developed previously, by Bobrow (1968), but our model is based on quite a different view of the process. In Bobrow's model, the main process was translation of the text into a set of simultaneous equations. This process of translation was based as much as possible on syntactic information,

and semantic processing occurred only when it could not be avoided. In our model, semantic processing is the main component of the understanding process. The system constructs a semantic network representing the information in the problem. To solve the problem, the system must select an arithmetic operation—for example, addition or subtraction. In our model, the operations are associated directly with structural representations, so there is no intervening process of constructing equations before the operation is chosen.

The processing of a problem by our system is based on a set of schemata that specify alternative structures of quantitative information. The analysis of these schemata has provided the most interesting result of our project thus far. The problems we have analyzed at this point all are solved by addition or subtraction of the numbers given in the problem. We have identified three distinct schemata that we believe are necessary and sufficient for understanding of all the problems that are solved by a single operation of addition or subtraction. I will refer to these three schemata as *cause/change*, *combination*, and *comparison*.

The cause/change schema is used for understanding situations in which some event changes the value of a quantity. For example, when Betty gives Jill some apples, there is a change in the number of apples that Jill has. The abstract schema that represents such situations is in Fig. 14.2. There are three main components. First, there is an initial quantitative state in which some object O is associated with some quantity P. Second, there is some action that involves a direction of change, increase or decrease, and an amount Q in the object O. Finally, there is a resulting state in which O has quantity R. For example, in the problem where Jill had three apples and got five more from Betty, the object is the set of apples in Jill's possession; the initial amount P is 3; the direction of change is increase; and the amount of increase Q is 5. The question indicates that the final amount R is unknown, and the problem is to find that quantity.

Figure 14.2 indicates that both addition and subtraction are related to the cause/change schema. This is because either operation can be required to solve problems in which the schema is used to represent the information. Consider two kinds of problems in which the unknown quantity is R, the amount in the final state, with numbers given as the values of P and Q. Addition is needed if the direction of the change is an increase, and subtraction is needed if the direction of the change is a decrease. For example, in the problem: ''Pat had eight flowers; he found three more flowers; how many flowers does Pat have now?'' P is 8, Q is 3, the direction is an increase, and the answer is found by adding 8 plus 3. In the problem: ''Pat had eight flowers; he lost three flowers; how many flowers does Pat have now?'' P is 8, Q is 3, the direction is a decrease, and the answer is found by subtracting 8 minus 3. Thus, both of the operations, addition and subtraction, are related to the semantic structure that represents changes in quantity, and the selection of an operation for solving a problem depends on the content that is found in a specific problem.

The second general schema for addition and subtraction problems is in Fig.
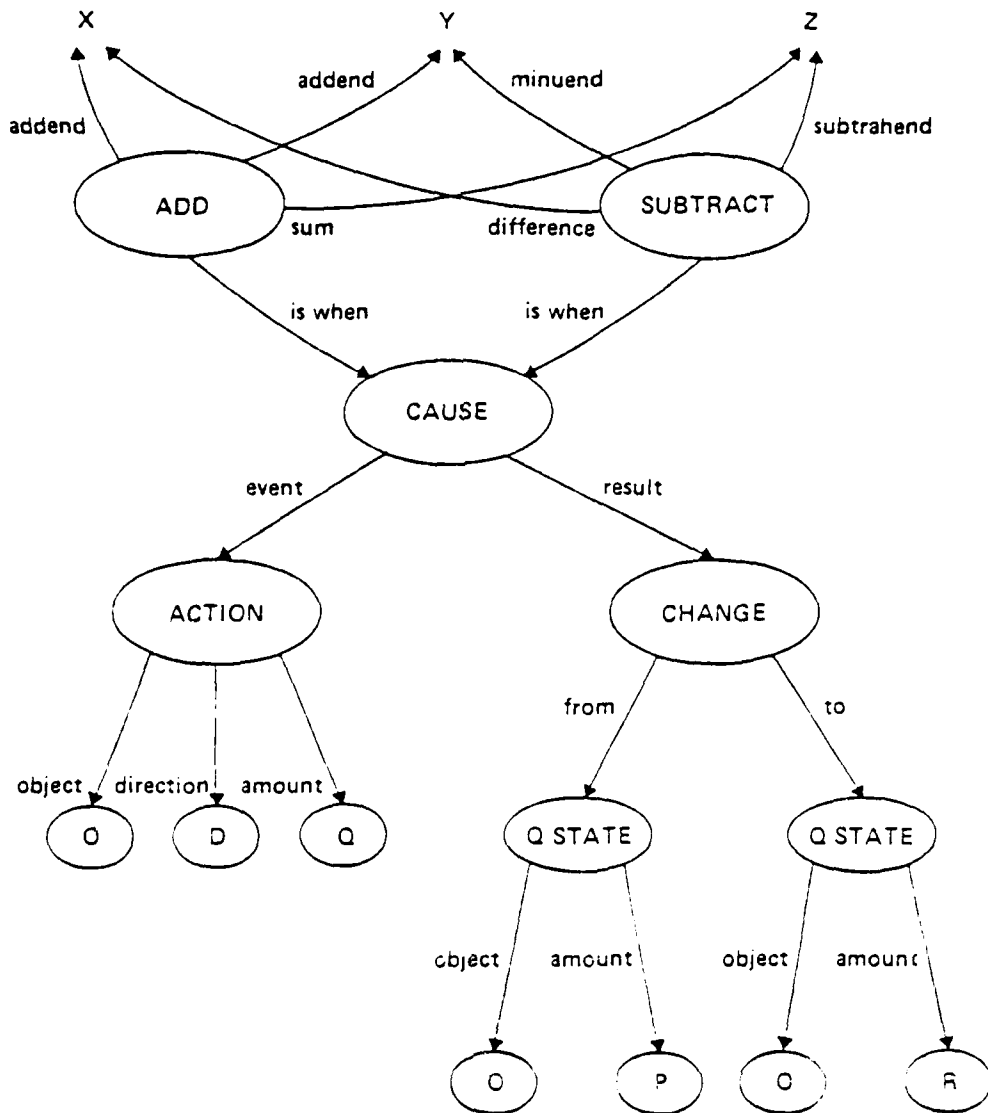
FIG. 14.2.   Schema for representing problems in which an event causes a change in quantity.

14.3. This schema is used to represent situations where there are two amounts, and they can be considered either separately or in combination. For example: "Sue has three apples; Betty has five apples; how many do they have altogether?" or "Sue has three apples; Betty has some apples; they have eight apples altogether. How many does Betty have?" The two separate amounts fill in the positions denoted by U and V in Fig. 14.3, and the combined amount fills the position denoted W. In this schema, the choice of an operation for answering a question depends on which of the three quantities is unknown in the question. If

the combined amount is unknown, it is found by adding the other two amounts. If one of the separate amounts is unknown, it is found by subtracting the known separate amount from the combined amount.

The third general schema for addition and subtraction is in Fig. 14.4. This involves two amounts that are compared and a difference between them. It would arise in a problem such as: "Sue has three apples; Betty has five apples; how many fewer apples does Sue have than Betty?" Betty's apples are the reference object 01, and their amount *J* is 5. Sue's apples are the comparison object 02,
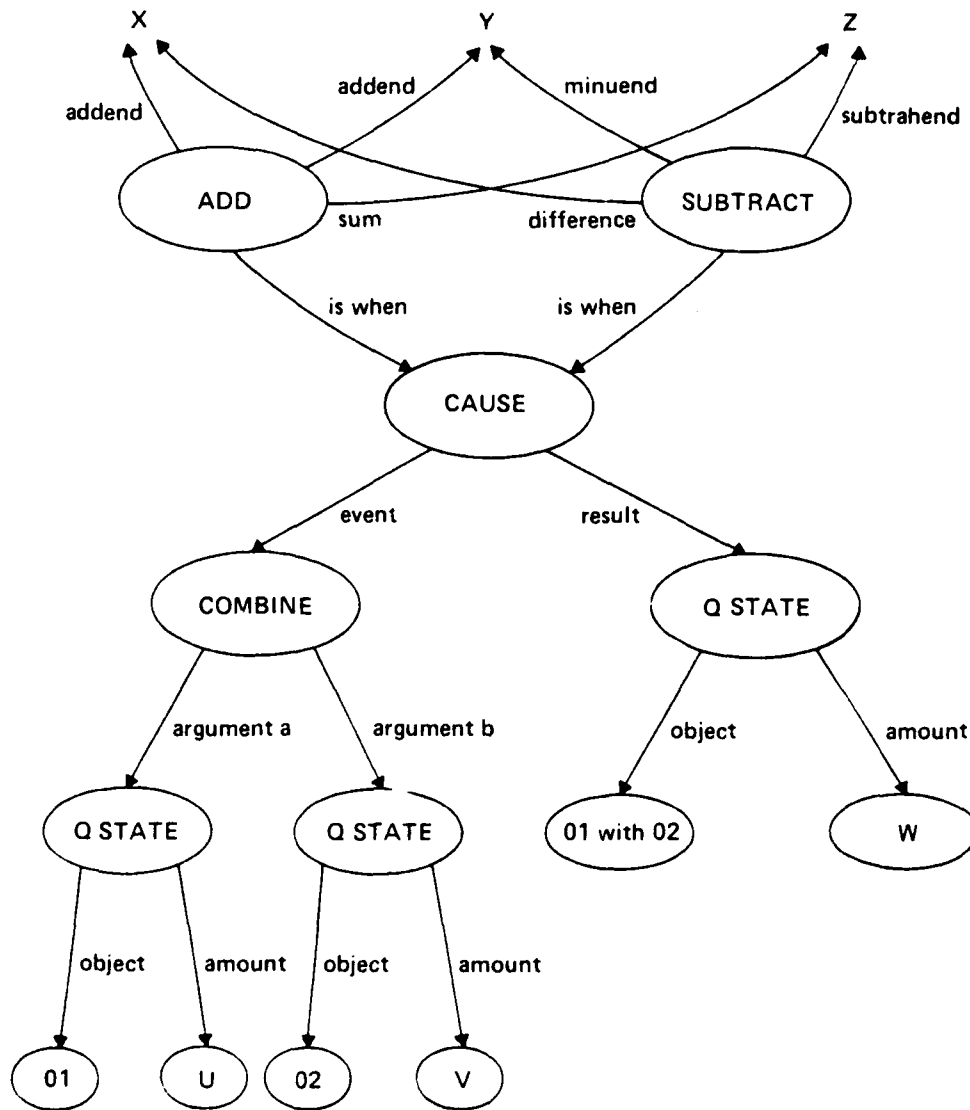
FIG. 14.3.   Schema for representing problems involving a combination of two quantities.
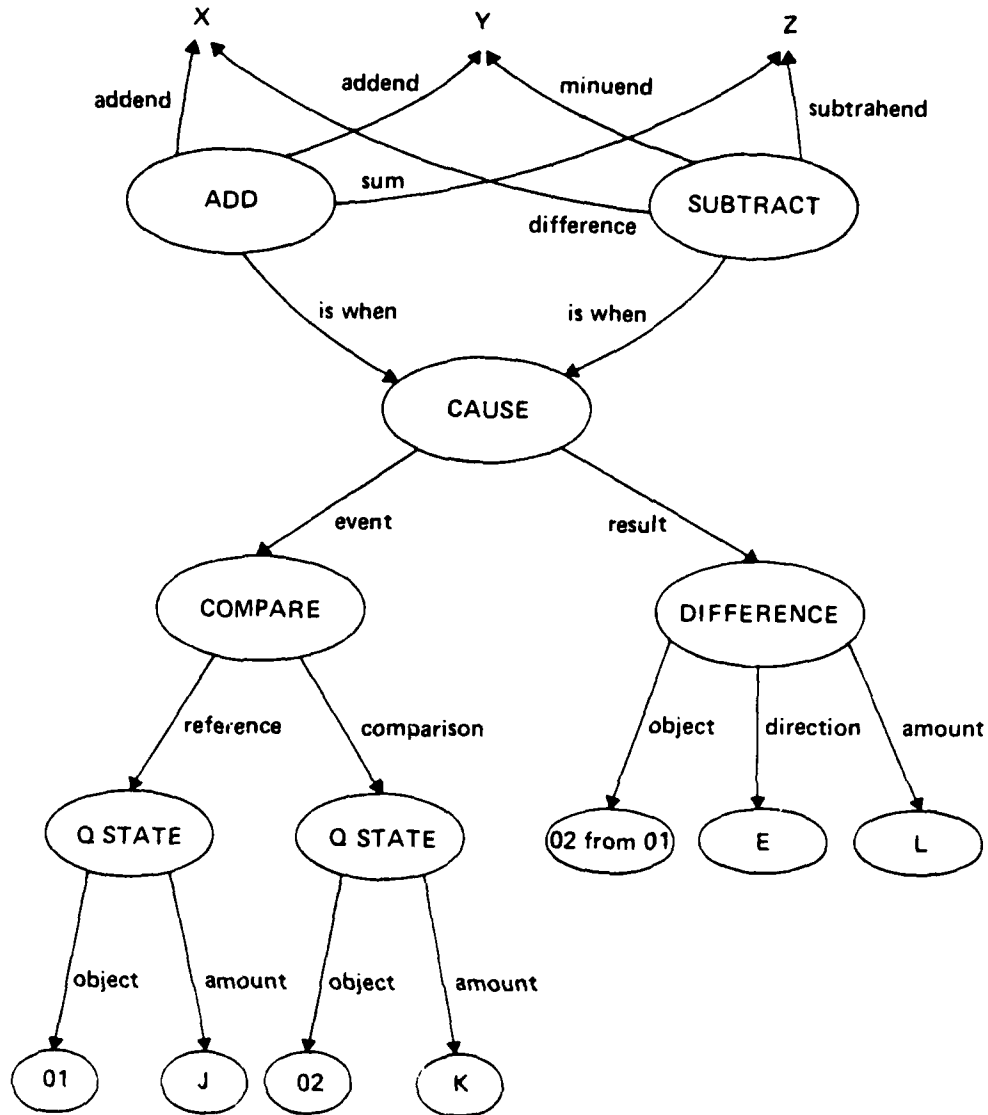
FIG. 14.4. Schema for representing problems involving a comparison of two quantities.

and their amount K is 3. The direction of the difference is fewer, and the amount of difference is unknown. Another problem that would be represented using this schema is: "Sue has three apples; Betty has five more apples than Sue; how many does Betty have?" In this case, the reference J, the number of Sue's apples, is given as 3; E, the direction of the difference, is given as more; L, the amount of difference, is given as 5; and K, the number of Betty's apples, is unknown. Notice that when the difference is unknown, the question is answered by subtracting J from K or K from J, depending on which is smaller. If the

difference is known, the question is answered by adding L to the known single quantity or by subtracting L from the known single quantity, depending on the direction given for the difference.

These three semantic schemata constitute three different meaning structures for addition and subtraction. I think it is appropriate to say that these arithmetic concepts are ambiguous. They have distinct and incompatible meanings. On the other hand, addition and subtraction are genuine abstractions in relation to the cause/change, combination, and comparison meaning structures. These, in turn, are relatively abstract themselves. For example, the cause/change structure applies to situations where many different events can occur that increase or decrease the number of objects in someone's possession, or to events that change the amount of some substance in a location (e.g., "There were five gallons of gasoline in the tank; I poured in three more gallons"). It is not hard to generate different verbs that refer to events that fit into the cause/change schema or different situations that fit into the combination or comparison schemata. There are also situations that can be interpreted naturally with more than one of the schemata. For example: "Jack built four birdhouses yesterday; today he built six more birdhouses," may most naturally be thought of as a combination. However, it also can be understood with the cause/change schema, considering the initial amount as the number of birdhouses built before, and the change as an increase in the number of birdhouses caused by today's work.

In our model of the problem-solving process, the input text is translated first into a parsed form, in Anderson's ACT formalism (Anderson, 1976). One of the three semantic structures is constructed, based on categorical information stored about the verbs in the sentences. Note that the construction of a semantic representation involves processing much like that involved in ordinary language processing, with inferences made in order to achieve a coherent structure. However, the inferences made in the context of arithmetic word problems are quite different from those made in other contexts, such as ordinary stories. If the sentence "Betty gave Jill five apples" were encountered in a story, the reader would probably be making inferences about Betty and Jill's friendship or about some general goal Betty had, such as a hope that Jill would reciprocate by sharing something that Betty wanted (cf. Schank & Abelson, 1977). In the context of an arithmetic problem, if a person already has the information that Jill had three apples before, then the sentence "Betty gave Jill five apples" produces the inference that a change occurred in the number of Jill's apples, that the direction of the change was an increase, and that the amount of the change was 5.

When a semantic representation has been constructed, the answer is obtained by applying an arithmetic operation. The first three columns in Table 14.1 specify 14 different structures that . . result from representing different addition and subtraction problems. One possible theory is that each of these is simply associated with one of the operations, along with a procedure for assigning the quantities in the problems as arguments of the procedures. The form of the model

TABLE 14.1
Selection of Arithmetic Operators

Schema	Direction	Unknown	Decision
Cause/Change(P,Q,R)	Increase	Result, R	Addition (P + Q)
Cause/Change(P,Q,R)	Decrease	Result, R	Subtraction P - Q)
Cause/Change(P,Q,R)	Increase	Change, Q	Transform to Combine(P,Q,R)
Cause/Change(P,Q,R)	Decrease	Change, Q	Transform to Combine(R,Q,P)
Cause/Change(P,Q,R)	Increase	Start, P	Transform to Combine(P,Q,R)
Cause/Change(P,Q,R)	Decrease	Start, P	Transform to Combine(R,Q,P)
Combine(U,V,W)	------	Combined Amount, W	Addition (U + V)
Combine(U,V,W)	------	Separate Amount, V	Subtraction (W - U)
Compare(J,K,L)	More	Difference, L	Subtraction (K - J)
Compare (J,K,L)	Fewer	Difference, L	Subtraction (J - K)
Compare(J,K,L)	More	Second Amount, K	Transform to Combine(J,L,K)
Compare (J,K,L)	Fewer	Second Amount, K	Transform to Combine(K,L,J)
Compare (J,K,L)	More	First Amount, J	Transform to Combine(J,L,K)
Compare(J,K,L)	Fewer	First Amount, J	Transform to Combine(K,L,J)

that we have programmed is based on a somewhat different intuition, which we consider plausible but not firm. The current model has direct associations from six of the semantic structures to operations. For the remaining structures, a transformation is required to obtain a representation that is associated with one of the operations. For example, for a problem such as: ''Jill had three apples; Betty gave her some more apples; now Jill has eight apples; how many apples did Betty give her?'' the model first generates a cause/change structure with 3 as the starting quantity, 8 as the final quantity, an increase as the direction, and the amount of increase unknown. This is the structure described on the third line of Table 14.1. Then this structure is transformed to a combine structure, with 3 as the first separate amount, 8 as the combined amount, and the second separate amount unknown. This is the structure shown on line 8 of Table 14.1. This new structure is associated with the operation of subtraction, so the system then chooses that operation.

The choice of combine as the canonical structure for missing addend problems is largely speculative on our part, though there is some suggestive evidence in Case's (1978) work that is consistent with our intuition. We consider the specific set of decision rules in Table 14.1 to be quite arbitrary, and probably different individuals have different decision rules associated with the semantic structures.

The nature of these decision and transformation processes remains an open question in our research, and Table 14.1 should be considered as illustrative of the kinds of procedures that seem plausible in the framework we are using.

The idea of a system that solves word problems without generating equations is encouraged by the fact that children can solve many word problems before they have any knowledge of equations. In fact, there are data showing that children can solve some word problems before they begin to learn arithmetic at all (Buckingham & Maclatchy, 1930). The supply of data about solution of word problems by young children is not large, perhaps because it is much more convenient to present word problems as test items to children who are able to read the problems from written text. One of our current projects involves collecting some systematic data to identify the abilities of young children to understand the kinds of information involved in simple word problems.

In one experiment conducted by Mary Riley (Riley and Greeno, 1978), second-grade children were asked to solve a series of word problems that were designed to provide information about the relative difficulty of the three semantic structures that we identified in the theoretical analysis described earlier. Examples of the problems used in the experiment are shown in Table 14.2. In the experiment, students were asked to solve the problems and were also asked to represent the problems using sets of blocks. Table 14.3 shows the structural descriptions of the nine kinds of problems used, and also shows the proportions of correct answers and the proportions of correct representations that the children produced with blocks.

The main finding is that the semantic schemata involved in problems were rather strong determiners of problem difficulty for these children. They had little difficulty with any of the problems with the cause/change structure. The combination problems with the combined amount unknown were all solved correctly, but the students were not as successful with the combination problems with one of the separate amounts unknown. This finding casts doubt on the assumption about decision rules, shown in Table 14.1, that missing addend problems are all transformed into combination structures. We are collecting further data on this matter, but if results like those in Table 14.3 are typical, we should revise our assumptions about the nature of transformations that are typically performed with cause/change and combination problems.

The most striking finding of this experiment is that all of the problems that have comparison structures were relatively difficult for these second-grade children. One interesting item was the discrepancy between the proportions of correct answers and correct representations in Problem Type 7 relative to Problem Types 8 and 9. The higher proportion of correct answers for Type 7 apparently was due to a tendency for students to add the numbers in the problems whether or not they understood the problems. When students were asked to show the relationships using blocks, these were the hardest problems of the set used. In the two remaining types of problems with comparison structures, representation

TABLE 14.2
Examples of Problems

Schema	Example
Cause/Change	1. Joe has 3 marbles. Tom gives him 5 more marbles. How many marbles does Joe have now?
	2. Joe has 8 marbles. He gives 5 marbles to Tom. How many marbles does Joe have now?
	3. Joe has 3 marbles. Tom gives him some more marbles. Now Joe has 8 marbles. How many marbles did Tom give him?
	4. Joe has 8 marbles. He gives some marbles to Tom. Now Joe has 3 marbles. How many marbles did he give to Tom?
Combination	5. Joe has 3 marbles. Tom has 5 marbles. How many marbles do they have altogether?
	6. Joe and Tom have 8 marbles altogether. Joe has 3 marbles. How many marbles does Tom have?
Comparison	7. Joe has 3 marbles. Tom has 5 more marbles than Joe. How many marbles does Tom have?
	8. Joe has 8 marbles. He has 5 more marbles than Tom. How many marbles does Tom have?
	9. Joe has 5 marbles. Tom has 8 marbles. How many more marbles than Joe does Tom have?

using blocks was more successful than problem solution, perhaps because the blocks provided a method of holding the quantitative information in external memory.

The analysis of semantic processing in solution of word problems provides an interesting suggestion regarding instruction. If we are correct, the process of solving a word problem often involves construction of a semantic representation that is only indirectly related to the operations of addition and subtraction that are used to solve the problems, but that is nonetheless an important component of the process. The suggestion that this hypothesis leads to is that students might be instructed to identify the various general semantic structures that occur in word problems and relate them to arithmetic operations in appropriate ways. In arithmetic, this would involve training students in representing problem situations as one of the three general schemata—change in a quantity, a combination, or a comparison—and teaching them the connections between those representations and the addition and subtraction operations. One approach that seems worth trying would be to use techniques of the kind used in concept formation tasks to train students to attend to the relevant dimensions of information. Many of the

training procedures used in experiments that have been concerned with training children to perform more successfully on Piagetian tests of cognitive development can be interpreted as concept formation procedures in which children learn to attend to the features of the situation that are relevant for the task. Gelman's (1969) study of training for number conservation is an important example in which the discrimination learning paradigm was adopted explicitly.

A second issue that arises involves the way in which computational skill is acquired. I have already discussed the fact that at the beginning of instruction in basic arithmetic, children have relatively sophisticated knowledge about counting, and that this is almost certainly an important knowledge base for their acquisition of basic arithmetic facts of addition and subtraction. Another issue involves children's understanding of these facts. The instructional materials used in primary grades emphasize use of manipulative materials, such as blocks or plastic counters, in providing alternative representations of addition and subtraction facts. The idea that seems to underlie this instruction is that students will be able to understand the operations performed with blocks and other concrete, manipulative materials relatively easily, and these will provide a cognitive basis for their understanding of arithmetic expressed in symbolic notation.

TABLE 14.3
Problem Structures and Proportions of
Correct Problem Answers and Representations

Problem Type	Schema	Direction	Unknown	Correct Answers	Representations
1.	Cause/Change	Increase	Result	1.00	1.00
2.	Cause/Change	Decrease	Result	1.00	1.00
3.	Cause/Change	Increase	Change	.83	.94
4.	Cause/Change	Decrease	Change	1.00	1.00
5.	Combine	-----	Combined Amount	1.00	1.00
6.	Combine	-----	Separate Amount	.67	.77
7.	Compare	More	Comparison Amount	.56	.28
8.	Compare	More	Reference Amount	.28	.50
9.	Compare	More	Difference	.42	.83

When we began our study of primary arithmetic in 1976, we planned to focus our attention on relationships between formal notation of arithmetic and manipulations of concrete materials such as blocks, plastic counters, and the number line. Our initial exploratory work using these materials was surprisingly discouraging. Rather than understanding operations on manipulative materials easily, children seemed to have considerable difficulty. The number line was especially troublesome as a medium for representing quantitative information, and we were informed that the children had not received much instruction involving the number line. We were led to wonder whether the children's general understanding of operations with concrete materials might depend rather strongly on the instruction they have received, rather than being something they comprehend easily and naturally. We have not pursued this issue in detail; however, the experience of our informal explorations was sufficiently discouraging that we moved our research program in another direction.

The direction in which we have developed our research is the study of processes of solving word problems, as I have described in this chapter. Children seem to have considerable ability to understand information that describes relationships among quantities in concrete situations involving changes in possession, location of objects, and so on. Our current conjecture is that children's ability to understand and solve word problems might be exploited much more than it is in present instructional practice as a part of the cognitive basis for the acquisition of arithmetic concepts and operations. Rather than basing instruction on relatively abstract representations such as blocks or the number line, we wonder whether addition and subtraction (and later, the more advanced topics of arithmetic) might be taught in relation to more concrete events and situations where people give things to each other, move objects from one room to another, and so on. This involves viewing problem solving as a basis for instruction in arithmetic, rather than as a skill that is more complex than arithmetic knowledge and that has to be built on top of the more basic knowledge of computation. The issue has ramifications that implicate fundamental aspects of the current structure of our teaching of mathematics in the schools, and we have only begun to touch the edges of some of these. However, the ideas seem plausible, and we look forward to a lively period of exploration and research in the years ahead.

CONCLUSIONS

In my concluding comments, I try to extrapolate from the kinds of results we have obtained in our studies of geometry and primary arithmetic. The kinds of issues that are raised by those findings arise in other domains as well, and it seems a reasonable conjecture that there are possibilities for exploring alternative methods of instruction in a number of different domains that correspond to the possibilities that I have been suggesting in the domain of mathematics.

First, the issue of teaching problem-solving strategies in geometry seems quite clearly applicable in other domains where students are trained in problem solving. Strategic knowledge in a problem-solving domain consists of knowledge of the kinds of subgoals that are useful in various problem situations and the plans that are helpful in achieving various goals and subgoals. One advantage of teaching that knowledge to a student in explicit form is that the student will then have a better understanding of her or his own problem-solving achievements (cf. Brown, Collins, & Harris, 1977). It would be reasonable to expect that this might facilitate transfer to other problem-solving tasks, although this conjecture remains to be tested. Explicit instruction in a problem-solving domain could have considerable facilitating effects on students' abilities to solve problems within the domain of instruction, but there may be potential hazards in making strategic knowledge too explicit if it reduces the educational benefits that at least some learners now receive by finding their own solutions for problems. It seems quite likely, however, that if a more detailed analysis of strategic knowledge in a problem domain were taught to individuals who are instructors in that domain, these individuals would have a better understanding of what their students are required to learn in order to succeed as problem solvers and could interact with their students more effectively in instructional situations.

The second general issue raised by the analysis I have presented is that of teaching students how to represent problem situations. There is a very large experimental literature on the process of learning the relevant attributes of a categorical concept, and an interesting extension to that literature has been given in Winston's (1975) analysis of acquisition of concepts in the blocks world. The general idea of analyzing the relevant features of problem domains and then giving specific training in identifying those features seems to be wi 'ely applicable. Recent studies by Larkin (1977) and by Simon and Simon (1978) have indicated that a major difference between expert and novice problem solvers in physics arises from the expert's construction of an abstract representation of the problem situation, in contrast to the novice's more direct attack on the problem. One interpretation of the result is that by achieving a coherent representation of the situation, the expert avoids the need for extensive problem-solving search, because the expert's representation contains information needed to select appropriate problem-solving operators directly. The well-known studies of expert chess and Go players' ability to encode complex game positions rapidly (Chase & Simon, 1973; Reitman, 1976) attest further to the importance of knowledge for representing problem situations to successful problem-solving performance.

Although the experimental literature on concept formation provides a useful starting point for a program of developing instructional technology for representational knowledge, we probably will encounter some important differences when we study concept formation in the domain of problem representation. Traditional study of concept formation emphasized features that permitted classification of stimuli and used simple perceptual features as much as possible.

In the representation of problem situations, the important thing is to find features that are relevant to the selection of a problem-solving method, rather than features that simply distinguish one category of situations from another. This means that the concepts to be acquired are components of a decision process, rather than simple labels. Further, the powerful representations that experts construct apparently involve complex and abstract relationships in the problem situation, rather than simple perceptual attributes. We need to extend our technology for teaching concepts considerably in the domain of problem-solving representations, but it seems a promising and generally applicable idea.

The third general issue raised by these analyses involves the acquisition of procedural knowledge in meaningful ways. It has always seemed reasonable to teach procedures in contexts that involved the situations in which the procedures were to be used to solve problems, both because that should make it more likely that the learner would be able to apply the knowledge appropriately, and because in that way, the new procedures would be more meaningful. However, the analysis of arithmetic problems and procedures may illustrate some of the reasons why that old truism is correct. The problem-solving contexts in which procedures are applied may indicate important semantic distinctions that should be considered as differences in meaning of the procedural concepts that are involved in the instruction. These distinctions are probably important for students to understand, because they are relevant components of the situations in which the students are expected to use procedures to solve problems. They also may be important mediating concepts that are needed to provide understanding of the nature of relationships between concrete problem situations and the abstract ideas involved in problem-solving methods.

ACKNOWLEDGMENTS

REFERENCES

Anderson, J. R. *Language, memory, and thought*. Hillsdale, N J: Lawrence Erlbaum Associates, 1976.

Bobrow, D. G. Natural language input for a computer problem-solving system. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, Mass: MIT Press, 1968.

Brown, J. S., Collins, A., & Harris, G. *Artificial intelligence and learning strategies* (BBN Rep. No. 3642). Cambridge, Mass.: Bolt Beranek and Newman, June 1977.

Buckingham, B. R., & Maclatchy, J. The number abilities of children when they enter grade one. In *29th Yearbook of the National Society for the Study of Education*. Bloomington, Ind.: Public School Publishing Company, 1930.

Case, R. Piaget and beyond: Toward a developmentally based technology of instruction. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Chase, W. G., & Simon, H. A. Perception in chess. *Cognitive Psychology*, 1973, *4*, 55-81.

Gelman, R. Conservation and acquisition: A problem of learning to attend to relevant dimensions. *Journal of Experimental Child Psychology*, 1969, *7*, 167-187.

Gelman, R. Counting in the preschooler: What does and does not develop? In R. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Greeno, J. G. Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976. (a)

Greeno, J. G. Indefinite goals in well-structured problems. *Psychological Review*, 1976, *83*, 479-491. (b)

Greeno, J. G. Process of understanding in problem solving. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (Vol. 2). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Greeno, J. G. A study in problem solving. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Groen, G. J., & Parkman, J. M. A chronometric analysis of simple addition. *Psychological Review*, 1972, *79*, 329-343.

Groen, G. J., & Poll, M. Subtraction and the solution of open sentence problems. *Journal of Experimental Child Psychology*, 1973, *16*, 292-302.

Groen, G., & Resnick, L. B. Can preschool children invent addition algorithms? *Journal of Educational Psychology*, 1977, *69*, 645-652.

Heller, J. I., & Greeno, J. G. *Semantic processing in arithmetic word problem solving*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1978.

Larkin, J. H. *Skilled problem solving in physics: A hierarchical planning model*. Unpublished manuscript, University of California at Berkeley, September 1977.

Reitman, J. S. Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology*, 1976, *8*, 336-356.

Riley, M. S., & Greeno, J. G. *Importance of semantic structure in the difficulty of arithmetic word problems*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1978.

Sacerdoti, E. D. *A structure for plans and behavior* (Artificial Intelligence Center Technical Note 109). Stanford, Calif.: Stanford Research Institute, 1975.

Schank, R. C., & Abelson, R. P. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Simon, D. P., & Simon, H. A. Individual differences in solving physics problems. In R. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Winston, P. H. Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Woods, S. S., Resnick, L. B., & Groen, G. J. An experimental test of five process models for subtraction. *Journal of Educational Psychology*, 1975, *67*, 17-21.

15 An Elaborative Conception of Learner Differences

William D. Rohwer, Jr.
University of California, Berkeley

Over the last several years, my colleagues and I have investigated a type of learning and memory that provides the underpinnings for a variety of human performances. Persons learn the English equivalents of words in other languages so that given *abend,* they can translate *evening.* They learn the ingredients for a favorite dish so that when hungry for a favorite roast chicken, they can stop on the way home from work and remember to buy mushrooms, parsley, shallots, garlic, chives, butter, and broth, as well as the centerpeice, the chicken. From reading a journal article, they not only learn the results of an experiment but also the methods used to obtain them, so that they can later conduct a replication. As these examples illustrate, we have focused on a type of learning and memory that involves the acquisition of connections between items of information and the capability of using cues to recover the information previously combined.

Much of our earlier work was undertaken to develop and evaluate a conception of how this type of learning proceeds, especially when it is effective (Rohwer, 1973). In brief, the conception is that effective learners construct relationships among items by elaborating events in which the items are integral components. Such events can be elaborated either physically—as when we actually use, or observe someone else using, various ingredients in preparing a dish—or mentally—when we imagine such an episode while reading a recipe (cf. Levin, 1976).

More recently, our interest in this elaborative conception has centered on its value in accounting for differences among persons in their performance on learning and memory tasks. Is the conception useful in explaining why persons differ markedly from one another in the proficiency with which they learn vocabulary equivalents or the relationships between experimental methods and results?

23

Moreover, does the conception have implications for attempts to improve the performance of those who are less proficient? These questions form the focus of the present discussion. The first step in addressing them is to describe the scope of differences among persons in their performance on learning and memory tasks. Next, a sketch is presented of an elaborative conception of the sources of these differences. Then, after evaluating this conception in the light of recent evidence, it is examined for implications about how to improve proficiency through instruction.

PERFORMANCE DIFFERENCES ON
LEARNING AND MEMORY TASKS

Psychological investigations of the type of learning considered here have typically involved the presentation of lists of unrelated items for study, followed by a test of eit! r cued or free recall. If the learners are heterogeneous in age, their performance usually varies over a very wide range. In one experiment, for example, persons sampled from several age levels, from age 6 to age 17, were asked to memorize a list of 36 pairs of familiar but unrelated nouns (Rohwer & Bean, 1973, Experiment II). After a single study opportunity, the 6-year-olds achieved an average performance level of less than 17% correct, whereas the 17-year-olds exceeded 67% correct. Age-related differences of this magnitude are common in studies of cued recall and also emerge on tasks requiring free recall.

Although performance often varies enormously across age, differences hardly disappear when age is held constant (Pressley & Levin, 1977). For example, in another experiment (Rohwer, Raines, Eoff, & Wagner, 1977, Experiment III), a sample of 17-year-olds was asked to learn two successive lists of noun pairs similar to those administered by Rohwer and Bean (1973). Performance on the first list was used to divide the group into thirds: high, middle, and low. On the second list, those in the lowest third averaged less than 25% correct, whereas those in the highest third averaged more than 65%. Individual differences, then, as well as age differences, abound in performances that depend on associative learning and memory.

Of these two types of differences, those related to age have received the larger share of attention in recent theoretical work (Kail & Hagen, 1977). The variety of explanations proposed thus far can be divided roughly into four categories of developmental factors: capacity, metamemorial knowledge, semantic knowledge, and strategies or operations. In the first category, for example, it has been argued that maturational increases in the capacity of short-term or working memory or in the capacity for mental attention (Pascual-Leone, 1970) might expand the opportunities for encoding information effectively enough for long-term storage and later retrieval.

In another vein, the knowledge human beings have of the reach and limits of their memory abilities, their metamemorial knowledge, may increase with the experience that comes with age and produce corresponding increases in performance (Flavell & Wellman, 1977). Similarly, it is a plausible hypothesis that the semantic content of long-term memory—a person's knowledge of words, concepts, entities, states, actions, and their properties—increases with age and experience and dramatically multiplies the possible relationships that can be drawn on in acquiring associative information (Campione & Brown, 1977; Moely, 1977). Finally, the range of procedures, whether conceived as structures of operations or as voluntary strategies, that individuals have for encoding and retrieving associative information may also increase with age, resulting in improved performance (Belmont & Butterfield, 1977; Hagen & Stanovich, 1977).

Although these factors have been formulated primarily to explain age differences, they might also be used to account for individual differences within age groups. Persons of comparable age might be expected to vary in metamemorial knowledge, semantic knowledge, and strategies. Even peculiarly developmental factors such as capacity and operational structures can be conceived to vary within age groups, given the assumption that individuals differ in rates of growth and the final levels they attain. This view, that both age differences and individual differences have common sources, characterizes the elaborative conception of the roots of performance variation.

AN ELABORATIVE CONCEPTION
OF LEARNER DIFFERENCES

The starting point for an elaborative conception of learner differences is an assumption about the character and organization of memory units. The assumption resembles a formulation offered by Schank (1975), according to whom the basic unit of memory is the "action-based conceptualization" in which "objects cannot be separated from the action sequences in which they occur [p. 295]." Similarly, the basic units of memory, according to the elaborative conception, are *events*, each of which is comprised of a beginning state and one or more entities involved in some action that changes that state. For example, one such event might consist of a saw, a board, the movement of the saw, and the resulting partitioning of the board.

Although the mental version of an event like that in the "saw-cutting-board" example is assumed to be unitary, it is also assumed to include more than one particular saw, board, action, and change of state. Mentally, this event might comprise a variety of woodcutting implements (handsaws, chain saws, perhaps even axes or knives), a variety of woods (boards, posts, logs, sticks), and a variety of actions (crosscut, rip, split, whittle), all organized together by means of what Rosch and Mervis (1975) have referred to as "family resemblance."

According to this view, objects form a single class, not because they all share one or more identical attributes or defining properties, but because of shared regularities in relationships among their properties. Similarly, the events in a family share a common relationship among their constituent beginning states (an undivided piece of wood), elements (a piece of wood and a dividing instrument), action (division), and change of state (divided piece of wood). Thus, the basic units of memory, events, are organized into families.

Mental events and event families are also subject to another form of organization. They can be interrelated by virtue of their joint occurrence in a more extended action sequence referred to as an episode. Episodes consist of two or more unitary events linked by one or more actions. A ''saw-cutting-board'' event, for example, can be linked with a ''hammer-nail-propel-board'' event, producing a ''building-a-flower-box'' episode. Like particularistic events, singular episodes may also be related to one another by family resemblance.

Beyond this fundamental assumption about the character and organization of memory units, the elaborative conception consists of a series of propositions. One proposition concerns the process by which associative information is encoded into events; another asserts that the retrieval of associative information involves a reactivation of previously stored events. The remaining propositions specify two factors that can produce differences among persons, both within and between age groups, in the proficiency with which they learn associative information.

The Encoding of Associative Information

According to the present conception, the effective learning of associative information involves the encoding of connections among entities by elaborating them into mental events. When a person interacts with the environment, either as a direct participant or as an observer, the interaction prompts this construction of mental events. One might be prompted to elaborate a ''saw-cutting-board'' event, for example, either by using or observing the use of a saw to cut a board. The power of such experiences is illustrated by research showing that the associative learning of object pairs by young children is enhanced dramatically when they either enact an event involving the pair members (Wolff & Levin, 1972) or observe such an enactment (Wolff, Levin, & Longobardi, 1972).

If elaboration were only activated by direct interaction with the environment, however, its role in the learning of associative information would be severely limited. Clearly, as their performance in most laboratory studies shows, human beings are capable of encoding appreciable amounts of information about relationships among entities even when entirely divorced from a natural context of interaction. Thus, the hypothesis is that elaboration can be activated by indirect prompts, ranging from explicit ones—such as seeing pictorial depictions or hearing verbal descriptions of events—to the prompts implicit in simple directions to learn and remember (cf. Rohwer, 1973).

The effectiveness of indirect prompts, however, varies from person to person. Research indicates, for example, that an indirect prompt, such as the instruction to elaborate imaginary events, has little effect on the performance of very young children but affords older children substantial benefit. Such age differences in the effectiveness of indirect prompts suggest the possibility that the mental construction of events is contingent on the prior construction of relevant event families through direct interaction. If so, elaborative prompts, in the form of event descriptions, should be effective only when they are congruent with events the learner has previously constructed. Rohwer and Levin (1968) tested this prediction in a study of noun-pair learning in fifth-grade children. In two of the conditions, sentences were presented to describe events involving the referents of their constituent nouns. One set of sentence descriptions was constructed to be consistent with events previously experienced by the children (e.g., *Fingers break sticks*), whereas the others described incongruous events (e.g., *Fingers break days*). As compared with appropriate control conditions, the incongruous descriptions did not improve performance, whereas the congruous ones resulted in substantial facilitation.

In the elaborative conception, then, effective encoding involves the construction of mental events that incorporate the associative information to be learned. This elaboration process can be activated, according to the conception, both directly, through environmental interaction or observation, and indirectly, by prompts of varying degrees of explicitness. Indirect activation, however, is contingent on the availability in memory of previously constructed events that can be used in elaborating the new information.

The Retrieval of Associative Information

In the elaborative conception, the retrieval of associative information depends on the generation of cues that are germane to previously encoded events. Such event-related cues may be either self-produced or generated in response to external reminders. One prediction that follows from this proposition is tha' e effect of providing elaborative prompts at encoding should be magnified \ ı study-list items are presented as reminders during recall tests. A second prediction is that in the absence of external reminders, as in free recall, prior elaborative encoding should not enhance initial access to study-list groups but should increase the number of items recalled when at least one member of such a group is retrieved. In contrast, then, to predictions implied by some alternative conceptions—trace independence models (e.g., Slamecka, 1968, 1969) and certain hierarchical models (e.g., Rundus, 1973; Slamecka, 1972), for example—the elaborative conception suggests that the presentation of study-list items as test-trial cues should facilitate recall.

Consistent with both of the predictions drawn from the elaborative conception, Prestianni and Zacks (1974) have reported that when college students are given encoding prompts, they perform better in cued than in free recall, and that

the effect of such prompts in free recall is more apparent on the index of items recalled per study-list group than in terms of the number of groups represented in recall. Similar confirmation comes from a series of unpublished studies we have conducted with students of a variety of ages, from 10-year-olds to 17-year-olds. Across a number of task manipulations, the results have shown that the greater the degree to which study conditions favor elaboration, the greater the superiority of cued to free recall. Moreover, in free recall the prompt effect invariably shows itself only in terms of the number of items recalled per study-list group, not in the number of groups represented in recall.

Learner Differences in Elaborative Propensity

According to the elaborative conception, one of the factors responsible for differences in learning proficiency is elaborative propensity: Persons vary in the explicitness of the indirect prompts required to activate their construction of mental events for encoding associative information. The developmental form of this proposition is that whereas interactive elaboration is inherent in the human being's commerce with the environment from birth onward, the propensity to elaborate under indirect conditions increases as a function of age, from early childhood to adulthood. This proposition arises from assumptions similar to those in Piagetian theory (Piaget & Inhelder, 1973) and in dialectical theories as well (e.g., Meacham, 1977). In brief, the rationale is that the mental procedures of mature intelligence have their origins in, and depend on, the overt interactive procedures of the immature.

The developmental proposition yields predictions that have now been confirmed in a number of empirical studies. One prediction is that age differences in performance will emerge under conditions of indirect prompting but not under conditions that prompt elaboration through direct interaction. In keeping with the prediction, Irwin (1971) found that kindergarten children learned as efficiently as sixth graders when instructed to enact an event physically for each of a list of object pairs. Additional confirmation was obtained by Wolff and Levin (1972), who asked kindergarten and third-grade children to learn a list of 16 object pairs in one of two ways. The children were instructed to create events involving the pair members either by generating images of interactions or by enacting such interactions with the actual objects. Given the indirect-prompt of imagery instructions, the older children produced 35% more correct responses than the younger, whereas with direct prompting in the enactive condition, the discrepancy was only 8%.

Another prediction concerns the effects of varying the explicitness of indirect prompts. Specifically, across the age range of adolescence, the expectation is that by the end of this period, the prompting implicit in simple study instructions should be sufficient to activate elaboration, whereas at the beginning of adolescence, more explicit prompts are required. The results of a study by Rohwer and

Bean (1973, Experiment II) confirmed this prediction. Sixth- and 11th-grade students were given either standard study instruction for learning a list of noun pairs or were directed to create sentence descriptions of events involving the pair members. The more explicit prompt instructions markedly facilitated the performance of the younger students, but the two conditions were equally effective for the older.

Converging evidence has recently been reported by Pressley and Levin (1977), who obtained descriptions from older and younger students of the activities they engaged in to encode pair members after simple study instructions. Elaboration-like activity was reported far more frequently in the older than in the younger samples. In addition, the frequency of reported elaborative activity was significantly related to recall performance at both age levels.

Despite these instances of confirming evidence, the developmental proposition is by no means sufficient to account for all of the relevant data. Pressley and Levin (1977), for example, found substantial variation in reported elaborative activity within age groups as well as between them. Furthermore, in a number of studies (Rohwer & Bean, 1973, Experiment I; Rohwer, Raines, Eoff, & Wagner, 1977, Experiments I and II), explicit prompt instructions have boosted the performance of late adolescents (17- and 18-year-olds) as much as that of preadolescents (11- and 12-year-olds).

Surmising that these irregular outcomes might be due to individual differences in elaborative propensity within age groups, Rohwer et al. (1977, Experiment III) conducted a further study to evaluate this possibility. Once again, samples of 11-year-olds and 17-year-olds were asked to learn a list of noun pairs after receiving either standard study instructions or instructions prompting them to create sentence descriptions of interactive events. Prior to the imposition of this prompt manipulation, however, all participants were asked to learn an initial list of pairs under standard study instructions. Performance on this first list was used as an individual-differences index to divide each age group into three levels of entering proficiency. Then, on second-list learning, the results revealed that of the six groups defined by the combination of age and proficiency, only one—the high-proficiency 17-year-olds—performed as well under standard as under prompt instructions. In fact, in terms of both absolute levels of performance and the magnitude of the prompt effect, the lowest third of the older sample was indistinguishable from the highest third of the younger students. Thus, it appears that elaborative propensity varies as markedly across individuals as it does across ages.

At present I can offer little more than speculation about possible sources of the presumed individual differences in elaborative propensity. Such differences might be thought to stem from corresponding differences in developmental rates; that is, variations in maturational pace and experiential opportunity might result in differences among persons in the point at which they successfully convert their interactive elaborative procedures to a functional mental form. Alternatively,

propensity differences may be a manifestation of variations in some more general cognitive dimension such as that of analytic-synthetic, for example. It would be premature, however, to pursue these speculations further without first examining the possibility that a second factor, event repertoire, interacts with proficiency to produce learner differences in performance.

Learner Differences in Event Repertoire

In addition to propensity variation, differences in associative learning proficiency should also arise, according to the elaborative conception, from variations in event repertoires—the mental events and episodes persons have previously constructed and stored in memory. This proposition is a corollary of the preceding proposition about the roots of elaborative encoding. If indirect elaboration depends on prior interactive elaboration, and if persons differ in the events and event families they have previously constructed through direct interaction, they should also differ in the kinds of new associative information they can learn effectively. Thus, performance should vary across persons who differ in the congruence of their event repertoires with the character of the associative information to be learned. But when elaborative propensity is equated and when the associative information to be acquired is equally congruent with the event repertoires of different persons, their performance should be equivalent.

This prediction should hold for learner variation in general, including differences across age as well as individual differences within age groups. As yet, however, none of these contentions about event repertoire effects have been evaluated empirically, mainly because of the need to devise and validate methods of assessing the congruence between event repertoire and the information to be acquired. Our preliminary efforts in this direction are described shortly. If such attempts are successful, they will be followed by experimental tests of the prediction about age differences and by studies of the relative contributions of propensity and repertoire to observed differences in performance.

RECENT RESEARCH ON ELABORATION
AND LEARNER DIFFERENCES

The bulk of our recent work has concerned the propensity factor as a source of age and individual differences in associative learning over the age range of adolescence. Accordingly, the principal experimental manipulations have consisted of varying the explicitness of elaborative prompts given prior to the presentation of associative learning tasks. In the first study described, this manipulation was used to investigate learner differences in the effectiveness and persistence of prompt instructions. The second study reported is a pilot attempt to devise and

validate a method for assessing the congruence between event repertoire and the associative information to be learned in a paired associate task.

Elaborative Propensity and Prompt Effectiveness

The results of previous research on developmental and individual differences in elaborative propensity have left a number of issues unresolved. Four of these were addressed in the present study conducted in collaboration with James Litrownik. One issue concerns the dependability of differences between preadolescent and late-adolescent students in the pattern of prompt effects observable in their performance on associative learning tasks. A second issue is whether these patterns are limited to a single type of task, paired associates, or extend to other arrangements of items as well. Third, in keeping with an interest in individual differences, prompt conditions were manipulated within, as well as between, subjects to address the issue of propensity variation within age groups. This provision of the design also provided a way of confronting a final issue—whether persistent effects of an instructional prompt could be produced more readily among older than among younger students of comparable initial performance levels.

Samples of 108 students were drawn from each of two grade levels, fifth and 11th, where the average ages were 11 and 17 years. To provide estimates of the generality of effects across task arrangements, each individual learned two types of word lists—lists in which 72 familiar nouns were presented for study in pairs, and lists in which 48 nouns were presented in tetrads, groups of four words at a time. Because test trials were conducted by a cued-recall procedure in which one word from each study-list group was presented as a reminder of the missing items, these different list lengths were used to equate for the number of words to be recalled, 36, across list types.

All students were administered a list of pairs and a list of tetrads for a single study-test cycle on each of four days: Monday, Tuesday, Wednesday, and Friday. Half received the lists of pairs first; the other half began with the lists of tetrads. Students were also assigned to one of three treatment conditions, distinguished by the sequence of task instructions given across days, as shown in Table 15.1.

In the baseline condition, prior to each list on all four days, students were given standard instructions that merely explained the procedure that would be followed, urged careful study, and offered short samples of the two types of list. As the sequence of specific lists was the same across all students, one purpose of including this baseline condition was to control for both generalized transfer and list difficulty. The condition also provided a standard for comparing individual-differences effects that were examined in subsequent regression analyses.

In the prompt condition, standard instructions were given on the first 2 days and on the final day. On Wednesday, however, the prompt group was asked to

TABLE 15.1
Instructional Conditions in the Elaborative Propensity Study

Condition	Days			
	Monday	Tuesday	Wednesday	Friday
Baseline	Standard	Standard	Standard	Standard
Prompt	Standard	Standard	Prompt	Standard
Repetition	Standard	Repetition	Prompt	Standard

construct, for each study-list group of nouns, a sentence describing an interaction among their referents, and this procedure was illustrated for items on the sample list.

Students in the repetition group also received standard instructions on the 1st and last day and prompt instructions on the 3rd day. On Day 2, however, rather than being asked simply to study the words, these students were directed to rehearse each study-list group repeatedly during intergroup intervals. The repetition condition was included mainly in the interest of construct validity. Our reasoning was that if students performed well under standard instructions because of a high degree of elaborative propensity, the activity of repetition should impede their learning; otherwise, it should have little effect. In the repetition condition, the relationship between Day 1 and Day 2 performance in the older sample was expected to be substantially attenuated in comparison with the younger.

Performance was indexed by the number of correct responses given on the test trial administered for each list. In terms of group averages, the results are shown in Table 15.2. As the mean values indicate, analysis of variance confirmed that the initial (Day 1) performance of the older students, on both the pair and tetrad tasks, far outstripped that of th. younger samples. The data for Day 2 produced significant interactions of grade and condition for both types of list, confirming the prediction that repetition instructions would impede the learning of the older students relative to the average for the baseline and prompt groups, but not that of the younger students.

The amount of facilitation produced by the prompt instructions was assessed by contrasting the Day 2-to-Day 3 gain in the prompt condition with that in the baseline condition. For the list of pairs, the predicted interaction of grade and condition was not significant, even though the relative gain in the younger samples (6.8 items) was nearly twice that in the older (3.8 items). On the tetrad list, however, the target interaction was significant; the relative amount of facilitation produced by the prompt instructions was substantially larger among the younger (11.3 items) than among the older students (2.3 items).

The two types of list also yielded different outcomes with reference to a final issue for which the data in Table 15.2 are pertinent—whether instruction would

produce enduring changes in propensity more readily among late adolescents than among preadolescents. In the present study, this implication translates into the prediction that the performance levels achieved in the prompt condition on Day 3, when instruction was provided, would be maintained on Day 4 to a greater degree in the older than in the younger groups.

This prediction was tested by contrasting, for each grade, the prompt and baseline conditions in terms of the difference in performance between Day 3 and Day 4. For the list of pairs, the prediction was not confirmed, as the differences were essentially zero; the fifth graders showed a net loss of 0.5 items and the 11th graders, a net gain of 0.3 items. Yet on the tetrad list, the predicted interaction was significant, reflecting a relative loss of 3.5 items among the preadolescents in contrast to a net gain of 1.2 items for the late adolescents. Although this result apparently confirms the prediction about age differences in susceptibility to instruction, at least for tetrads, a question of interpretation remains. The age effect may be a regression artifact. The persistence of the instructional effect among the older students may have been due mainly to those whose performance level was high prior to receiving prompt instructions.

To appraise this possibility and to assess two other issues about individual differences in elaborative propensity, three regression analyses were conducted for each of the two list types. Within the relevant cells in the design, estimates were obtained of the slopes in the functions relating Day 1 performance to each of three dependent variables. Appropriate contrasts were then formed among the estimated slopes and were tested for significance.

The first issue examined by this method was whether the deleterious effect of repetition instructions varied with initial performance levels in similar ways across grades. Slope estimates were obtained for the regression of Day 2 performance on Day 1 performance. Within each grade, these estimates were computed

TABLE 15.2
Mean Numbers of Correct Responses as a Function of Grade,
Condition, List Type, and Days[a]

| | | List Type | | | | | | | |
| | | Pairs | | | | Tetrads | | | |
Grade	Condition	1	2	3	4	1	2	3	4
	Baseline	8.6	8.7	7.9	9.6	4.4	5.6	4.0	4.2
5	Prompt	9.5	7.9	13.9	15.1	4.2	4.8	14.5	11.2
	Repetition	8.9	6.4	14.0	12.9	5.9	3.7	13.5	9.1
	Baseline	20.5	20.6	20.0	21.3	13.7	16.2	17.6	18.5
11	Prompt	21.8	21.8	25.0	26.6	12.6	19.3	23.0	25.1
	Repetition	23.2	12.1	24.2	25.7	15.0	5.8	25.4	24.6

[a] Maximum number of correct responses = 36.

separately for the repetition condition and for a combination of the baseline and prompt conditions, in which a common set of standard instructions had been given on both days.

The results are shown in Fig. 15.1 for the lists of pairs and in Fig. 15.2 for the lists of tetrads.[1] The lengths of the estimated regression lines indicate the range of Day 1 scores observed in each condition, and the crosshatches represent 1 standard deviation on either side of the group means. Contrasts among the estimated slopes confirmed the visual impression gained from these plots, yielding an interaction of grades and conditions that was significant, but only for the tetrad list. In the fifth-grade sample, the estimated regression lines are virtually parallel for the repetition and the baseline-prompt conditions. In the 11th-grade sample, however, the slope for the repetition condition is considerably less steep than for the baseline-prompt condition. Consistent with predictions for the older students, then, the deleterious effect of repetition instructions is more pronounced, the higher the students' initial performance levels and, presumably, the higher their elaborative propensity.

The second issue addressed here concerns individual differences in the effects of prompt instructions. According to the developmental proposition in the elaborative conception, the 17-year-old samples were expected to include students representing an extensive range of elaborative propensity, from very low to very high. For these older students, therefore, it was predicted that the amount of facilitation produced by prompt instructions would be a decreasing function of initial performance level, the operational index of propensity. In contrast, the younger students were expected to represent a very restricted range of propensity, including few, if any, genuinely high-propensity students. For these students, then, the magnitude of the prompt effect was predicted to be comparatively independent of initial performance.

These predictions were tested by computing, for the prompt and baseline conditions within each grade, the estimated slopes in the functions relating the number of items gained from Day 2 to Day 3 with the number correct on Day 1. Analysis of the results for the lists of pairs, displayed in Fig. 15.3, revealed that the predicted interaction of grade and condition was significant. In the baseline conditions for both grades, performance changes from Day 2 to Day 3 appear negligible and virtually constant across the range of initial performance. Furthermore, in the younger sample, the amount of facilitation produced by the prompt instructions also appears constant. In the older sample, however, facilitation was a negative function of initial performance level, as would be expected if profi   nt learners need only an implicit prompt to elaborate.

---

[1]In all regression lines in all figures, numerals designate mean performance levels; line length represents range of actual scores on Day 1; and crosshatches mark 1-standard-deviation distances from Day 1 means.
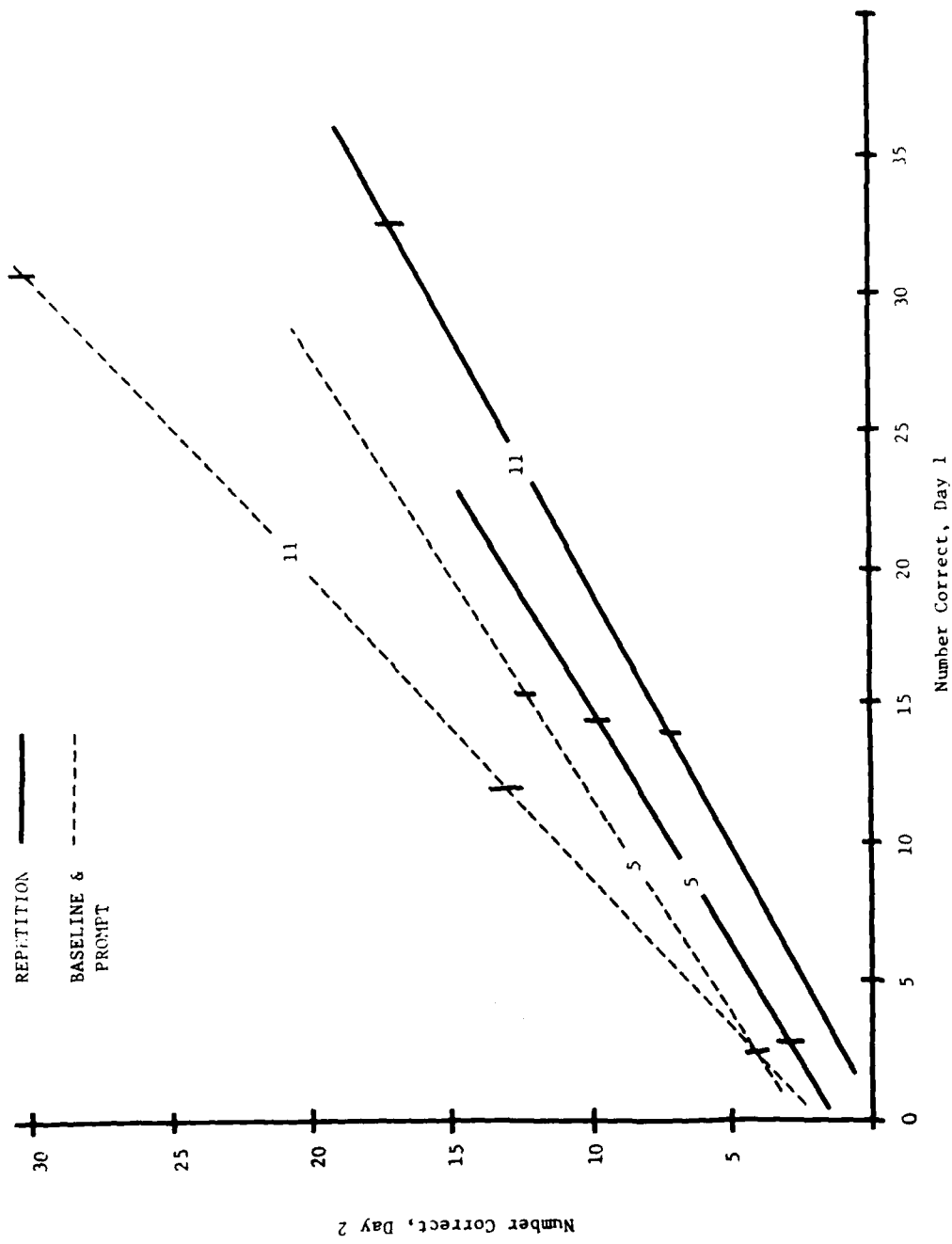
FIG. 15.1. Estimated regression lines as a function of grades and conditions: Number of correct pairs, Day 2, on number of correct pairs, Day 1
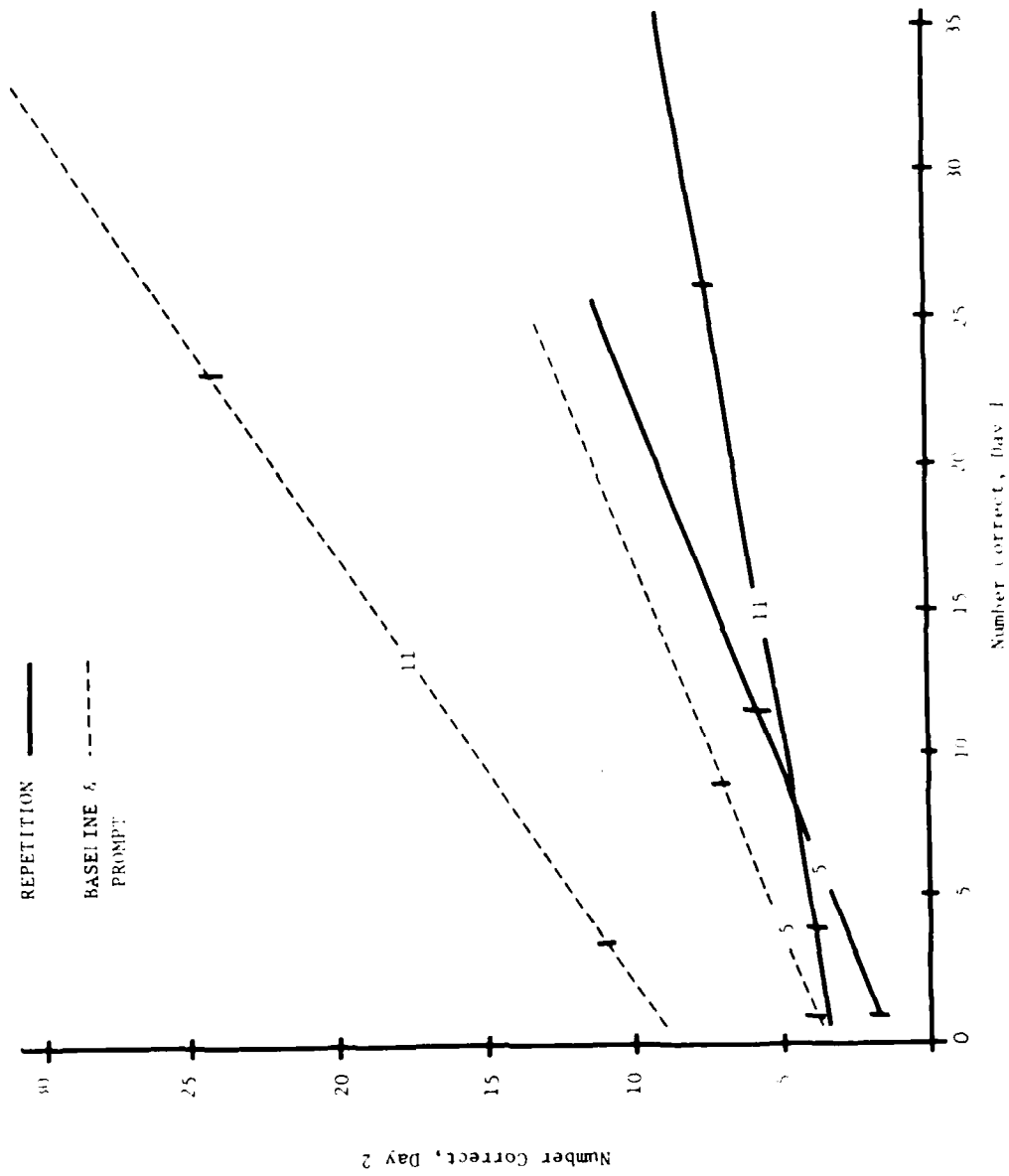
FIG. 15.2. Estimated regression lines as a function of grades and conditions. Number of correct tetrads, Day 2, on number of correct tetrads, Day 1.

FIG. 15.3. Estimated regression lines as a function of grades and conditions: Change from Day 2 to Day 3 in number of correct pairs on number of correct pairs, Day 1.

This same interaction was also significant for the tetrad lists, but as an inspection of Fig. 15.4 indicates, its form was quite different. Among the fifth graders, the relative benefit of prompt instructions was larger, the higher the initial performance level; that is, for these younger students prompt instructions increased rather than decreased the range of individual differences in tetrad performance. In contrast, among older students the explicit prompt, as predicted, diminished individual differences in tetrad performance just as it did for pairs.

A final regression analysis was conducted to resolve the interpretive question raised by the apparent age differences in the persistence of the effect of instructional prompting on tetrad performance. Accordingly, changes in performance from Day 3 to Day 4 were computed for the prompt and baseline conditions within each grade. The slopes of the estimated functions relating this variable to Day 1 performance were then obtained and contrasted.

The results are shown in Fig. 15.5 for the lists of pairs and in Fig. 15.6 for the lists of tetrads. For lists of pairs, the previous analysis of variance indicated that the relative benefit of prompt instructions was maintained equally across grades. This outcome was further extended by the present analysis of the slopes of the estimated regression lines, which showed that the persistence of instruction was also virtually constant across individual differences within grades; that is, no significant differences resulted from contrasts among the four slopes.

Similarly, analysis of the slopes for the tetrad lists revealed no significant contrasts. Apparently, then, the greater persistence of instructional effects among the older than among the younger students cannot be attributed merely to a continued maintenance of high performance by the initially proficient 17-year-olds. Thus, a possible conclusion is that a given amount of instruction has a larger payoff for older than for younger students.

In many respects, the results of the present study offer substantial support for key propositions in the elaborative conception of learner differences. In other respects, however, the results suggest caution, especially in interpreting effects associated with prompt instructions. For example, these instructions failed to diminish individual differences among the younger students in their learning of the lists of pairs and even magnified such differences on the tetrad lists. Perhaps the instructions were inadequate for fully engaging the elaborative capabilities of these preadolescent learners and could not, therefore, compensate for differences in propensity within the fifth graders, nor for differences between these younger students and the older 17-year-olds. And, indeed, the advantage of the older students on the list of pairs was as large with prompt instructions at it was initially with standard instructions, contrary to the proposition that propensity is a major source of age differences in performance. Finally, inadequacies in the prompt instructions might also explain the drop in tetrad performance, from Day 3 to Day 4, among fifth graders in the prompt condition, making it premature to conclude that instructional effects are less enduring among younger than among older students.
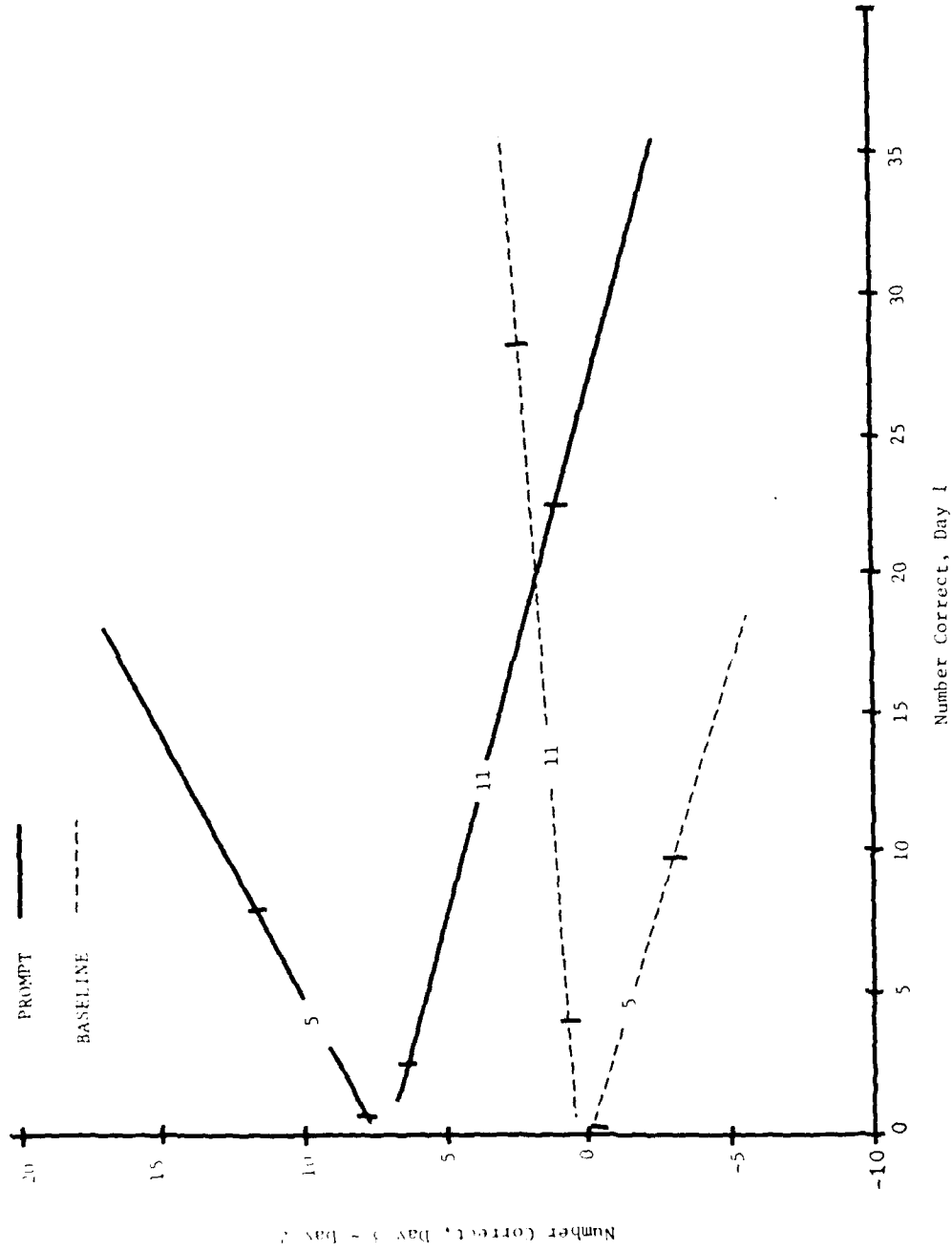
FIG. 15.4. Estimated regression lines as a function of grades and conditions: Change from Day 2 to Day 3 in number of correct tetrads on number of correct tetrads, Day 1.

39
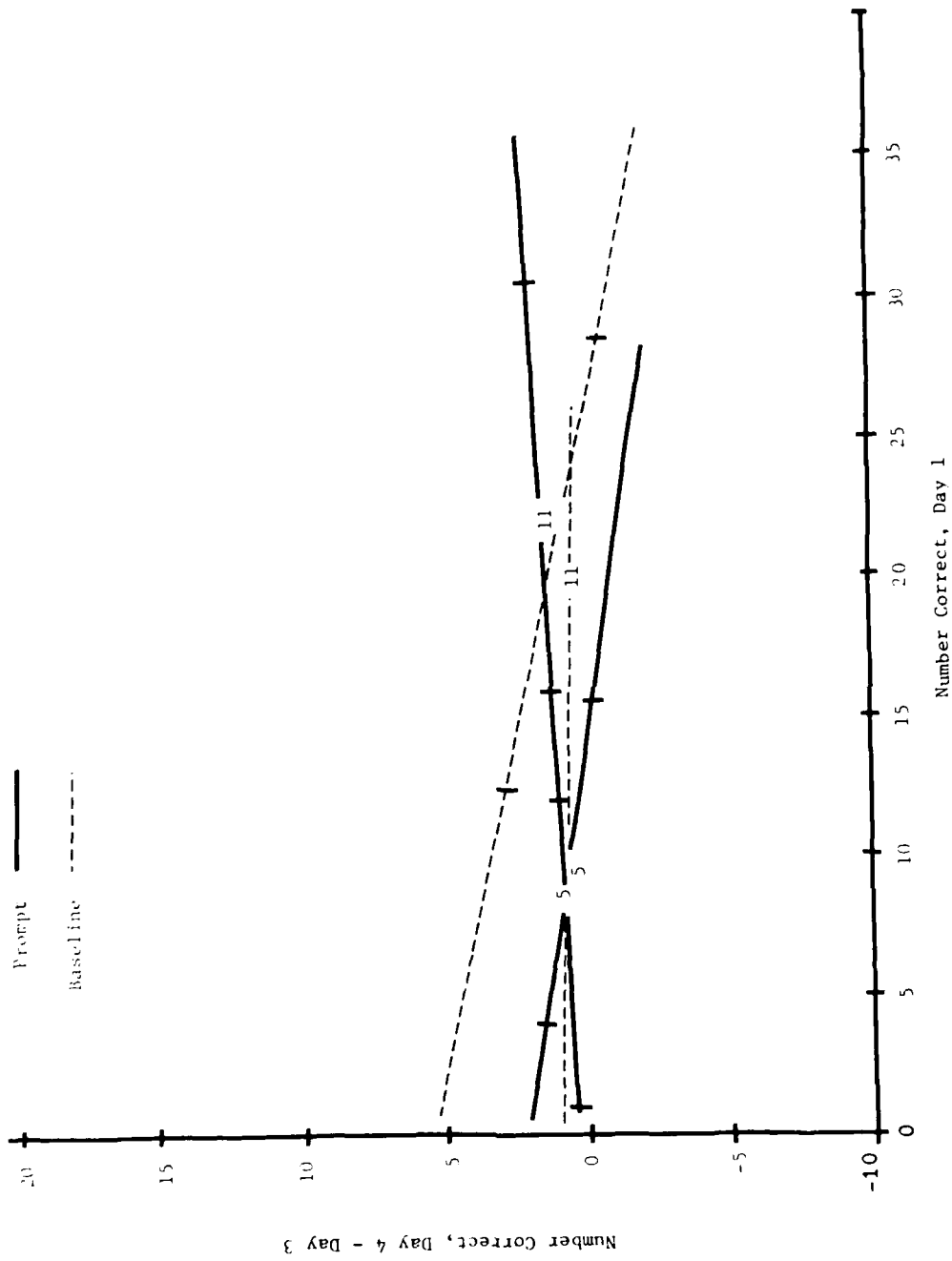
FIG. 15.5. Estimated regression lines as a function of grades and conditions: Change from Day 3 to Day 4 in number of correct pairs on number of correct pairs. Day 1.

40
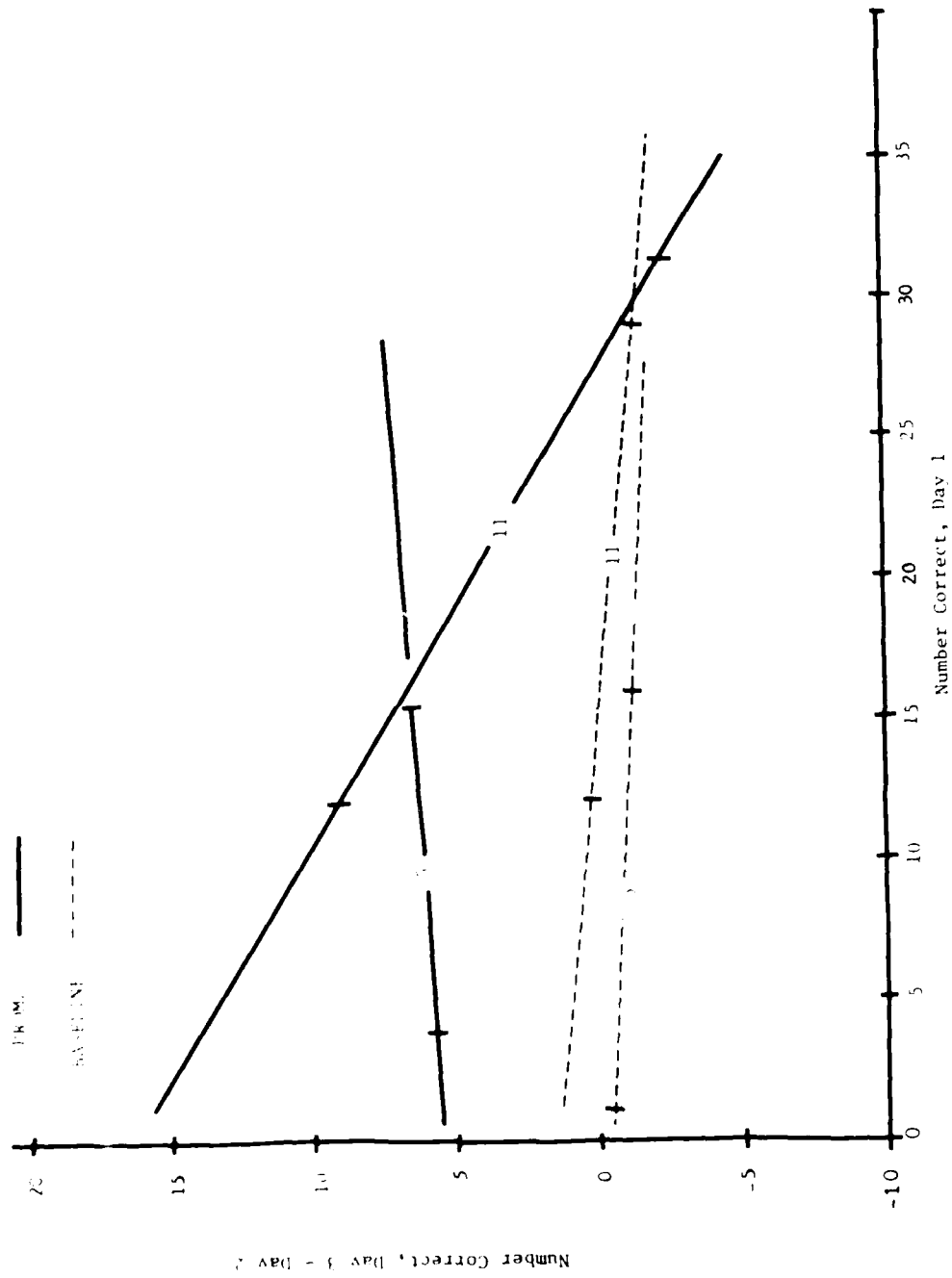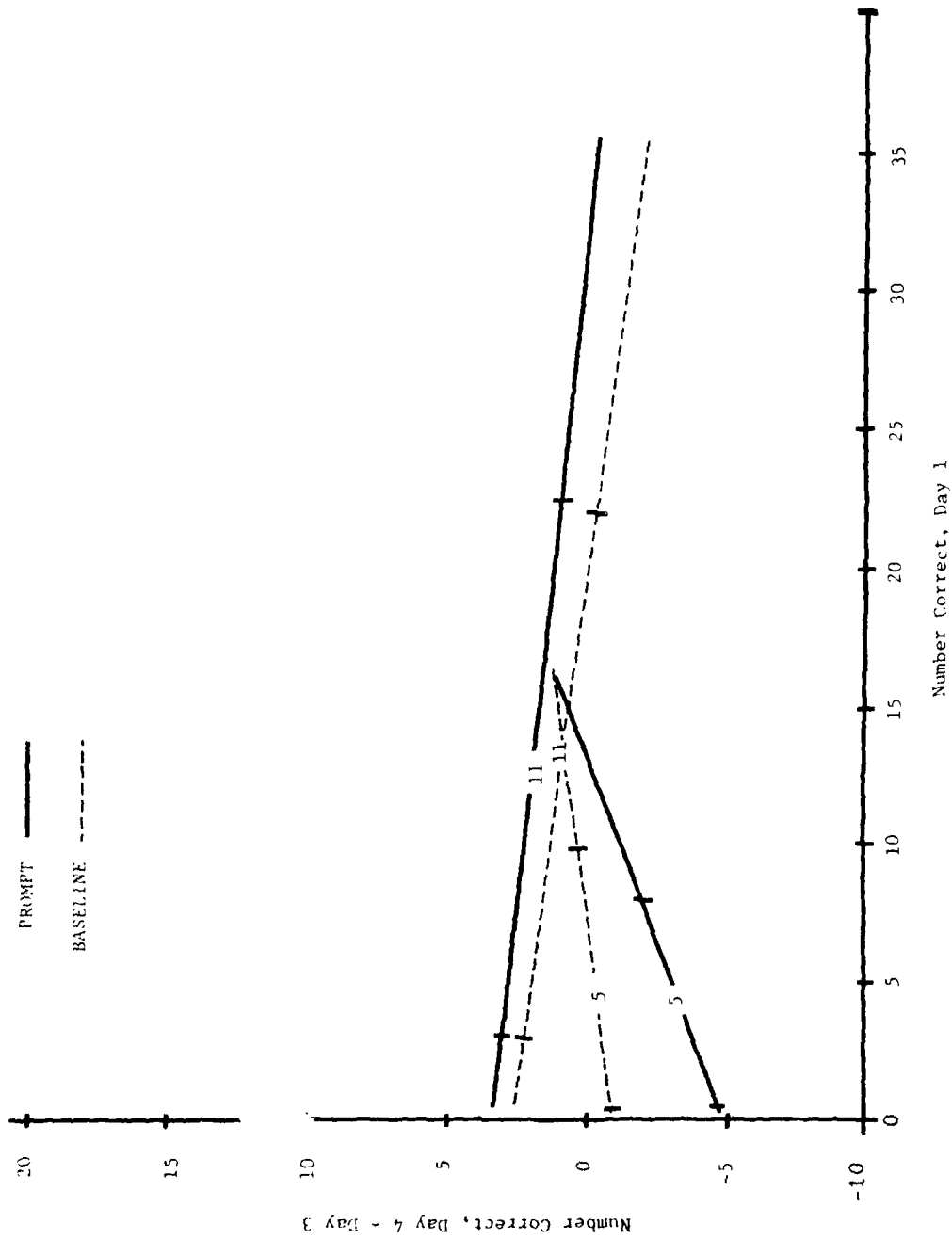
FIG. 15.6. Estimated regression lines as a function of grades and conditions: Change from Day 3 to Day 4 in number of correct tetrads on number of correct tetrads, Day 1.

41

As a preliminary check on the efficacy of the prompt instructions used in the present experiment, we recently completed a pilot study of an alternative prompting procedure. Ten students were drawn randomly from the same fifth-grade population that furnished the previous sample. Furthermore, each student was administered the same list of noun pairs in the same method as that used on Day 3 of the preceding study.

Prior to learning this target list, however, the students were given elaborative instruction that was considerably more extended than that used before. The session began with the learning of a sample list of pairs under standard instructions. Then a procedure was explained for elaborating the referents of pair members into interactive events. This explanation emphasized four steps: (1) conceiving a context that might plausibly encompass both referents; (2) imagining an incident that jointly involved the referents; (3) also imagining the state that would result from the incident; and (4) checking to determine whether the name of each referent was an effective reminder of the event generated. Next, this procedure was illustrated by the experimenter for a second sample list. Finally, the students practiced the procedure on yet a third list of examples.

Immediately after this instructional sequence was completed, the students were asked to learn the test list of 36 noun pairs during a single study-test cycle. Their performance was impressive—an average of 24.3 correct responses. This more extended instruction produced markedly higher scores than the average of 13.9 correct given by prompted fifth graders and was nearly equivalent to the average of 25.0 correct attained by the 11th-grade students in the previous study.

In view of this result, one of our current priorities is to conduct a formal experiment using extended elaborative instruction to determine the limits of the propensity factor in accounting for age and individual differences in associative learning. A second priority raised by the results of another pilot study is to make a similar determination for the factor of event repertoire.

As yet, we have been unable to devise a method for directly assessing the event repertoires individuals have constructed and stored in long-term memory. Meanwhile, Mitchell Rabinowitz, James Litrownik, and I have begun to explore the potential value of an indirect method in which informants rate the ease of elaborating noun pairs into interactive events. The rationale for this method is that an informant's event repertoire should determine his or her judgment of the relative difficulty involved in creating events for the referents of unrelated words. To gain assurance that such judgments are determined primarily by event repertoire rather than other factors, independent validation is a necessity.

The results of an initial pilot study suggest that an attempt to validate the method may be worthwhile. The study involved two phases. In the initial phase, each of 24 fifth-grade students was presented with 37 sets of five familiar but unrelated nouns. Within a word set, students were asked to consider the four pairs that could be assembled by thinking of the first noun in conjunction with each of the four remaining nouns. In considering the four pairs, the students were to imagine an interactive event for each and then, having done so, were to rank

the pairs in terms of the relative difficulty they experienced in constructing the events. This procedure was then illustrated with four sample sets of words, after which the students ranked the test sets of words at their own pace.

The resulting rankings were used in the second phase of the pilot study to construct two lists—one consisting of the pairs rated easiest on the average, and the other of pairs rated most difficult. The lists were administered in a cued-recall task given to independent samples of 10 fifth graders each. Students in both samples received the brief prompt instructions used previously by Rohwer and Litrownik, followed by a single study-test cycle.

The results revealed substantial differences in the difficulty of learning the two lists. Students given the easy list made an average of more than 23 correct responses, whereas those given the difficult list averaged less than 12 items correct. Viewed in comparison to the results obtained by Rohwer and Litrownik, the prompted performance of the present fifth graders on the easy list is remarkably high, emphasizing the unusually low performance of those given the difficult list. Evidently, the ranking method is quite sensitive to some factor that bears a strong relationship to associative learning, but the question remains whether this factor can be identified with the construct of event repertoire.

An attempt to answer this question is now under way in the form of a valida-tion study. Once again, the study consists of two phases, consisting of a ra g task and a paired-associate learning task. Ratings are being obtained from s -ples of fifth-grade students under one of two kinds of instructions: promp - structions similar to those used in the pilot study, and standard instructions that ask the students to estimate the difficulty they would experience in learning the alternative pairs in their customary way.

The second phase of the study will involve a three-factor design. Inde c dent samples of fifth graders will be given either prompt or standard instructions for learning one of four lists of noun pairs: pairs rated easiest and most difficult under prompt ranking instructions, and pairs rated easiest and most difficult under standard ranking instructions. If the ranking method is a valid indicator of event repertoire, the results should form a three-way interaction. For lists ob-tained under prompt ranking instructions, the difficulty manipulation should have a larger effect in the prompt than in the standard learning conditions, whereas the direction of this interaction should reverse for lists obtained under standard ranking instructions. Unfortunately, the data are not yet in, leaving the story incomplete.

INSTRUCTION AND DIFFERENCES
IN ELABORATIVE PROFICIENCY

In considering how research on elaboration and learner differences might bear on instructional issues, it is important to acknowledge that many fundamental ques-tions about elaboration itself still remain to be answered. With the stipulation that

the discussion is provisional, however, it is in order to consider possible answers to two principal questions. Can instruction be designed that would effectively increase elaborative propensity and, hence, associative learning proficiency? If so, what instructional provisions should be made for learner differences?

Any answer to the question about the feasibility of improving elaborative propensity through instruction presupposes a prior decision about the value of doing so. I do not take a position on this issue, since I have an obvious conflict of interest in the way it is resolved. Given a positive decision, however, the results obtained by Rohwer and Litrownik would encourage me about the possibility of designing effective instruction, even though the implications of these results are quite limited. To review, a 30-minute instructional experience produced substantial performance gains, and although they varied with age and list type, these gains largely persisted over a 48-hour interval. Though not imposing in themselves, such facts seem to warrant optimism, especially when compared with the outcomes of numerous other attempts to improve proficiency in the use of strategylike learning procedures (Belmont & Butterfield, 1977; Campione & Brown, 1977; Hagen & Stanovich, 1977; Rohwer & Ammon, 1971). Nevertheless, the research available to date is limited in two ways. First, we have no evidence about the persistence of elaborative instructional effects over longer periods of time. Second, we have little empirical basis for expecting transfer of such instructional effects beyond the tasks used in training, much less to actual subject-matter materials (cf. Atkinson & Raugh, 1975; Pressley, 1977).

If it is too early to answer the feasibility question, it is even more premature to offer answers to the question about how elaborative instruction should be designed to accommodate learner differences. Such answers depend critically on how large a role differences in event repertoire play in determining functional elaborative proficiency. If repertoire effects are substantial, they might imply, for example, the need for long-term instructional experience, perhaps of kinds that have yet to be devised. On the other hand, if propensity is the main determining factor, more circumscribed instruction might suffice to increase this tendency, and even if not, instruction can be designed to offer immediate prompts for those students who need them. In either case, recommendations must await the creation of alternative instructional sequences and evaluations of their effectiveness for different types of learners.

Despite the imposing magnitude of what is not yet known, research to date makes it possible to speculate that elaborative instruction would necessarily take certain factors into account. It seems clear, for example, that variations should be available for learners of different ages, especially if the task demands are relatively difficult, as in the comparison of pairs and tetrads of items. Similar variations should be made available for students within age groups, depending on their initial performance levels. Finally, some students may need assistance in developing their repertoire of events, whereas others may need help mainly in increasing their elaborative propensity. Each of these principles, if it is to be

transformed into prescriptive treatments for individual students, presupposes substantial diagnostic capabilities. These capabilities depend not only on techniques for assessing the propensities and repertoires of learners but also on a deeper knowledge of the processes of learning and the organization of memory as well. Instructional improvement, therefore, awaits further progress in both domains.

ACKNOWLEDGMENTS

REFERENCES

Atkinson, R. C., & Raugh, M. R. An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *104*, 126-133.

Belmont, J. M., & Butterfield, E. C. The instructional approach to developmental cognitive research. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Campione, J. C., & Brown, A. L. Memory and metamemory development in educable retarded children. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Flavell, J. H., & Wellman, H. M. Metamemory. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Hagen, J. W., & Stanovich, K. G. Memory: Strategies of acquisition. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Irwin, M. H. *A developmental study of elaboration modality effects in paired-associate learning*. Unpublished doctoral dissertation, University of California, Berkeley, 1971.

Kail, R. V., & Hagen, J. W. (Eds.). *Perspectives on the development of memory and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Levin, J. R. What have we learned about maximizing what children learn? In J. R. Levin & V. L. Allen (Eds.), *Cognitive learning in children*. New York: Academic Press, 1976.

Meacham, J. A. Soviet investigations of memory development. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Moely, B. E. Organizational factors in the development of memory. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Pascual-Leone, J. A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, 1970, *63*, 301-345.

Piaget, J., & Inhelder, B. *Memory and intelligence*. New York: Basic Books, 1973.

Pressley, M. Children's use of the keyword method to learn simple Spanish vocabulary words. *Journal of Educational Psychology,* 1977, *69,* 465-472.

Pressley, M., & Levin, J. R. Developmental differences in subjects' associative-learning strategies and performance: Assessing a hypothesis. *Journal of Experimental Child Psychology,* 1977, *24,* 431-439.

Prestianni, F. L., & Zacks, R. T. The effects of learning instructions and cueing on free recall. *Memory & Cognition,* 1974, *2,* 194-200.

Rohwer, W. D., Jr. Elaboration and learning in childhood and adolescence. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 8). New York: Academic Press, 1973.

Rohwer, W. D., Jr., & Ammon, M. S. Elaboration training and learning efficiency in children. *Journal of Educational Psychology,* 1971, *62,* 376-383.

Rohwer, W. D., Jr., & Bean, J. P. Sentence effects and noun-pair learning: A developmental interaction during adolescence. *Journal of Experimental Child Psychology,* 1973, *15,* 521-533.

Rohwer, W. D., Jr., & Levin, J. R. Action, meaning and stimulus selection in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1968, *7,* 137-141.

Rohwer, W. D., Jr., Raines, J. M., Eoff, J., & Wagner, M. The development of elaborative propensity in adolescence. *Journal of Experimental Child Psychology,* 1977, *23,* 472-492.

Rosch, E., & Mervis, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology,* 1975, *7,* 573-605.

Rundus, D. Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 43-50.

Schank, R. C. The structure of episodes in memory. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Slamecka, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology,* 1968, *76,* 504-513.

Slamecka, N. J. Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology,* 1969, *81,* 537-560.

Slamecka, N. J. The question of associative growth in the learning of categorized material. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 324-332.

Wolff, P., & Levin, J. R. The role of overt activity in children's imagery production. *Child Development,* 1972, *43,* 537-548.

Wolff, P., Levin, J. R., & Longobardi, E. T. Motoric mediation in children's paired-associate learning: Effects of visual and tactual contact. *Journal of Experimental Child Psychology,* 1972, *14,* 176-183.

16 Event-Related Potentials: Approaches to Cognitive Psychology

Emanuel Donchin

Jack B. Isreal
Cognitive Psychophysiology Laboratory
University of Illinois, Champaign-Urbana

INTRODUCTION

The welcome revival of mentalism in psychology did not eliminate one of the important constraints imposed on psychology during the behavioristic interlude. It remains the case that the data on which theories and models rest must be derived from the measurement of observables. Psychological discourse may concern images, mental rotations, transfers from short- to long-term memory, or unconscious pathway activations. Yet the investigator in each case must anchor these concepts by recording, measuring, and analyzing such tangibles as the narratives related by subjects, the reaction times, or the number of errors committed during task performance. The methodological basis of cognitive psychology, then, is as dependent on the availability of overt indices—that is, on subjects' responses—as was that of the most thoroughgoing and strict behaviorists. In this context it is important to note that the repertoire of observables to which cognitive concepts are coordinated is not overly rich. It is, of course, rich enough to support the explosive development of cognitive science in the last two decades. Yet when enumerated, the class of available responses seems rather meager. Subjects can speak or write, and the content and form of their discourse can be examined. Subjects can also manipulate devices, and the speed and accuracy with which they do so can be monitored. Saying this, we have exhausted the repertoire. Pachella (1974) has succinctly stated the problem: "The events of interest to a Cognitive Psychologist usually take place when the subject is not engaged in any overt activity. They are events that often do not have any overt behavioral component. Thus, reaction time is often chosen as a dependent variable by default: there simply isn't much else that can be measured [p. 43]."

47

Although the response time can be easily recorded, it is of little interest in and of itself. It is only because of what it reveals about the processes antecedent to it that the response is of any interest. For example, experimental manipulations that are presumed to affect the component processes additively are introduced, and their effects on the distribution of response times are examined so as to permit inferences about the processes (Sternberg, 1969). But the component processes are rarely, if ever, amenable to direct observation. It is precisely here that psychophysiology may be of use. A class of psychophysiological indices, known collectively as event-related brain potentials (ERPs), can serve, we think, as manifestations of some of the component processes. It may be possible to develop a cognitive psychophysiology in which "behavioral" and psychophysiological measures are used in combination to elucidate cognitive processes. It is our purpose in this chapter to review evidence that the incorporation of psychophysiological indices in the study of cognition strengthens the psychologists' armamentarium in a useful way. We argue this thesis using one of many ERP components, the so-called P300. This component has been studied in some detail at the Cognitive Psychophysiology Laboratory at the University of Illinois, so we can speak from a fairly rich data base. For more details and for a consideration of the entire gamut of ERP studies, the reader may consult the summaries presented in Callaway, Tueting, and Koslow (1978). In the following pages, we first review some of the background that is necessary for understanding ERP studies, and we introduce some of the ERP components. We then discuss data supporting the assertion that both the amplitude and the latency of the P300 component can be used to study the subject's cognitions.

EVENT-RELATED BRAIN POTENTIALS

The event-related brain potential is a transient response to specific events that is embedded within the human electroencephalogram (EEG). Without the aid of signal-extraction techniques, these potentials are difficult to detect in scalp recordings, as the magnitude of the ERP may be considerably smaller than the magnitude of the ongoing EEG. The development of digital signal averagers in the late 1950s (Clynes & Kohn, 1960) made it possible to obtain useful estimates of the ERP. Figure 16.1 illustrates the utilization of signal averaging to extract from the EEG the brain response to a tone. In panel A are shown several individual records of EEG. The records are aligned by the time of presentation of a brief tone. It is clearly impossible to see a consistent response to the tone in the individual records. In panel B, photographs of 60 of such single trials are superimposed. This superimposition technique was pioneered by Dawson (1954) in England. This procedure enhances aspects of the data that are consistent across the trials. Thus voltage changes that are time-locked to the stimulus should be emphasized by the superimposition. A consistent negative–positive pattern indeed appears immediately following the tone. A more refined estimate of the
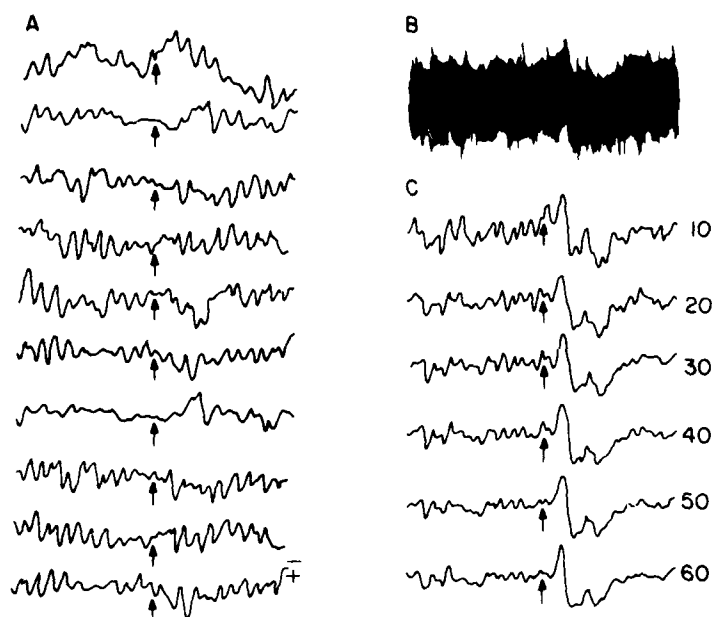
FIG. 16.1.   An illustration of signal averaging as applied to electroencephalographic data. (A) Raw EEG records from the vertex. An auditory stimulus was presented to the subject at the point indicated by the arrows. Note that no consistent response to the stimulus can be detected in these "single-trial" EEG records. (B) The superimposition of 60 single-trial records. The ERP waveform can be seen, but with low resolution. (C) Sequence of cumulative averages of the same data, successively adding 10 single trials to each average. Note the increasing clarity of the waveforms. (From Donchin, 1975.)

ERP is obtained when a signal average is computed, as seen in panel C. The voltage oscillations of the ongoing EEG are not time-locked to the tones and tend to average out, leaving the signal, or ERP. This noise reduction is proportional to the square root of the number of trials contributing to the average. The improvement in the estimate can be seen as additional batches of 10 trials are added to successive averages.

In examining the ERPs, it is important to recognize that it is inappropriate to refer to *the* evoked potential as if it were a unitary entity. Consider, for example, the array of studies conducted in the 1960s purporting to identify the relation between attention and ERP amplitude. Many experiments were reported, all sharing a similar design. Subjects were placed in conditions in which "attention" was known (or believed) to vary, and the ERPs were recorded; ERPs elicited by the same physical stimuli were obtained while the subjects were "attending" or "not attending" to the stimuli. "Attention" was reported by some investigators to cause increases in the amplitude of the ERP (Chapman & Bragdon, 1964; Debecker & Desmedt, 1966; Donchin & Lindsley, 1966; Ritter & Vaughan, 1969; Satterfield, 1965). Others, however, reported no effects of attention on the ERP or even decreases in ERP amplitude with attention (Hartley,

1970; Naatanen, 1967; Satterfield & Cheatum, 1964). ᵢhis confusing state of affairs may be attributed, in part, to a tendency to treat the ERP as if it were a global representation of the state of cortical tissue. Many investigators felt it was sufficient to report the overall "amplitude" of the ERP as if it did not matter which particular feature of the wave was modulated by the experimental variables. It has proven, however, far more fruitful to consider the ERP as a sequence of overlapping components, each possibly representing the activity of different populations of nerve cells and each standing in different, often orthogonal, relations to experimental variables.

The ERP elicited by a tone of adequate intensity can last for many hundreds of milliseconds and may contain many components. An example of the ERP elicited by a moderately loud click is shown schematically in Fig. 16.2. Seven small but very consistent waves will appear (after averaging thousands of trials) within the first 10 msec after the tone (Je·vett, Romano, & Williston, 1970). The next 50 msec will reveal four or five more oscillations, with larger amplitude. In later segments of the recording epoch, considerably larger potentials may appear.
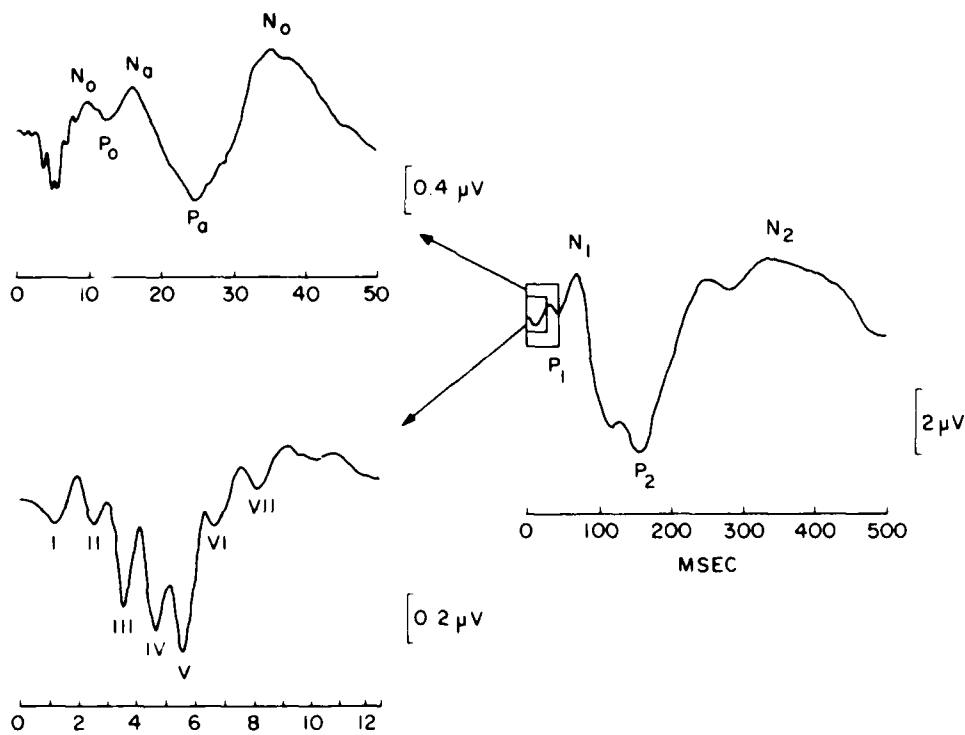
FIG. 16.2.   A schematic presentation of the configuration of the event-related potential elicited by a click of moderate intensity. Note the different time bases and different calibration signals in each of the three insets. Note also that different peak nomenclatures are used in the three cases and that component labelings for the data shown for the last 500 msec are different from those used in the rest of this chapter. (After Picton, Hillyard, Krausz, & Galambos, 1974.)

These later peaks and troughs are denoted by a letter and number combination. The letter (N or P) indicates the polarity; the number expresses the minimum latency of the peak (see Donchin, Callaway, Cooper, Desmedt, Goff, Hillyard, & Sutton, 1977). Note that negative-going peaks are displayed here as *upward* deflections.

Thus, the ERP is a sequence of components. We assume here that it is possible to record an ERP component because a population of neurons, either narrowly localized or dispersed within the cranium, has been synchronously activated in response to the specific needs of the information-processing system. That the activity of such a population can be recorded between electrodes placed on the scalp suggests that upon the synchronous activation of the elements, the geometry of these elements causes their field potentials to summate. This, of course, is only a working hypothesis. Yet in terms of this hypothesis, we can present our approach, in defining the vocabulary of the ERP, as the enumeration of the conditions under which specific populations of neurons are activated. We hope, of course, that the data we obtain will also serve as a guide for neurophysiological research that will identify the populations whose existence we postulate. However, for our present purposes, this physiological information is not necessary.

Investigators often see a component in each peak or trough in the ERP that appears with some regularity at specific points in time. Thus, in order to study the components, a measurement procedure is applied to each of the components. The experimental results are then described as *functional relationships* between the component measurements and the independent variables. This procedure, though useful, presents some difficulties. Donchin (1966, 1969) has commented on the inadequacy of visual inspection as the sole guide to component identification and on the disadvantages inherent in any technique that does not assure that the components are defined, identified, and measured objectively. (For a detailed discussion, see Donchin & Heffley, 1979.)

It is particularly important, in considering procedures for defining and measuring evoked-potential components, to realize that the positive and negative potential swings observed in the ERP waveform are not necessarily independent. It is conceivable, and indeed quite probable, that the scalp-recorded waveform is not produced by the linear summation of several independent generating processes. The degree of interaction among the various generating mechanisms cannot be assessed directly from measurements of peaks and troughs in the potentials. For example, investigators seem to encounter a persistent difficulty in determining whether component amplitudes should be expressed as baseline-to-peak, or as peak-to-peak, measurements. This problem has become especially recalcitrant with the increasing interest in "slower" potential shifts, such as the contingent negative variation (CNV), which led to an increasing use of the lower end of the frequency spectrum in the recording of ERPs. The fact that the faster positive-negative swings often ride on low-frequency components makes the definition and study of the ERP a very complex matter.

To measure components appropriately, one must have a clearly developed idea of what the components are. We view a component as a set of potential changes that can be shown to be *functionally related to an experimental variable* or to a combination of experimental variables. A component can be assumed to exist only if it has been shown to vary systematically as a function of some independent variable. Given this definition, changes in a component must be uncorrelated with the effects that a given experimental variable has on other ERP components. Thus, rather than defining components in terms of peaks and troughs in the waveform, using the morphology of the wave as our primary datum, we dissect the morphology in terms of manipulated experimental variables. Different neuronal aggregates might be activated at different times to different extents by different values of our critical variables. Other neuronal aggregates may or may not be activated during our recording session; but to us, they are transparent.

A useful heuristic distinction can be made between two categories of ERP components, the *exogenous* and the *endogenous* (Sutton, Braren, Zubin, & John, 1965). The early components elicited by auditory stimuli mentioned earlier are typical exogenous components. That is, they represent the response of brain tissue to the activation of a peripheral sense organ by an external event. They are *obligatory* responses to stimuli. If a stimulus is presented to a living person with an intact auditory system, these potentials will invariably appear. In fact, if these potentials do not appear, we can assume the person to have some hearing loss (Davis, 1976). The exogenous components are very sensitive to the sensory characteristics of the eliciting stimulus. Their form and their distribution on the scalp are quite dependent on the modality of the stimulation and relatively independent of psychological variables such as attention and expectancy (Regan, 1972).

The exogenous components are often followed by endogenous components. These components are not obligatory responses to stimuli. They are manifestations of the cortical information-processing activities invoked by the demands imposed by the subject's task. The variance in the characteristics of the endogenous components (i.e., their amplitudes, latencies, and scalp distributions) is normally accounted for by variation in the tasks assigned to the subject (see Donchin, 1979, and Donchin, Ritter, & McCallum, 1978, for more detailed discussions of the componential approach to ERPs).

THE P300 COMPONENT AND HUMAN INFORMATION PROCESSING

We focus in this chapter on one endogenous component, the P300. The P300 is easily recorded in the "oddball" experimental paradigm that underlies many of the experiments described in this chapter. In the auditory version of this

paradigm, the subject wears headphones through which he or she hears a random sequence of 1500-Hz and 1000-Hz tones (or any discriminable stimuli). One of the two tones is presented frequently (say 80% of the time), whereas the other is presented rarely (20% of the time). The subject is asked to count (covertly) the number of times one of the two tones has been presented. It is invariably found that the rarer stimulus elicits a much larger P300 than does the frequent stimulus. If the probability of the rare stimulus is increased, or if the subject is instructed to perform a task that is unrelated to the tones, the difference between the ERPs elicited by rare and frequent stimuli is diminished (Duncan-Johnson & Donchin, 1977). An interesting characteristic of P300 is that its elicitation does not require the physical presentation of a stimulus. If the series of tones contains only one tone that is occasionally omitted, then the omission of the stimulus will elicit a P300 (Ruchkin & Sutton, 1973; Sutton, Tueting, Zubin, & John, 1967). This is a powerful demonstration that P300 *is* endogenous.

To understand the manner in which we have attempted to elucidate the nature of P300, we must first dispense with a somewhat outdated model of the human operating system that has confused some of the investigators attempting this task. This is the S-R (stimulus-response) model, which is illustrated in Fig. 16.3. This view considers the organism as inert at all times other than during the interval between a stimulus and the overt response it evokes. Responses are selected through a series of information-processing stages that are activated solely by the nominal stimuli. All information processing, accordingly, takes place in the interval between stimulus onset and response termination. Thus, to be considered as a manifestation of an information-processing activity within this framework, an ERP component must appear and terminate within this interval. The fact that the P300 may at times *follow* the execution of an overt response consequently casts doubt on the validity of any asserted relationship between P300 and information processing. A reasonable approach, more consistent with current thinking about the human information processor, is labeled in Fig. 16.3 as "cognitive." This view assumes that the subject brings strategies, memories, expectations, and so forth into any stimulus-response interaction (Pribram & McGuinness, 1975; Sokolov, 1969). He or she continues to process data delivered in the past. Memory is being reordered. Old, unsolved problems are treated. Stimuli presented on a trial interact with this stream and invoke a variety of serial and parallel processes. Several of these processes might lead to an overt response. Others will result in no overt response on the specific trial but will change a subject's strategies in ways that will be manifested only on successive trials. Human information processing is thus viewed as an ongoing process, not chunked into "trials."

As is discussed later, the P300 in some situations seems to index aspects of information processing that are opaque to traditional behavioral analyses of performance. Hence, findings of low correlations between the characteristics of P300 and behavioral criteria do not imply that the functional role of the process

S - R
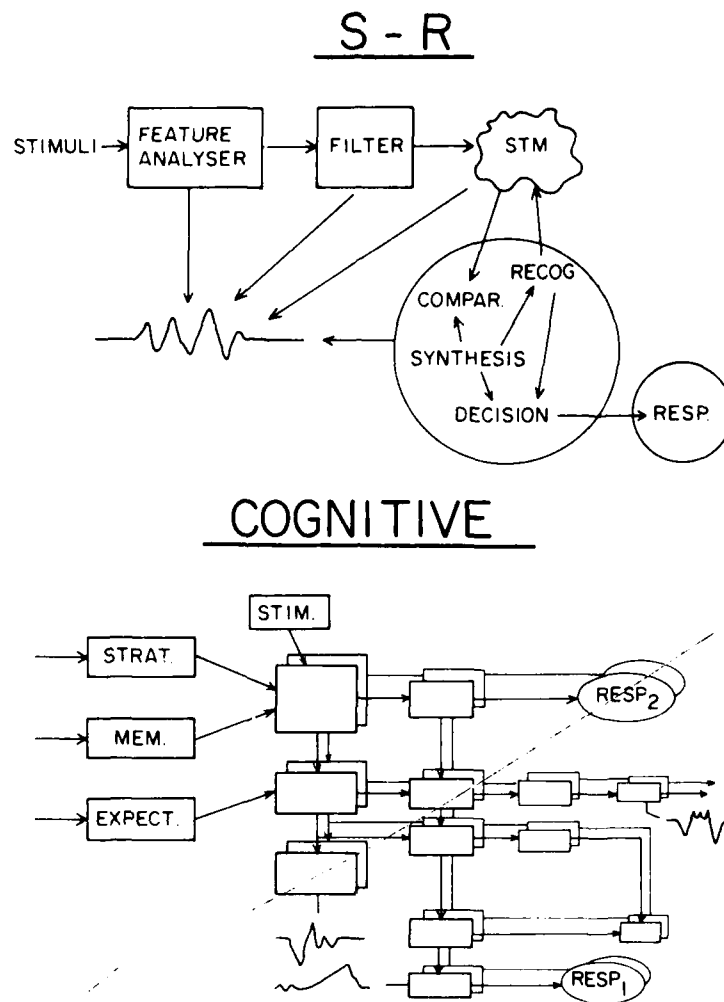
COGNITIVE

FIG. 16.3.   Two conceptualizations of the events that occur in an experimental trial. (From Donchin, 1979.)

underlying P300 is irrelevant to the problems addressed by cognitive psychology. Quite to the contrary, such dissociations signify that behavioral and psychophysiological criteria may provide complementary and often orthogonal information regarding the individual's interactions with the environment. In the following sections we review recent studies conducted in the Cognitive Psychophysiology Laboratory that have contributed to the understanding of P300 and, in doing so, have established the P300 as a valuable tool in the study of human information processing. For more detailed reviews of the current literature on P300, see Donchin et al. (1978); Picton, Campbell, Baribeau-Braun, and Proulx (1978); and Tueting (1979).

TASK RELEVANCE AND THE P300

One of the major factors determining the amplitude of P300 is the task relevance of the eliciting stimulus. This is illustrated in an experiment described by Heffley, Wickens, and Donchin (1978). The subject views a screen on which targets move slowly and in random directions. The targets traverse the screen, disappear, and then reappear at random locations, and this cycle then repeats. Half of the targets are small squares; the other half are small triangles. Every 6 to 8 seconds, the brightness of one of the targets is slightly increased for 200 msec. At any one time, then, the subject sees a number of moving targets on the display. The individual's task is to monitor one class of targets and count their intensifications while ignoring the other class. So if the subject is counting triangles, all intensifications of squares are irrelevant, and intensifications of the triangles are relevant stimuli. The experiment also varied the number of targets on the screen and the probability that a square or triangle would intensify on any one trial.

In Fig. 16.4 are shown the ERPs averaged over six subjects. There are two groups of averages—one elicited by the uncounted stimuli, and one elicited by the counted stimuli. The difference between the ERPs in the two groups is striking. The relevant stimuli elicit a considerably larger P300 than that elicited by the irrelevant stimuli. Note that these ERPs were elicited by essentially the same physical stimuli. There was almost no brightness difference between the triangles and the squares. Both stimuli appeared in the same region of the visual field at about the same time. Yet the amplitude of the P300 was enhanced when the stimulus was task relevant.

Task relevance has also been manipulated by instructing subjects to perform a task unrelated to the stimuli that are to be ignored. Under these conditions, no P300 is elicited (Courchesne, Hillyard, & Galambos, 1975; Duncan-Johnson & Donchin, 1977; Ford, Roth, & Kopell, 1976; Squires, Donchin, Herning, & McCarthy, 1977). A few investigators, however, have reported large P300s to be associated with unpredictable stimulus shifts in task-irrelevant stimuli (Ritter, Vaughan, & Costa, 1968; Roth, 1973; Roth, Ford, Lewis, & Kopell, 1976; Vaughan & Ritter, 1970). In all studies in which an "ignored" stimulus was reported to elicit a P300, the investigators instructed subjects to ignore the stimuli or else told subjects to perform an unrelated task (e.g., read a book), the performance of which was not measured. Such experiments illustrate what Sutton (1969) has termed "the role of subjects' options." Sutton was justifiably criticizing experimental designs that rely solely on instructions to produce complex psychological states, and in which no independent evaluation validates whether or not these states have in fact been produced. Chapman (1973) has also addressed this issue: "From the standpoint of studying the psychophysiology of thinking, it would seem advantageous to control *what* thinking takes place and

NOT
COUNTED

COUNTED

Counted / Not	Counted	
#	Prob.	
———	2/2	50/50
— — —	2/2	20/80
—·—	2/8	50/50
··········	2/8	20/80
———	5/5	50/50
— — —	5/5	20/80

IO μv

+

FLASH
ONSET

0.0   0.5   1.0   1.5   2.0

TIME (sec)
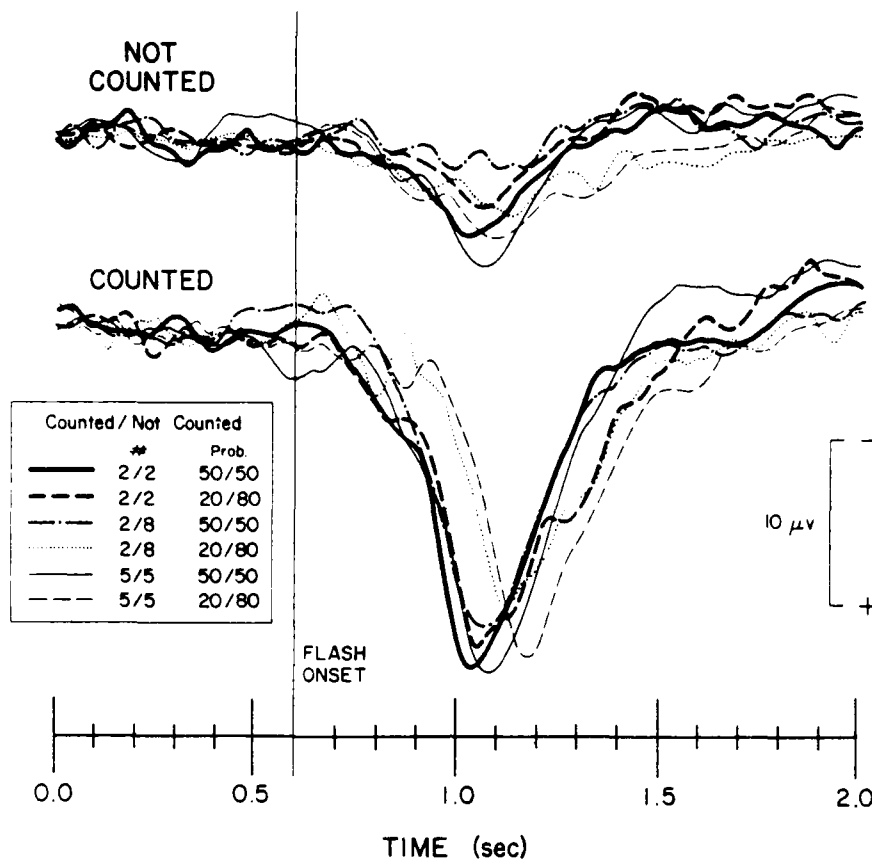
FIG. 16.4. ERPs elicited by relevant (counted) and nonrelevant (uncounted)
target intensifications averaged over six subjects, for various probability and dis-
play load conditions. Note the marked difference in P300 amplitude between ERPs
elicited by relevant and nonrelevant events. Note also the effect of display load on
P300 latency.

when it takes place. . . . Averaged evoked potential experiments often appear to
have these two characteristics, but unfortunately, both the *what* and the *when* in
many experiments have lacked precision [p. 70]."

A recognition of the subjects' control over their options leaves the experi-
menter two alternatives. He or she may either attempt to control the subject's
information-processing strategies or may allow them to vary freely, provided that
means are available for evaluating the nature of the subject's strategies. These
two approaches are illustrated in a study reported by Johnson and Donchin
(1978), who attempted to clarify some theoretical difficulties presented by
Adams and Benson (1973). Adams and Benson reported that the amplitude of
P300 elicited by a stimulus that indicated to the subject successful task perfor-
mance (S +) varied with the intensity of the corresponding failure indicator (S −).
Adams and Benson used a 30-dB SL tone as an S +, whereas S − varied in

intensity from series to series. (The sensation level |SL| of a tone refers to its pressure level in decibels above its threshold of audibility for a particular individual or for a specified group of subjects.) The smaller the difference between S+ and S−, the smaller was the P300 amplitude elicited by S+. These data were puzzling because there was no apparent change in either the probability or the task relevance of the S+ with the changing S−. Johnson and Donchin (1978) replicated and considerably extended this experiment, arriving at altogether different conclusions. The prime differences between Johnson and Donchin's experiment and that run by Adams and Benson were that: (1) the new experiment examined the ERPs to both S+ and S−; (b) the experiment evaluated the extent to which the effect depended on the intensity difference between S+ and S−, rather than on the absolute intensity of S−; and (3) the subject's performance was evaluated to determine if task performance was sensitive to the same variables that affected P300. The results are quite conclusive in showing that the effect depends entirely on the intensity *differences*. Moreover, the subject's task performance (time estimation in this case) deteriorated with decreasing difference between the intensities of S+ and S−. Johnson and Donchin interpreted the data as follows: The P300 indexes the task relevance of the stimulus, which in turn depends on the feedback value of a stimulus. Feedback here is interpreted rather strictly as referring to the extent to which the consequences of past actions can affect future performance (Donchin, 1975). Thus a stimulus providing data about task performance can only be considered as ''feedback'' if it, in fact, has an effect on subsequent performance. As subjects' time estimation deteriorated with decreased S+/S− differences, it is plausible to assume that subjects relied less on the S+/S− as they become less discriminable. This will lead to a degradation of the time estimation performance, at the same time yielding a smaller P300. In short, the larger the feedback value of the stimulus, the larger the P300 is. This interpretation was buttressed by data acquired while the subjects were counting the number of S+ stimuli, rather than estimating a time interval. The counting task requires that the necessary discriminations be made, leaving the subject with no option. The P300s in these conditions were of equal amplitude at all levels of S+/S− differences.

We are implying here a definition of task relevance that depends on subject options as much as it depends on the experimenter's instructions. The defining framework is the subject's task. Data needed for successful task performance are carried by stimuli that the subject may, or may not, so process. A stimulus is task relevant to the extent that the subject is processing it so as to increase his or her potential success in performance.

The evidence that P300 is related to task relevance suggests that stimuli must in some way engage the subject's attention for P300 to be elicited (Donchin et al., 1978). The Johnson and Donchin (1978) experiment demonstrates that this engagement is not an all-or-none affair. This conclusion is entirely consistent with current views of the processing system that conceive of attention as a

processing resource of limited supply, which may be allocated in graded quantities to processing activities (Kahneman, 1973; Navon & Gopher, 1979; Norman & Bobrow, 1975). Isreal, Wickens, Chesney, and Donchin (1980) investigated the extent to which P300 amplitude is affected by variation in the available supply of processing resources. The oddball paradigm described earlier was used as a "secondary task" (Kerr, 1973; Posner & Boies, 1971; Rolfe, 1971). However, this time the tone series served as probes, presented concurrently with each of two variants of the aforementioned display-monitoring task developed by Heffley et al. (1978). The P300s elicited by the probes were examined to reveal the effect of the difficulty, or resource demands, of each primary task. As in the Heffley et al. (1978) experiment, the subject viewed a screen on which either four or eight targets moved about. Half of the targets were squares, and half were triangles. Every 6 to 10 seconds, one of the targets (either a square or a triangle) briefly increased in brightness. Also every 6 to 10 seconds, one of the targets changed its direction of movement. In one set of conditions, the subject's task was to monitor one class of targets (squares) and to signal their intensifications with a button press. In other conditions, subjects monitored the squares for changes in their courses. Task "difficulty" was varied by manipulating the number of targets to be monitored. Figure 16.5 shows the ERPs elicited by the probes, averaged across eight subjects, during each monitoring task as well as a count-only control condition. It can be seen that introducing either of the display-monitoring tasks leads to a substantial reduction in P300 amplitude.
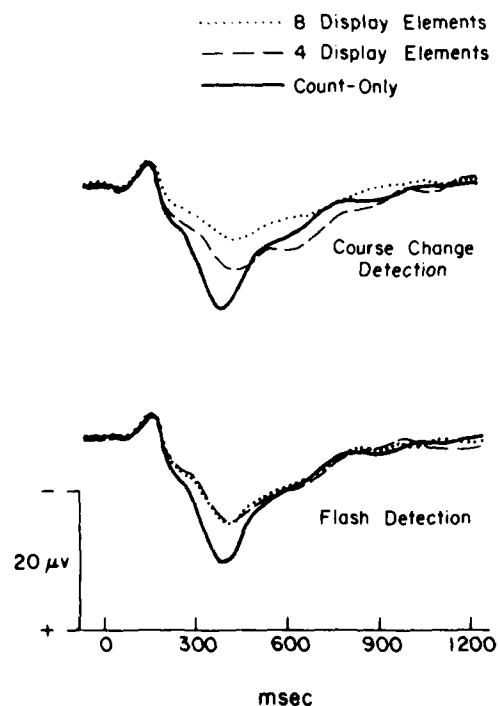
FIG. 16.5. ERPs (averaged over eight subjects) elicited by counted probe tones during the concurrent performance of two display-monitoring tasks. Note how P300 amplitude is reduced when either visual task is introduced, and note also the further attenuation of P300 amplitude accompanying an increase in the number of targets to be monitored for "course changes."

Increasing the number of targets to be monitored diminished P300 further only when subjects were detecting course changes. Visual detection accuracy and performance on a secondary reaction-time task paralleled the P300 amplitude results in showing that whereas the course-change monitoring task was made more difficult by increasing the number of targets, this manipulation was ineffective in changing the difficulty of the intensification-detection task. Thus P300 amplitude seems to be a sensitive and valid index of the perceptual processing demands of a concurrent task. (For descriptions of other ERP experiments utilizing secondary tasks, see Isreal, Chesney, Wickens, & Donchin, 1980; and Wickens, Isreal, & Donchin, 1977.)

SUBJECTIVE PROBABILITY AND P300

Although task relevance plays a major role in determining the amplitude of P300, it is by no means the only controlling variable. There is considerable evidence that the magnitude of P300 elicited by task-relevant events depends on the subjective probabilities of the events. Consider the data reported by Duncan-Johnson and Donchin (1977). In this experiment, the oddball paradigm was used. Subjects heard a series of high- and low-pitched tones and were instructed to count the high tones. The prior probabilities associated with the tones were varied over a wide range, from .10 to .90. Some of the ERPs obtained in this experiment are shown in Fig. 16.6. Regardless of whether frequent or infrequent stimuli were counted, the rare stimulus elicited a large P300, and the frequent stimulus produced a smaller P300. And as the prior probability of the infrequent stimulus increased, the P300 it elicited decreased, and the P300 to the other stimulus proportionately increased. Note that when the subject was asked to solve a word puzzle and the tones became irrelevant, the potentials elicited by the rare tones and by the frequent tones were not significantly different; both showed no P300. Thus the magnitude of P300 appears to index the "surprise" value of a stimulus.

Can we determine from a single trial whether the subject was presented with a rare or with a frequent event? Success in this endeavor depends, in part, on the ratio of the magnitude of the P300 to the power of the ongoing EEG. Fortunately, for P300, this ratio is favorable, and a single-trial analysis of P300 is quite feasible. This we accomplish through a stepwise discriminant analysis (Dixon, 1975; Donchin & Heffley, 1979; Donchin & Herning, 1975). We know that on the average, the rare stimuli elicit a large P300, whereas the P300 elicited by the frequent stimuli is small. Using this information, a stepwise discriminant analysis is used to develop a classification rule with which we can classify each trial of new data into one of the two categories. Squires and Donchin (1976) showed that about 80% of the trials can be correctly classified as being either rare or frequent. The formulation of the question in terms of the *external* stimulus, however, is somewhat misleading. It is inconsistent with the assumption that the
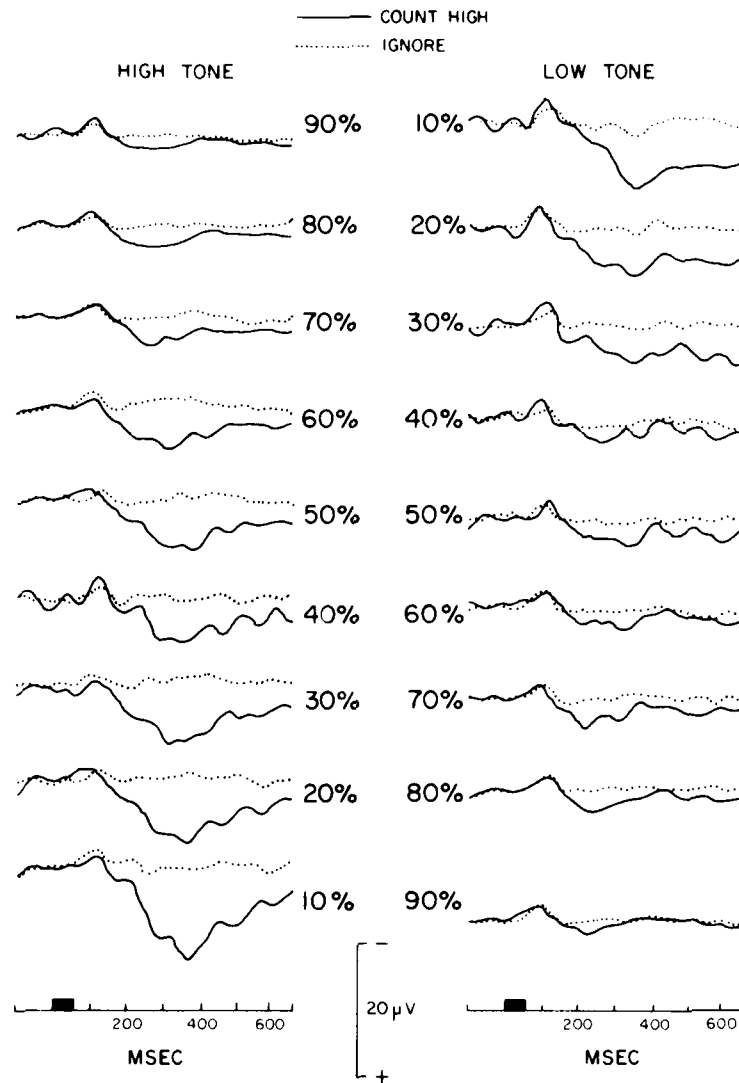
FIG. 16.6.   Averaged ERPs from an experiment in which the probabilities of tones in a Bernoulli series were parametrically manipulated. Subjects either counted the high-frequency tones (solid lines) or solved a word puzzle (dotted lines). P300 amplitude varies systematically with the prior probability of the tones only when they are task relevant. (From Duncan-Johnson & Donchin, 1977.)

P300 reflects an internal, endogenous response of the subject to the stimulus rather than the properties of the external stimulus. The 20% ''error'' may be due to different responses by the subjects to the same stimuli. Rare stimuli may, at times, be perceived as frequent, and vice versa. In fact, if the trials that were ''misclassified'' are averaged, it turns out that rares classified as frequents do not elicit a P300 but that frequents classified as rares do. One explanation for these

data is that subjects indeed respond to some frequent stimuli as if they were rare and to some rare stimuli as if they were frequent.

A search for a rule that will govern this trial-by-trial variability in the subject's behavior led to the consideration of the effects that stimuli presented on preceding trials have on the ERP elicited by each stimulus. Squires, Wickens, Squires, and Donchin (1976) reported an oddball experiment in which 50% of the tones were high-pitched and 50% were low-pitched. The average of all the high tones yielded a small P300. However, when the high tones were classified into two groups, one containing high tones preceded by a high tone, the other containing high tones preceded by a low tone, it was found that the high tones elicit a much larger P300 when they are preceded by the low tones. These data are summarized in Fig. 16.7. A measure of the amplitude of the P300 is plotted against the length of the sequence of preceding stimuli. It can be seen that there is a systematic relationship between the amplitude of the P300 elicited by the stimulus on a given trial and the sequence of stimuli that precede it. Furthermore, if the prior probabilities of the two stimuli are altered so that they are .30 and .70, the P300 amplitude is displaced upward for the rare event and downward for the frequent event. Yet within that relationship, the sequential effect is still quite clear.

A model that attempts to account for this P300 amplitude variance is shown in Fig. 16.8. This model states that the amplitude of P300 depends on the "surprise" value of the stimulus, which is the reciprocal of the expectancy, or subjective probability, associated with the stimulus. That is, the amplitude of the P300 elicited by a high tone depends on the expectancy for a high tone, which in turn depends on how often and how long ago a high tone was previously presented. The model postulates that expectancy is proportional to a combination of three factors: the prior probability of the event, a "memory" factor, and an "alternation" factor. The prior probability factor accounts for the relative displacements of the sequential dependency trees shown in Fig. 16.7. The memory factor assumes that whenever a high (or low) tone is presented, the subject expects it to repeat. Forgetting occurs at an exponential rate, however, so that a stimulus presented further in the past will contribute less to expectancy than will a more recently presented stimulus (M_2 vs. M_3 in the lower portion of Fig. 16.8). The expectancy at any time for a high tone is equal to the sum of the expectancies contributed by all previous presentations of a high tone. If the solid bars represent occurrences of a high tone in the example portrayed in Fig. 16.8, then the expectancy for a high tone on Trial 5 would equal $M_2 + M_3$. The "alternation" factor was included in the model because in the special case of alternating series of events (e.g., ABABA or BABAA), subjects seem to "chunk" alternating pairs of stimuli into single units and expect these units to recur. Under these circumstances, then, a stimulus repetition (e.g., BABAA) elicits a larger P300 than a stimulus alternation (ABABA) because subjects are basing their expecta-
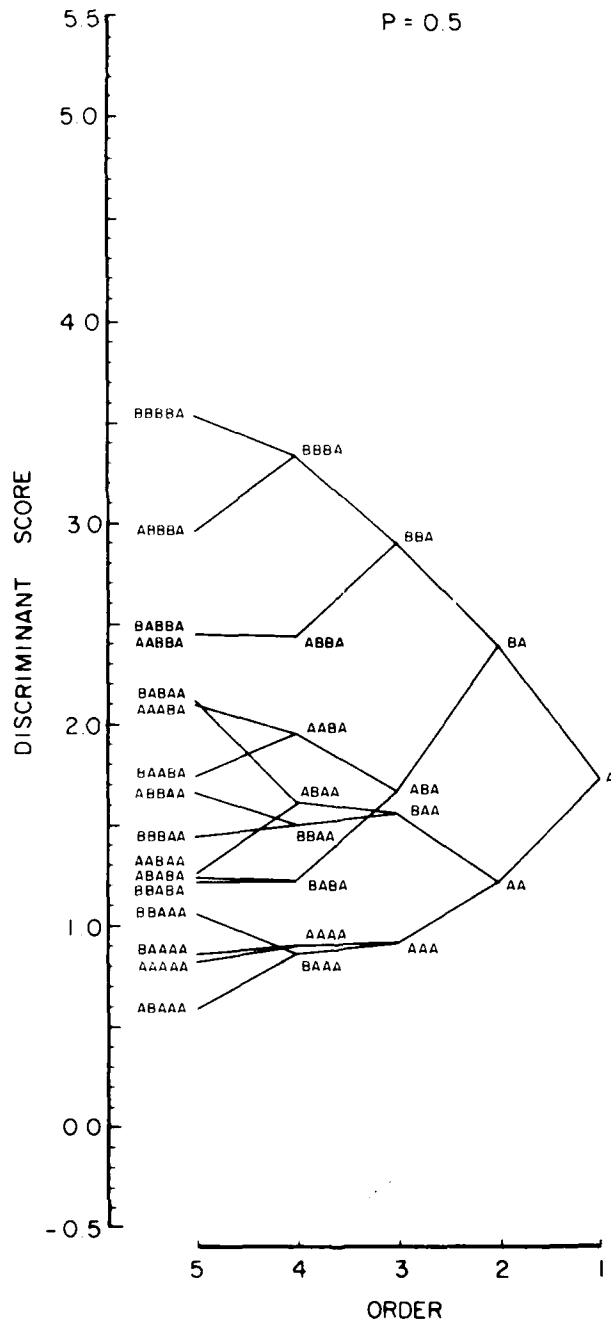
FIG. 16.7.   Tree diagrams of P300 amplitude (measured in terms of discriminant (scores) as a function of the preceding stimulus sequence. Each tree corresponds to a different prior probability of the counted stimulus .4. (From K. Squires et al., 1976.)
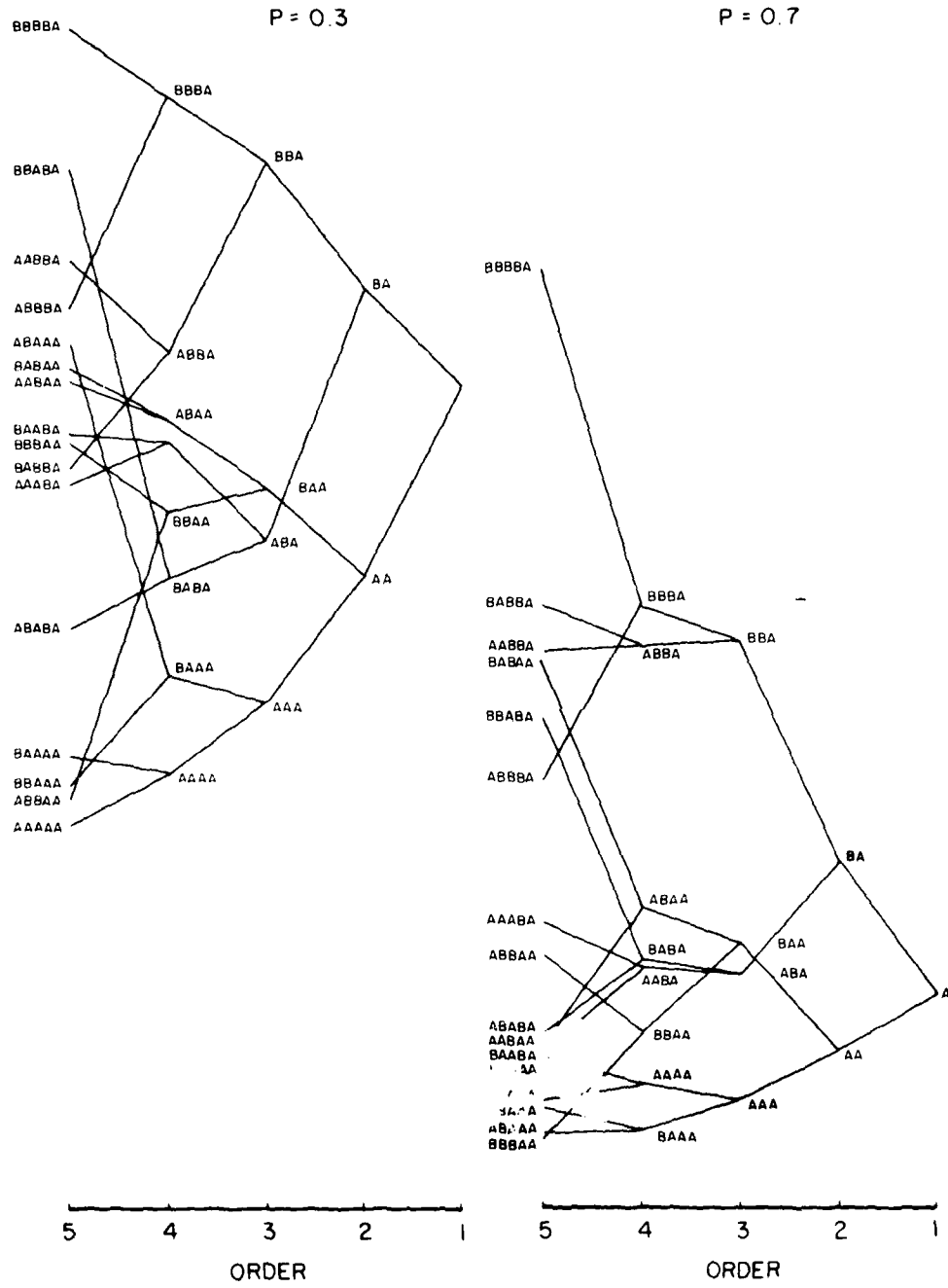
FIG. 16.7.   *continued*

tions on higher-order stimulus integrations. To some extent, the degree to which past events influence the amplitude of P300 to any given stimulus can be used to measure the "memory capacity" of an individual at a given point in time (see Wickens et al., 1977). Note that this is not a model about ERPs. It is a model about the manner in which people process information and develop expectancies. It is a model about cognition that is validated using ERP data, given the assumption that P300 is an index of surprise.

It can, of course, be argued that the variations of P300 with stimulus sequence do not reflect a cognitive process but rather reflect peripheral, or adaptational, interactions between stimuli. To address this argument, we carried out a number of experiments. All experiments used the same strategy: Similar stimulus sequences were presented, but the instructions to the subject were varied. By varying the instructions, the manner in which the events are to be categorized is affected; if the sequential dependencies we observed in P300 amplitude are related only to the physical-stimulus series characteristics, and are independent of instructions, then it is unlikely that we are dealing with a cognitive process. If for the same physical-stimulus structure, the sequential effects are modified by the instructions, then it cannot be argued that it is the physical-stimulus sequence alone that produces changes in P300. For example, Johnson and Donchin (1980) used a variant of the oddball paradigm in which the stimulus series was constructed from three equiprobable stimuli—a 1000-Hz, a 1400-Hz, and an 1800-Hz tone. The subject was instructed to count only one of the three stimuli. The question was whether P300 amplitude would vary as if there were three equiprobable categories defined by the physical stimuli, or rather as if the "uncounted" category had a .67 probability and the "counted" category had a .33 probability. The data show that the two uncounted events are treated by the subject as if they come from a single category whose probability is .67. It appears, then, that the *category* to which stimuli belong, rather than their physical properties, determines the sequential effect. Thus the P300 seems to manifest processing that occurs *after* the subjects have categorized the physical stimuli into classes determined by the particular tasks.

Another study, described by Duncan-Johnson and Donchin (1978), gives additional support to this view. Series of paired stimuli, separated by a 400-msec interval, were presented to subjects tachistoscopically. The first stimulus in each pair (S1) was either a star, the letter *H*, or the letter *S*. The second stimulus (S2) could be either the letter *H* or the letter *S*. Two different types of conditions were compared in this experiment. In one, the three S1 warning stimuli were presented with equal probability and provided no probabilistic information concerning which S2 would be presented. Thus the series of S2 stimuli constituted a Bernoulli series. Figure 16.9a shows the expectancy trees from this control condition. The estimates of the probability of S2 in this case develop on the basis of the past history of the sequence of S2 stimuli, repli ating the results already de-

Expectancy Model

$$E \propto M + P + A$$

E = Expectancy
M = "Memory" Factor
P = Global Probability
A = Alternation Factor

$$M_{AN} = \sum_{i=N-1}^{N-m} \alpha^{N-i} S_i$$

where

$$S_i = \frac{0 \text{ for } (S_i = B)}{1 \text{ for } (S_i = A)}$$
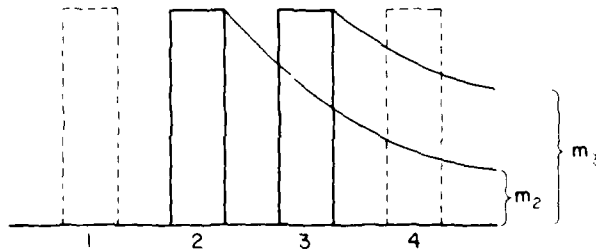
FIG. 16.8. A model that attempts to account for variation in a subject's expectancy that a given stimulus will occur on a given trial of a Bernoulli sequence as a function of the specific sequence of preceding stimuli and the prior probability of the stimuli. This model is described by K. Squires et al. (1976). (From Donchin, 1979.)

scribed for Bernoulli series. In the second set of conditions, the warning stimulus on each trial indicated the likelihood that the second stimulus would be an *H* or an *S*. In this case, the past history of the stimulus sequence is not needed for developing the probability estimates, and consequently, the effect of stimuli on the ERPs elicited by succeeding stimuli should be greatly diminished. The results from these conditions, shown in Fig. 16.9b, confirm these predictions. The amplitude of the P300 and, by inference, the expectancies for the stimuli depend on the probabilistic information conveyed by the warning stimulus, rather than on the sequential structure of the series. Thus Figs. 16.9a and 16.9b demonstrate that in the absence of explicit information on which subjective probabilities can be based, subjects will derive inferences from the past history of the stimulus

(a)
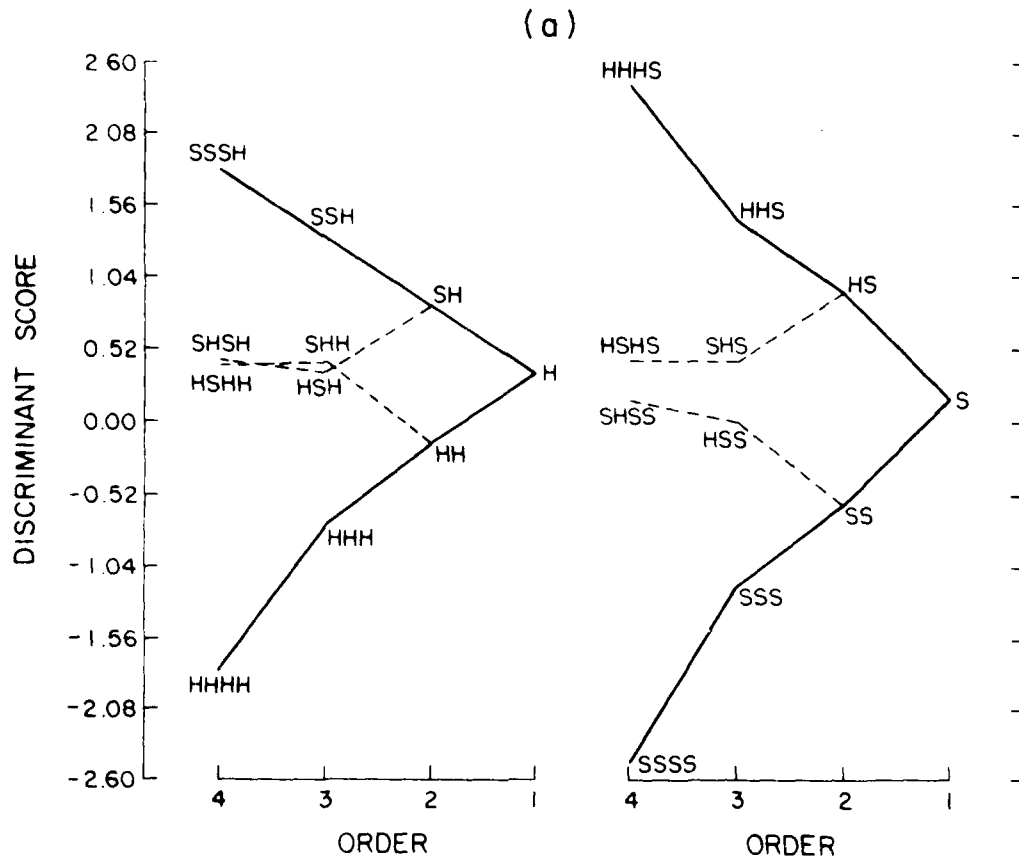
FIG. 16.9a.   Expectancy trees showing amplitudes of P300s elicited by *H*s and *S*s (presented visually) that were preceded by a warning stimulus that provided the subject with no probabilistic information. The amplitude of the P300 elicited by a stimulus is affected by the sequence of stimuli presented on preceding trials. (From Donchin, 1979.)

sequence, and a sequential dependency tree will appear; if they receive the information by other means, the "tree" will collapse. These results cannot be explained in terms of receptor adaptation or habituation. We seem to be tapping the manner in which individuals utilize external information, which may be related or unrelated to objective event probabilities, to impose a subjective probability structure on the environment.

In all the studies we have described, the eliciting stimuli were rather simple tone bursts. The relation between P300 amplitude and the surprise value of the stimulus holds, however, for complex stimuli in fairly elaborate paradigms. Consider, for example, a study reported by Horst, Johnson, and Donchin (1979). The subjects were assigned a paired-associates learning task. They had to learn,

(b)
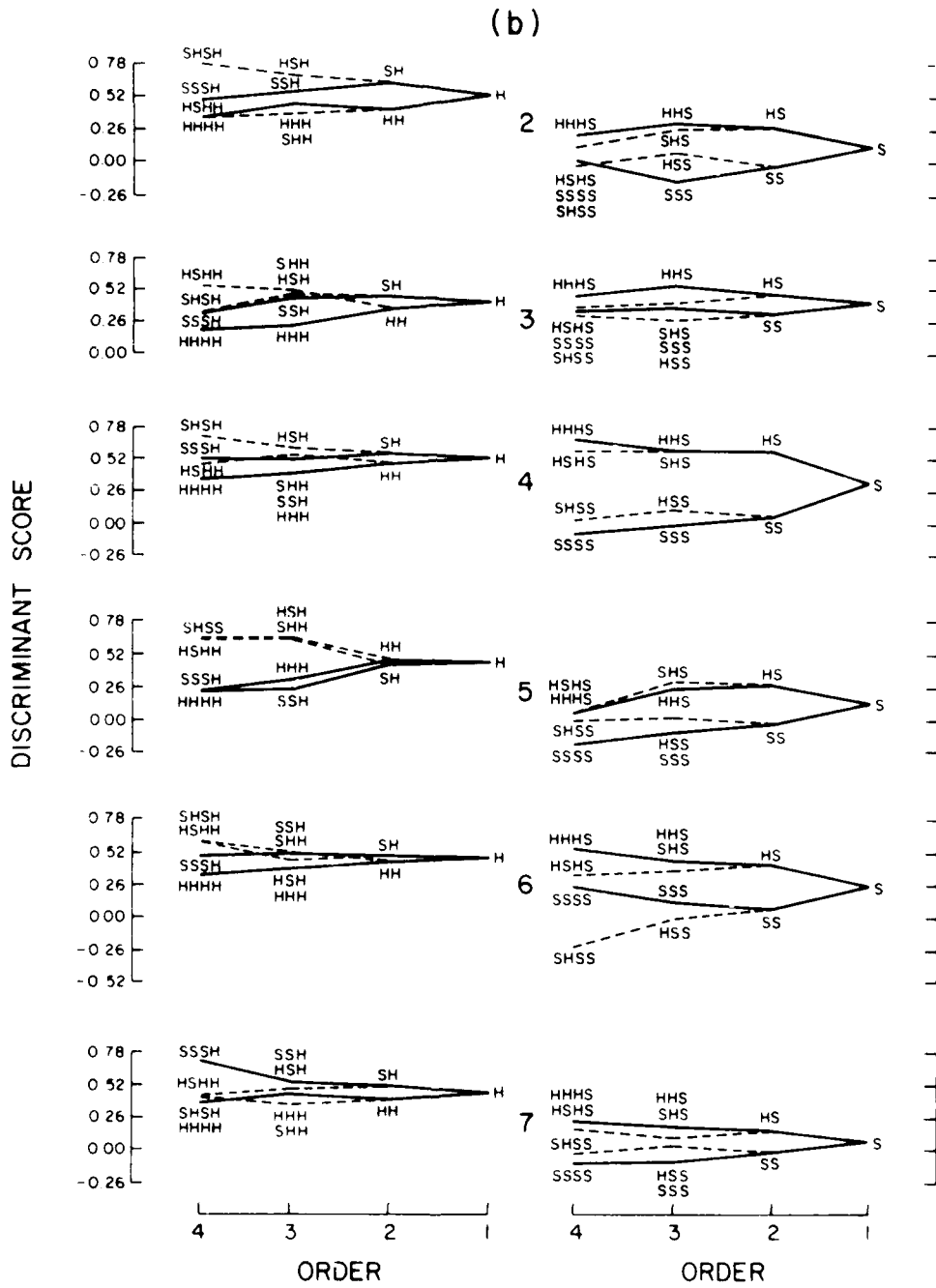
FIG. 16.9b.   Expectancy trees from six conditions in which the warning stimulus preceding each stimulus indicated the probability with which an *H* or *S* would occur. The sequential effects virtually disappear. (From Donchin, 1979.)

for example, that every time they were presented with the syllable *wok*, they had to respond with *zok*. A subject had to learn the "correct" responses to the stimulus syllables during the course of the experiment. When the subjects were presented with the syllable *wok*, they typed what they believed to be the correct response. They also gave confidence judgments ranging from 1 to 100. These confidence judgments were grouped into four ranges of confidence that were comparable across subjects. A confidence judgment in range 1 indicated that the subject was absolutely certain that his or her response was incorrect, whereas a judgment in range 4 meant absolute certainty that his or her response was correct. After responding, the subject was presented with the correct response syllable. The ERPs elicited by these "feedback" syllables were averaged according to the subjects' response-confidence levels. Figure 16.10 shows that if the subjects believed their response to be incorrect, and the feedback following the response indicated that they were instead correct, a bigger P300 was elicited than if the feedback indicated that they were indeed incorrect. On the other hand, when the subjects indicated that they thought their responses were correct and the feedback contradicted them, the P300 was larger than it was when the feedback confirmed their beliefs.

The relationsnips between confidence judgments, feedback, and P300 amplitude are perfectly consistent with the interpretation of P300 as reflecting surprise. We assume that the confidence judgment accurately reflects the subjects' expectations concerning the outcome of the trial. Therefore, feedback that confirms these expectations elicits a smaller P300 than does feedback that disconfirms expectations. In this situation, then, confidence judgments bear a direct relationship to expectancy. Consider, however, a situation in which subjects are presented with a Bernoulli series and are asked to predict, prior to each trial, which stimulus they believe is most probable. On the assumption that such predictions are also valid indicators of expectancy, stimuli that disconfirm pre-
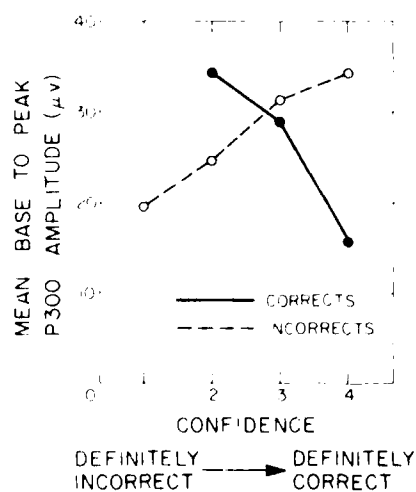
FIG. 16.10.   Amplitudes, averaged over five subjects, of P300s elicited by "correct" and "incorrect" feedback syllables as functions of subjects' response confidence. Correct feedback elicits a larger P300 when associated with low response confidence, whereas the reverse is apparent for feedback indicating an incorrect response.

dictions should elicit a larger P300 than should stimuli that confirm predictions. The evidence for this relationship is quite inconsistent, however. Whereas disconfirmations elicit a larger P300 in some experiments (e.g., Chesney, 1976; Tueting, Sutton, & Zubin, 1971), in others there is no difference or even the opposite effect (e.g., Levit, Sutton, & Zubin, 1973). Either P300 does not reflect surprise, or the assumption that subjects' predictions faithfully reflect their expectations is faulty.

Recent experiments by Chesney and Donchin (1979) shed light on this issue. Subjects were presented a Bernoulli series of crosses and squares on a display screen. In one condition subjects were required to count the number of occurrences of one of the characters, as in the oddball studies previously discussed. In another condition subjects had to predict before each stimulus whether it would be a cross or a square. Thus on every trial in this condition, the stimulus either confirmed or disconfirmed the subject's prediction. The sequential effects on P300 are illustrated for each condition in Fig. 16.11. In this figure, A represents a cross, and B represents a square. Regardless of which task the subjects were performing, the amplitude of the P300 elicited by a cross or square is strongly affected by the structure of the preceding stimulus sequence. However, when the trials were sorted for averaging according to the preceding sequence of prediction outcomes (i.e., confirmation or disconfirmation), no sequential effect is obtained, as shown in the prediction-outcome portion of Fig. 16.11.

The previous failure to demonstrate a consistent relationship between P300 amplitude and prediction outcome was also confirmed in this experiment. Fig. 16.12 shows that stimuli that confirm predictions elicit a P300 of magnitude equal to the P300 elicited by stimuli that disconfirm. This implies that subjects are, on the average, as surprised to find that their predictions have been confirmed as they are to find that they have been disconfirmed. If the subjective probabilities are governed by our expectancy model, then subjects expect stimuli to repeat and should be surprised when stimuli alternate, *whether or not these expectancies are transmitted to overt predictions.* When the confirmations and disconfirmations are sorted according to the stimulus sequence, as in Fig. 16.12, we find that this is indeed the case. Disconfirmations elicit larger P300s only if the predictions are congruent with expectancy (i.e., a repetition prediction). On the other hand, if subjects predicted an alternation, then a confirmation of that prediction elicits the larger P300. In other words, P300 amplitude in this context varies as a function of the surprise value of the stimuli, which is determined by the stimulus sequence only.

The experimental literature is replete with references to a dissociation between the perceived expectancy of an event and the predictions declared by the subject, especially when the consequences of different types of errors are not equal (Beach, Rose, Sayeki, Wise, & Carter, 1970; Myers, 1976; Neimark & Shuford, 1959; Reber & Millward, 1968; Vlek, 1970). However, no index other than the subject's overt predictions has, until now, been available to assess his or her

GRAND AVERAGES FOR 6 SUBJECTS

——— DISCONFIRM
········· CONFIRM
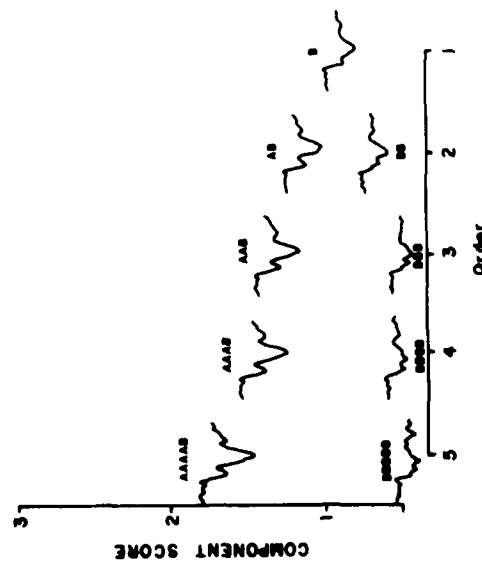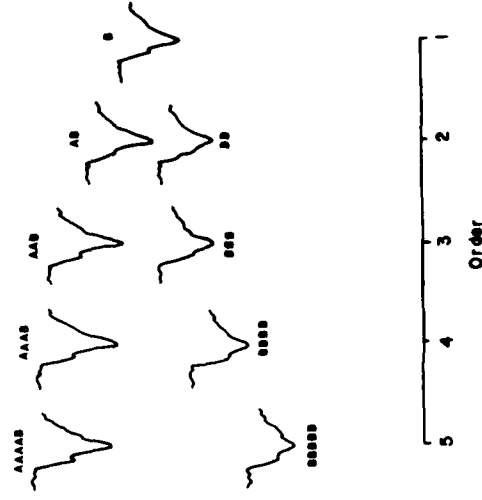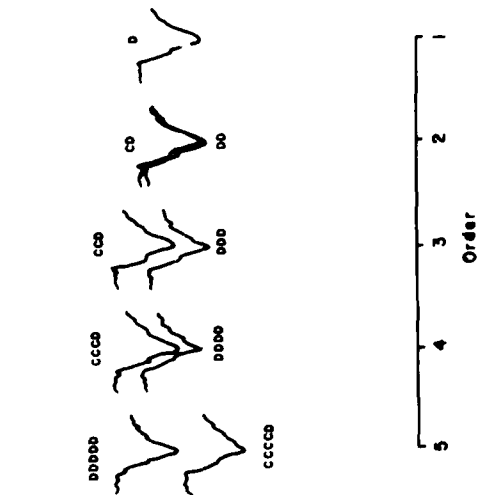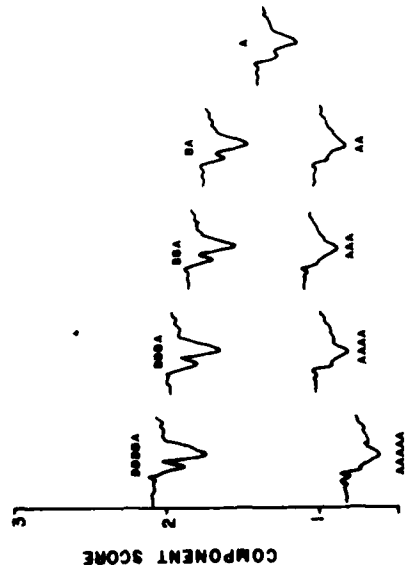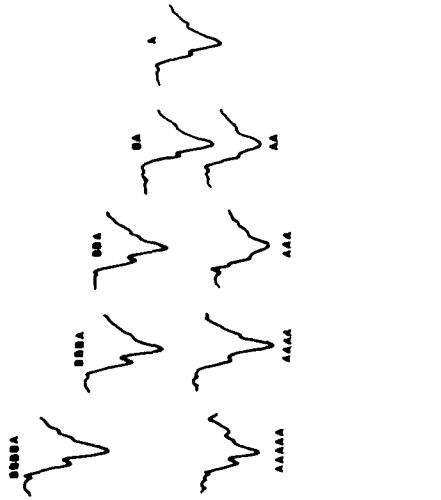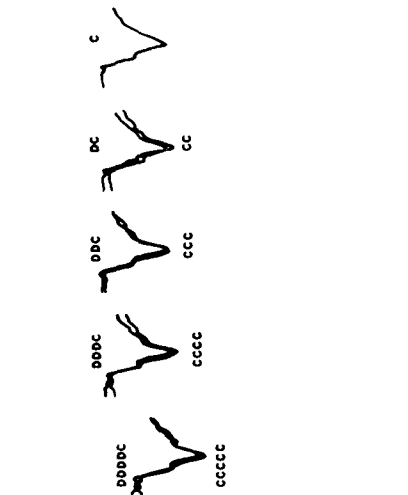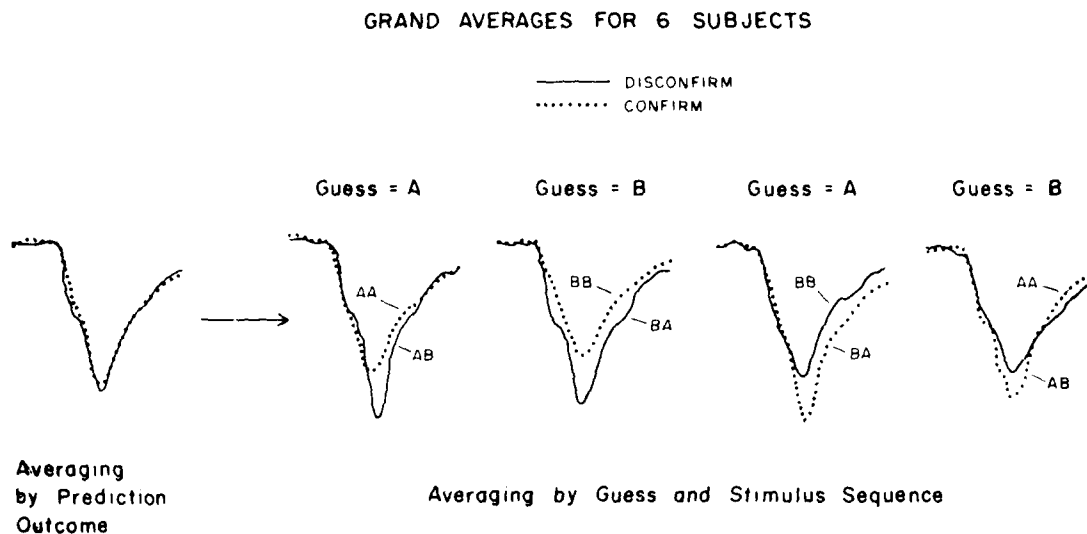
FIG. 16.12.   ERPs demonstrating that although the amplitude of the P300 component is the same whether the stimulus confirms or disconfirms the subject's prediction, large differences in amplitude obtain if the data are sorted by the specific prediction made by the subject and the degree to which these predictions are confirmed or disconfirmed. (From Donchin, 1979.)

expectancies, and so the existence of this dissociation has usually been inferred from contradictions in the data rather than from direct observables (see Messick & Rapoport, 1965). It may at first seem puzzling that when subjects are asked to indicate which of two events seems more likely to occur, they do not always do so in accordance with their expectations. This becomes less perplexing when prediction behavior is viewed as resulting from a multiplicity of factors, only one of which is subjective probability. Kahneman and Tversky (1973; Tversky & Kahneman, 1974) have identified various "heuristics" that may interact with the perceived probability to bias individuals' predictions. The major implication of the Chesney and Donchin results is that the P300 component may not be subject to the effects of these heuristics and may therefore serve in many situations as a more reliable indicator of subjective probability than subjects' overt prediction behavior. Thus the degree to which predictions are coupled to expectancy will depend on the degree to which heuristics are employed. The extent of this coupling will determine the relationship between prediction outcomes and P300 amplitude.

The studies we have reviewed illustrate the thesis we presented in the "Introduction." The amplitude of the P300 component is a manifestation of the activity

FIG. 16.11.   Expectancy trees for the "count" and "guess" conditions. Stimulus sequential effects on P300 are similar for the two tasks. Sequences of stimuli that confirm and disconfirm predictions do not affect the P300.

of an intracranial process. This process is of considerable psychological interest, as it reflects the degree to which stimuli engage the subject's attention. Furthermore, the amplitude of P300 can also serve as an index of subjective probability. By designing appropriate studies, it is possible to extend our understanding of the manner in which individuals allocate attention and of the rules whereby the subjective probability of events is determined.

RELATIONSHIPS BETWEEN P300 LATENCY
AND REACTION TIME

In the studies in which simple tones were presented to subjects, the peak amplitude of P300 followed the eliciting stimuli by 300 to 400 msec. However, it is not unusual to find P300 latencies on the order of 600 msec or longer (e.g., see Fig. 16.4). If P300 is elicited by surprising events whose category membership is defined by the task (see Johnson & Donchin, 1980), then the processing manifested by P300 cannot be invoked until the stimulus has been evaluated to the extent that it can be identified and categorized. Evidence that P300 latency depends on stimulus evaluation time has been reported by several investigators (Ford, Roth, Dirks, & Kopell, 1973; Gomer, Spicuzza, & O'Donnell, 1976; N. Squires, Donchin, K. Squires, & Grossberg, 1977).

A major implication of the hypothesis that P300 latency is proportional to stimulus evaluation time is that P300 latency and reaction time should be positively correlated. Although several investigators have found such a correlation (Bostock & Jarvis, 1970; Rohrbaugh, Donchin, & Eriksen, 1974; Roth, Kopell, Tinklenberg, Darley, Sikora, & Vesecky, 1975; Wilkinson & Morlock, 1967), others report a dissociation between P300 latency and reaction time (Karlin & Martz, 1973; Karlin, Martz, Brauth, & Mordkoff, 1971). In fact, as prior stimulus probability (Fitts, Peterson, & Wolpe, 1963; Hyman, 1953) and sequential relationships in stimulus series (Remington, 1969) are known to affect reaction time, several of the studies we reported earlier that found P300 amplitude effects of these variables without any corresponding latency variability (Duncan-Johnson & Donchin, 1977, 1978; K. Squires et al., 1976) represent further instances of a dissociation between P300 latency and reaction time.

A detailed analysis of the relationship between P300 and reaction time has been presented by Kutas, McCarthy, and Donchin (1977). If P300 latency represents stimulus evaluation time, then its relation to reaction time should depend primarily on the extent to which the subject's reaction time depends on stimulus evaluation. It is well known that subjects can trade speed for accuracy (Pachella, 1974). Speed, in this case, implies that responses are emitted without waiting for full evaluation of the stimulus. It can be predicted, therefore, that the correlation between reaction time and P300 latency would depend on the subject's strategy.

Kutas et al. (1977) tested this hypothesis by requiring subjects to discriminate between stimuli of varying degrees of complexity under both speed- and accuracy-maximizing regimes. They presented subjects with series of words on a screen. In different series, the words were varied. In one series, the names *Nancy* and *David* were presented on 20% and 80% of the trials, respectively. In order to assess the latency of P300 on each trial, Kutas et al. adopted a cross-correlation technique developed by Woody (1967). These latency measures could then be compared with the reaction times recorded on the same trials. Scattergrams of P300 and reaction time in the speed and the accuracy regimes for all subjects and all experimental conditions are presented in Fig. 16.13. Large *X*s depict the trials on which the subject's response was incorrect. The correlation between reaction time and P300 was lower (.476) during the speed regime than during the accuracy regime (.660). Note that whenever the subject makes an error, P300 latency exceeds reaction time. The conclusion from these data was that at least two processes are initiated by a stimulus. One is a response selection and execution process, which is indexed by the overt response. The other, a stimulus evaluation process, is indexed by the P300 component. Under accuracy instructions, response selection is contingent upon stimulus evaluation, and so the two processes are tightly coupled, and reaction time is frequently longer than P300 latency. When subjects operated under speed instructions, stimulus evaluation was more loosely coupled with response selection. Responses may be generated before the stimulus has been fully evaluated. This can be summarized by noting that when the subject is trying to be accurate, he or she thinks before acting, so the P300 occurs earlier in time because the subject does not press the button until he or she finishes processing the stimulus. However, when the subject is trying to be fast, he or she acts before thinking, and a longer-latency P300 is obtained. The thrust of the data is clear. The correlation, or lack thereof, between reaction time and P300 latency cannot be interpreted without reference to the specific tasks the subjects are performing and the strategies they adopt. The theoretical difficulties concerning P300-latency-reaction-time correlations that Tueting (1979) summarizes disappear if proper consideration is given to the different processes that underlie reaction time. At present, it would be reasonable to view changes in P300 latency as reflecting changes in stimulus evaluation time.

On the basis of the relationships between P300 latency and reaction time, a procedure was developed for error correction using P300 latency and reaction time. McCarthy, Kutas, and Donchin (1979; see also Donchin & McCarthy, 1980) described a study in which an individual received and responded to messages while sitting before a terminal. Of the messages, 90% were relatively unimportant, but 10% were important and should not have been missed. The cost of misses of the important event was great, whereas the cost of missing frequent events was not so large. Speed of response was reasonably important. An interesting aspect of this situation is that "misses," for the most part, will be due to

ACCURATE-RT


```
CORRELATION= .6604
SLOPE      = .57
Y-INTERCEPT= 155.93
F RATIO    = 185.65
PROBABILITY= .0000
```
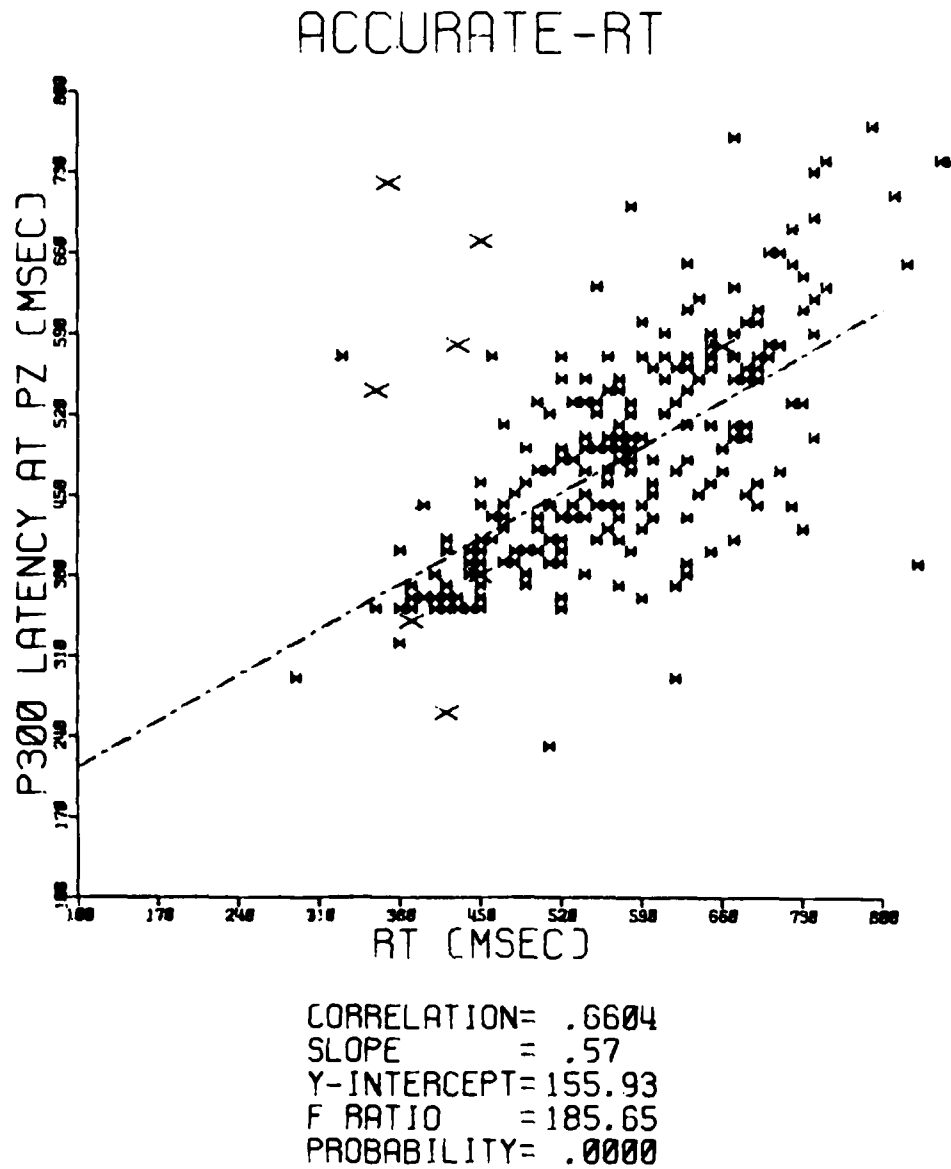
FIG. 16.13.   Scatter diagrams showing the relation between choice RT and P300 latency in speed- and accuracy-emphasis conditions in the Kutas et al. (1977) experiment. Trials in which errors were made are indicated by large X's and are not included in the determination of the regression line. Note the deceased correlation between RT and P300 latency in the speed-RT condition relative to the accuracy condition. (From Kutas et al., 1977.)

response bias. That is, the subject is more likely to press the "frequent" button, even when a rare event occurs, than to press the "rare" button when a frequent event occurs.

SPEED-RT
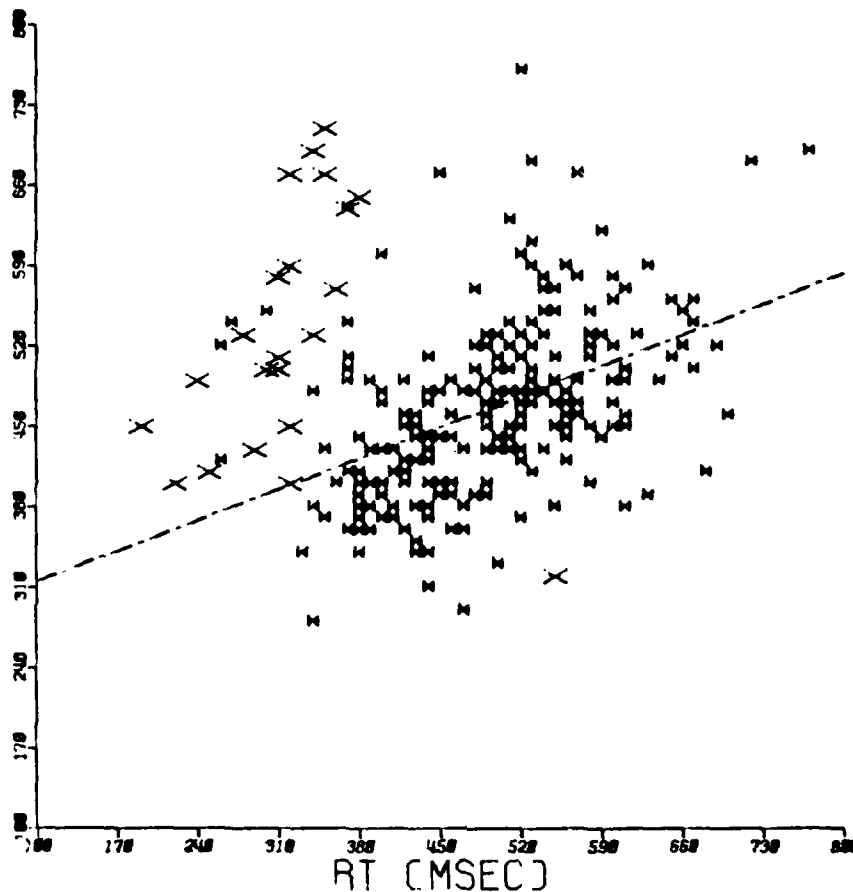

```
CORRELATION= .4756
SLOPE       = .38
Y-INTERCEPT=276.19
F RATIO     =69.86
PROBABILITY= .0000
```

FIG. 16.13. *continued*

The data from the experiment again illustrate the trade-off between speed and accuracy. If accuracy is stressed, the subject can classify 92% of the rare events correctly; but if speed is stressed, then he or she misclassifies 40% of the rare events. McCarthy et al. (1979) tried to determine if it is possible to add an ERP channel into the decision algorithm so that the 40% incorrectly classified rare items will be correctly identified. The logic works as follows: If the subject responds "rare," he or she is assumed to be correct, and no further analyses are needed. But when the subject responds "frequent," the ERPs are inspected to

determine if that individual is not in error. The latency of P300 for incorrectly classified rare events is considerably longer than P300 latency on the correctly classified rare events. And when subjects are correct, their reaction times are longer than when 'hey are incorrect. So the difference in time between the occurrence of P300 and the occurrence of the reaction-time press will discriminate very well between correct and error trials. Therefore, the following type of decision rule can be adopted. If the subject says rare, it is accepted as rare. If the subject says frequent, a three-step analysis is conducted. If there was no P300 at all, the trial is accepted as frequent. If there was a P300, the relation between P300 latency and reaction time is assessed before the final classification. The final result can be seen in Fig. 16.14. With this biocybernetic-aided classification, the error rate in the speed-stressed situation is reduced to the same level as that of the accuracy-stressed situation. Of course, this does not occur without an increase in the percentage of frequent items that are misclassified as rares. However, these misclassifications are not troublesome, because their cost, according to our presupposition, is low.
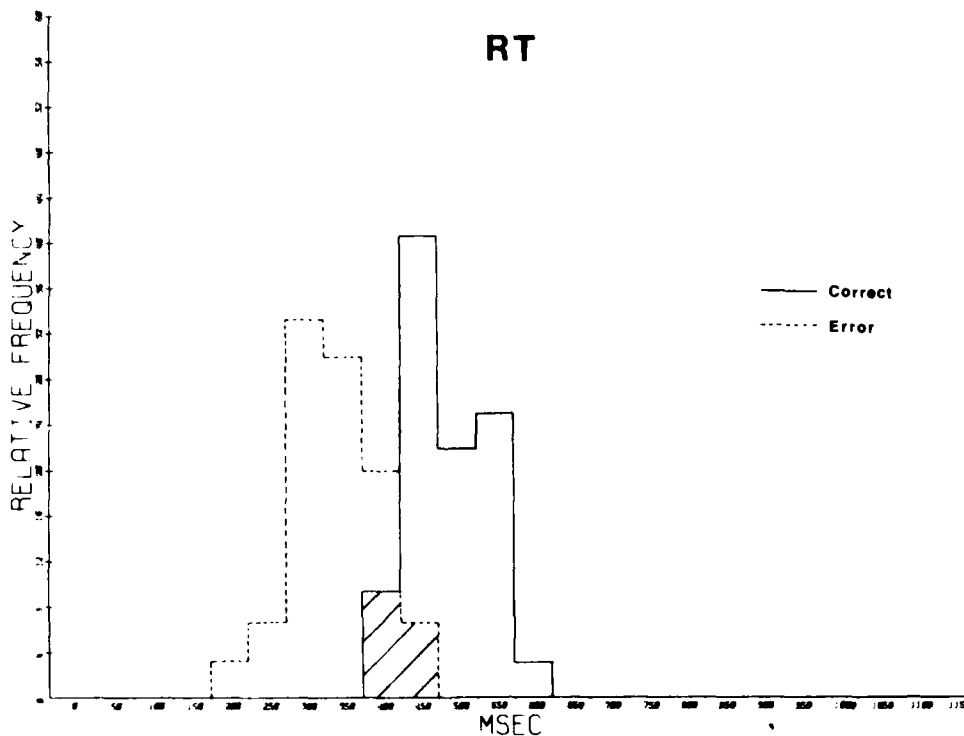
FIG. 16.14.   Histograms of RT, P300, and the difference between P300 and RT for each single trial in the speed-RT condition for one subject in the McCarthy et al. (1979) experiment. The dotted lines represent error trials, and the solid lines represent correct trials. Overlap in the distributions is indicated by the hatched lines. Note that RTs are shorter for error trials than for correct trials, whereas P300 latency shows the converse trend. (From Donchin & McCarthy, 1980.)
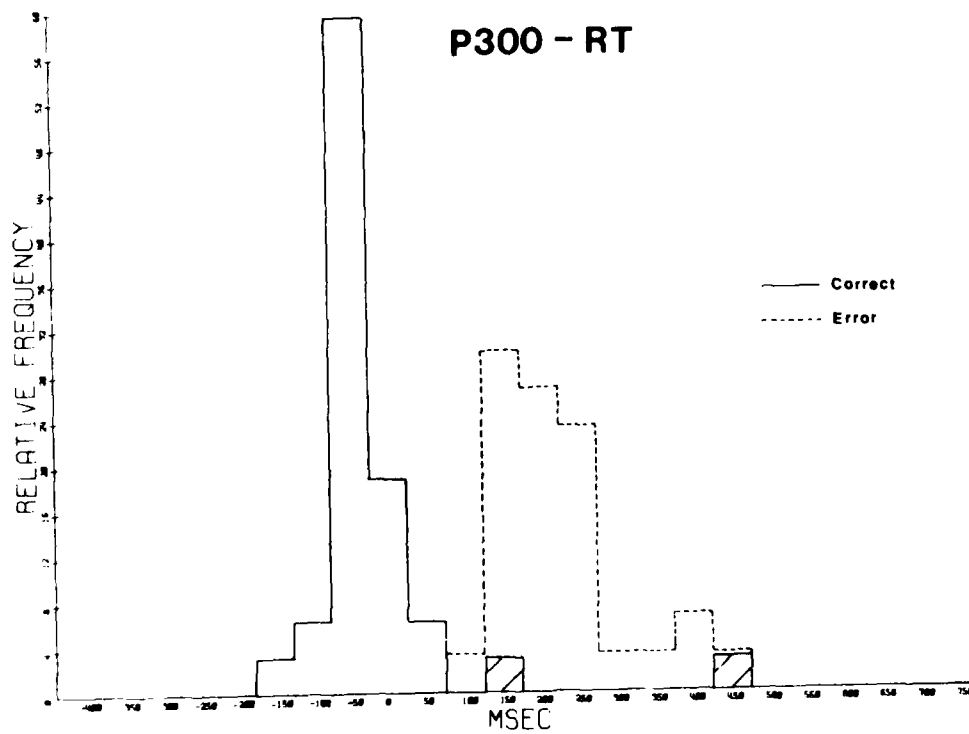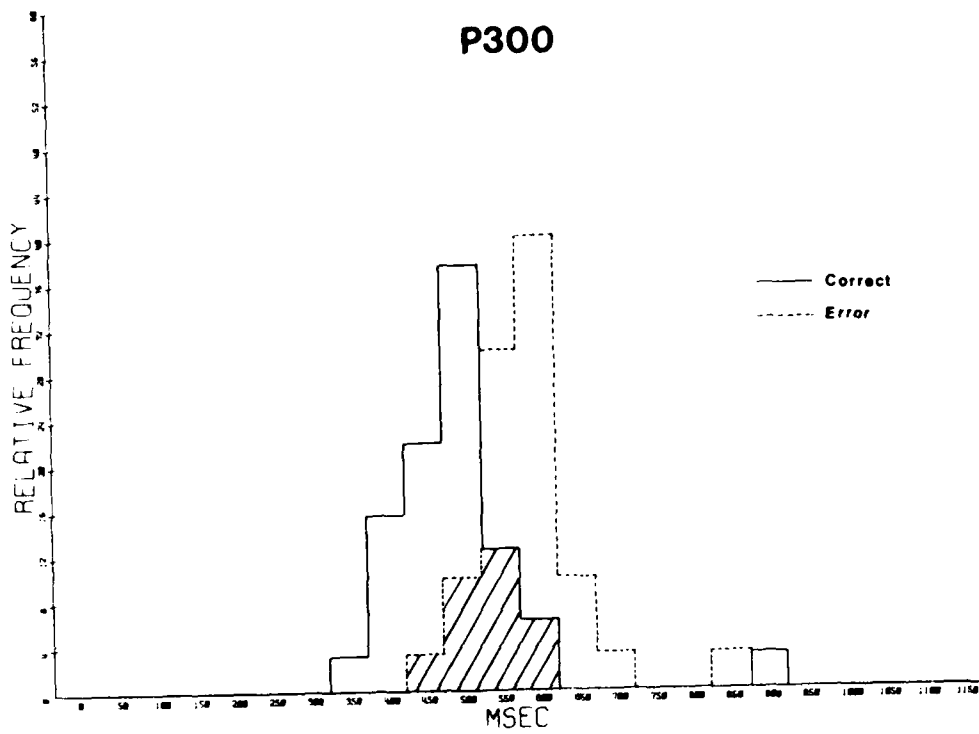
FIG. 16.14. *continued*

CONCLUDING REMARKS

The case for the utility of P300 as a tool in the study of human information processing is strong. The reader should note that several other components of the ERP, though not yet examined with the detail accorded P300, are likely to be of equal interest. Our intent here was not to review the field exhaustively (a task admirably undertaken by Callaway et al., 1978), but rather to convey a sense of this very active field of research and its possibilities. It is important to point out that in our concentration on work from the Cognitive Psychophysiology Lab, we have ignored an extensive body of research conducted in several other active laboratories. Fortunately, the results of most investigators converge to a consistent set of conclusions.

A few words may be in order on the primary topic of this conference, the differences between individuals. We have ignored this aspect in the previous pages. The research program at the Cognitive Psychophysiology Laboratory (Donchin, 1979) has tended to emphasize those aspects of the ERP that apply to the fictional average subject. Furthermore, the economics of our work argues against using large groups of subjects. In point of fact, the phenomena we report do hold with exquisite reliability for virtually all subjects we have studied. Yet it is equally clear that there are considerable individual differences in the pattern of the specific relations we describe. All subjects may show the sequential effects. For some, however, these will be more pronounced than for others. Similarly, the latencies, and the correlations between latencies and reaction times, often show aspects characteristic of the subject. It is trite to conclude the chapter by calling for "more research." It is nevertheless the case that few, if any, attempts have been made to utilize the endogenous components of the ERP in the study of individual differences. The task will require new approaches, new experimental paradigms, and novel analytical attitudes. The return on this investment promises to be high. Therefore, a call for research may well be an appropriate conclusion to this chapter.

ACKNOWLEDGMENT

REFERENCES

Adams, J. C., & Benson, D. A. Task-contingent enhancement of the auditory evoked response. *Electroencephalography & Clinical Neurophysiology*, 1973, *35*, 249-257.

Beach, L. R., Rose, R. M., Sayeki, Y., Wise, J. A., & Carter, W. B. Probability learning:

Response proportions and verbal estimates. *Journal of Experimental Psychology*, 1970, *86*, 165–170.

Bostock, H., & Jarvis, M. J. Changes in the form of the cerebral evoked response related to the speed of simple reaction time. *Electroencephalography & Clinical Neurophysiology*, 1970, *29*, 137–145.

Callaway, E., Tueting, P., & Koslow, S. (Eds.). *Brain event-related potentials in man*. New York: Academic Press, 1978.

Chapman, R. M. Evoked potentials of the brain related to thinking. In F. J. McGuigan & R. A. Schoonover (Eds.), *The psychophysiology of thinking*. New York: Academic Press, 1973.

Chapman, R. M., & Bragdon, H. R. Evoked responses to numerical and non-numerical visual stimuli while problem solving. *Nature*, 1964, *203*, 1155–1157.

Chesney, G. L. *The effects of prior expectancy and prediction outcome on the P300 component of the human average evoked potential*. Unpublished master's thesis, New Mexico State University, 1976.

Chesney, G. L., & Donchin, E. Predictions, their confirmation and the P300 component. *Psychophysiology*, 1979, *16*, 174. (Abstract)

Clynes, M., & Kohn, M. The use of the Mnemotron for biological data storage, reproduction, and for an average transient computer. Philadelphia: *Abstracts of Fourth Annual Meeting of the Biophysics Society*, 1960, *23*.

Courchesne, E., Hillyard, S. A., & Galambos, R. Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography & Clinical Neurophysiology*, 1975, *39*, 131–143.

Davis, H. Brain stem and other responses in electric response audiometry. *Annals of Otology, Rhinology and Laryngology*, 1976, *85*, 1–12.

Dawson, G. D. A summation technique for the detection of small evoked potentials. *Electroencephalography & Clinical Neurophysiology*, 1954, *6*, 65–84.

Debecker, J., & Desmedt, J. E. Rate of intermodality switching disclosed by sensory evoked potentials averaged during signal detection tasks. *Journal of Physiology*, 1966, *185*, 52–53.

Dixon, W. J. *BMDP-biomedical computer programs*. Los Angeles: University of California Press, 1975.

Donchin, E. A multivariate approach to the analysis of average evoked potentials. *IEEE Transactions on Bio-Medical Engineering*, 1966, BME-13, 131–139.

Donchin, E. Data anlaysis techniques in average evoked potential research. In E. Donchin & D. B. Lindsley (Eds.), *Average evoked potentials: Methods, results and evaluations*. Washington, D.C.: U.S. Government Printing Office, 1969.

Donchin, E. Brain electrical correlates of pattern recognition. In G. F. Inbar (Ed.), *Signal analysis and pattern recognition in biomedical engineering*. New York: Wiley, 1975.

Donchin, E. Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked potentials in psychiatry*. New York: Plenum Press, 1979.

Donchin, E., Callaway, E., Cooper, R., Desmedt, J. E., Goff, W. R., Hillyard, S. A., & Sutton, S. Publication criteria for studies of evoked potentials (EP) in man. In J. E. Desmedt (Ed.), *Attention, voluntary contraction and event-related cerebral potentials. Progress in clinical neurophysiology* (Vol. 1). Basel: Karger, 1977.

Donchin, E., & Heffley, E. Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research*. EPA-600/9-77-043. Washington D.C.: U.S. Government Printing Office, 1979.

Donchin, E., & Herning, R. I. A simulation study of the efficacy of stepwise discriminant analysis in the detection and comparison of event-related potentials. *Electroencephalography & Clinical Neurophysiology*, 1975, *38*, 51–68.

Donchin, E., & Lindsley, D. B. Average evoked potentials and reaction times to visual stimuli. *Electroencephalography & Clinical Neurophysiology*, 1966, *20*, 217–223.

Donchin, E., & McCarthy, G. Event-related brain potentials in the study of cognitive processes. In C. Ludlow & M. E. Doran-Quine (Eds.), *Proceedings of Symposium on Neurological Bases of Language Disorders in Children: Methods and direction for research*. Washington, D.C.: U.S. Government Printing Office, 1980, NINCDS Monograph No. 22, 109-128.

Donchin, E., Ritter, W., & McCallum, C. Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain event-related potentials in man*. New York: Academic Press, 1978.

Duncan-Johnson, C. C., & Donchin, E. On quantifying surprise: The variation in event related potentials with subjective probability. *Psychophysiology*, 1977, *14*, 456-467.

Duncan-Johnson, C. C., & Donchin, E. Series-based vs. trial-based determinants of expectancy and P300 amplitude. *Psychophysiology*, 1978, *15*, 262. (Abstract)

Fitts, P. M., Peterson, J. R., & Wolpe, G. Cognitive aspects of information processing. Adjustments to stimulus redundancy. *Journal of Experimental Psychology*, 1963, *65*, 423-432.

Ford, J. M., Roth, W. T., Dirks, S. J., & Kopell, B. S. Evoked potential correlates of signal recognition between and within modalities. *Science*, 1973, *181*, 465-466

Ford, J. M., Roth, W. T., & Kopell, B. S. Auditory evoked potentials to unpredictable shifts in pitch. *Psychophysiology*, 1976, *13*, 32-39.

Gomer, F. E., Spicuzza, R. J., & O'Donnell, R. D. Evoked potential correlates of visual item recognition during memory-scanning tasks. *Physiological Psychology*, 1976, *4*, 61-65.

Hartley, L. R. The effects of stimulus relevance on the cortical evoked potentials. *Quarterly Journal of Experimental Psychology*, 1970, *22*, 531-546.

Heffley, E., Wickens, C., & Donchin, E. Intramodality selective attention and P300 — Reexamination in a visual monitoring task. *Psychophysiology*, 1978, *15*, 269-270. (Abstract)

Horst, R., Johnson, R., Jr., & Donchin, E. Event related potentials and subjective probability in a learning talk. *Memory and Cognition*, in press.

Hyman, R. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 1953, *45*, 188-196.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, 1980, *17*, 259-273

Isreal, J. B., Wickens, C. D., Chesney, G. L., & Donchin, E. The event-related brain potential as an index of display-monitoring workload. *Human Factors*, 1980, *22*, 212-224.

Jewett, D., Romano, H. N., & Williston, J. S. Human auditory evoked responses: Possible brain stem components detected on the scalp. *Science*, 1970, *167*, 1517-1518.

Johnson, R., Jr., & Donchin, E. On how P300 amplitude varies with the utility of the eliciting stimuli. *Electroencephalography & Clinical Neurophysiology*, 1978, *44*, 424-437.

Johnson, R., Jr., & Donchin, E. The P300 and stimulus categorization: Two plus one is not so different from one plus one. *Psychophysiology*, 1980, *17*, 167-178.

Kahneman, D. *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review*, 1973, *80*, 237-251.

Karlin, L., & Martz, M. J., Jr. Response probability and sensory-evoked potentials. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

Karlin, L., Martz, M. J., Jr., Brauth, S. E., & Mordkoff, A. M. Auditory evoked potentials, motor potentials and reaction time. *Electroencephalography & Clinical Neurophysiology*, 1971, *31*, 129-136

Kerr, B. Processing demands during mental operations. *Memory & Cognition*, 1973, *1*, 401-412

Kutas, M., McCarthy, G., & Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, 1977, *197*, 792-795

Levit, R. A., Sutton, S., & Zubin, J. Evoked potential correlates of information processing in psychiatric patients. *Psychological Medicine*, 1973, *3*, 487-494

McCarthy, G., Kutas, M., & Donchin, E. Detecting errors with P300 latency. *Psychophysiology*, 1979, *16*, 175. (Abstract)

Messick, D. M., & Rapoport, A. A comparison of two payoff functions on multiple-choice decision behavior. *Journal of Experimental Psychology*, 1965, *69*, 75-83.

Myers, J. L. Probability learning and sequence learning. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 3). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Naatanen, R. Selective attention and evoked potentials. *Annales Academiae Scientiarum Fennicae*, 1967, *151*, 1-226.

Navon, D., & Gopher, D. On the economy of the human processing system. *Psychological Review*, 1979, *86*, 214-255.

Neimark, E. D., & Shuford, E. H. Comparison of predictions and estimates in a probability learning situation. *Journal of Experimental Psychology*, 1959, *57*, 294-298.

Norman, D., & Bobrow, D. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, *7*, 44-64.

Pachella, R. G. The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Picton, T. W., Campbell, K. B., Baribeau-Braun, J., & Proulx, G. B. The neurophysiology of human attention: A tutorial review. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Picton, T. W., Hillyard, S. A., Krausz, H. I., & Galambos, R. Human auditory evoked potentials I: Evaluation of components. *Electroencephalography & Clinical Neurophysiology*, 1974, *36*, 179-190.

Posner, M. I., & Boies, S. J. Components of attention. *Psychological Review*, 1971, *78*, 391-408.

Pribram, K. H., & McGuinness, D. Arousal, activation, and effort in the control of attention. *Psychological Review*, 1975, *82*, 116-149.

Reber, A. S., & Millward, R. B. Event observation in probability learning. *Journal of Experimental Psychology*, 1968, *77*, 317-327.

Regan, D. *Evoked potentials in psychology, sensory physiology and clinical medicine*. London: Chapman & Hall, 1972.

Remington, R. J. Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology*, 1969, *82*, 250-257.

Ritter, W., & Vaughan, H. G., Jr. Averaged evoked responses in vigilance and discrimination: A reassessment. *Science*, 1969, *164*, 326-328.

Ritter, W., Vaughan, H. G., Jr., & Costa, L. D. Orienting and habituation to auditory stimuli: A study of short-term changes in average evoked responses. *Electroencephalography & Clinical Neurophysiology*, 1968, *25*, 550-556.

Rohrbaugh, J. W., Donchin, E., & Eriksen, C. W. Decision making and the P300 component of the cortical evoked response. *Perception & Psychophysics*, 1974, *15*, 368-374.

Rolfe, J. M. The secondary task as a measure of mental load. In W. T. Singleton, J. G. Fox, & D. Whitfield (Eds.), *Measurement of man at work*. London: Taylor and Francis, 1971.

Roth, W. T. Auditory evoked responses to unpredictable stimuli. *Psychophysiology*, 1973, *10*, 125-138.

Roth, W. T., Ford, J. L., Lewis, S. J., & Kopell, B. S. Effects of stimulus probability and task-relevance on event-related potentials. *Psychophysiology*, 1976, *13*, 311-317.

Roth, W. T., Kopell, B. S., Tinklenberg, J. R., Darley, C. F., Sikora, R., & Vesecky, T. B. The contingent negative variation during a memory retrieval task. *Electroencephalography & Clinical Neurophysiology*, 1975, *38*, 171-174.

Ruchkin, D. S., & Sutton, S. Visual evoked and emitted potentials and stimulus significance. *Psychonomic Society Bulletin*, 1973, *2*, 144-146.

AD-A099 209   NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN D--ETC F/G 5/9
               CONFERENCE PROCEEDINGS; APTITUDE, LEARNING, AND INSTRUCTION. VO--ETC(U)
               JAN 81   R E SNOW, P FEDERICO, W E MONTAGUE
UNCLASSIFIED   NPRDC/ONR-TR-81-5-VOL-2   NL

2 OF 4
AD
A099209
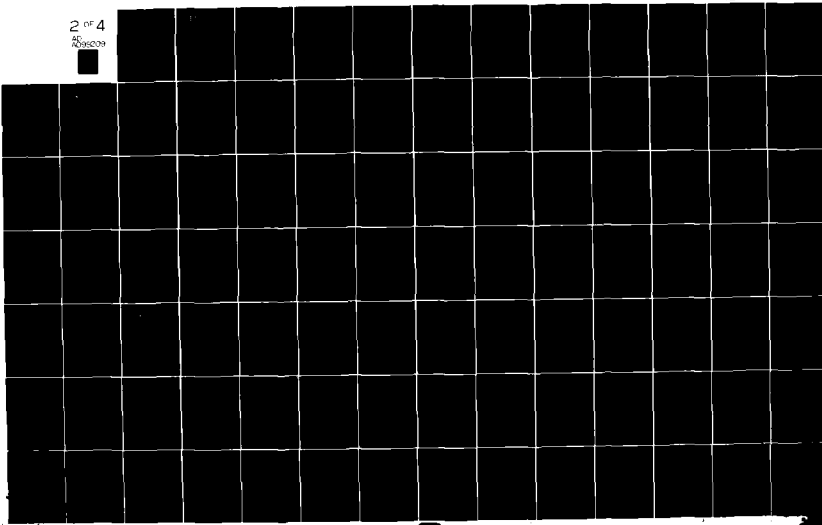
Satterfield, J. H. Evoked cortical response enhancement and attention in man: A study of responses to auditory and shock stimuli. *Electroencephalography & Clinical Neurophysiology*, 1965, *19*, 470–475.

Satterfield, J. H., & Cheatum, D. Evoked cortical potential correlates of attention in human subjects. *Electroencephalography & Clinical Neurophysiology*, 1964, *17*, 456.

Sokolov, E. N. The modeling properties of the nervous system. In I. Maltzman & K. Cole (Eds.), *Handbook of contemporary Soviet psychology*. New York: Basic Books, 1969.

Squires, K. C., & Donchin, E. Beyond averaging: The use of discriminant functions to recognize event related potentials elicited by single auditory stimuli. *Electroencephalography & Clinical Neurophysiology*, 1976, *41*, 449–459.

Squires, K. C., Donchin, E., Herning, R. I., & McCarthy, G. On the influence of task relevance and stimulus probability on event-related potential components. *Electroencephalography & Clinical Neurophysiology*, 1977, *42*, 1–14.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science*, 1976, *193*, 1142–1146.

Squires, N. K., Donchin, E., Squires, K. C., & Grossberg, S. Bisensory stimulation: Inferring decision-related processes from the P300 component. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 299–315.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 1969, *30*, 276–315.

Sutton, S. The specification of psychological variables in an average evoked potential experiment. In E. Donchin & D. B. Lindsley (Eds.), *Average evoked potentials: Methods, results and evaluations*. Washington, D.C.: U.S. Government Printing Office, 1969.

Sutton, S., Braren, M., Zubin, J., & John, E. R. Evoked-potential correlates of stimulus uncertainty. *Science*, 1965, *150*, 1187–1188.

Sutton, S., Tueting, P., Zubin, J., & John, E. R. Information delivery and the sensory evoked potential. *Science*, 1967, *155*, 1436–1439.

Tueting, P. Event-related potentials, cognitive events, and information processing. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office, 1979.

Tueting, P., Sutton, S., & Zubin, J. Quantitative evoked potential correlates of the probability of events. *Psychophysiology*, 1971, *7*, 385–394.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, *185*, 1124–1131.

Vaughan, H. G., Jr., & Ritter, W. The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography & Clinical Neurophysiology*, 1970, *28*, 360–367.

Vlek, C. A. J. Multiple probability learning. In A. F. Sanders (Ed.), *Attention and performance III*. Amsterdam: North Holland, 1970.

Wickens, C., Isreal, J., & Donchin, E. The event-related cortical potential as an index of task workload. In A. S. Neal & R. F. Palasek (Eds.), *Proceedings of the Human Factors Society 21st Annual Meeting*. Santa Monica, Calif.: Human Factors Society, October 1977.

Wilkinson, R. T., & Morlock, H. C., Jr. Auditory evoked response and reaction time. *Electroencephalography & Clinical Neurophysiology*, 1967, *23*, 50–56.

Woody, C. D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical & Biological Engineering*, 1967, *5*, 539–553.

17 Discussion: Process Analyses of Learning and Problem Solving

Henry M. Halff
Office of Naval Research

The most sensible approach to a discussion of this type might be to determine what common core or theme holds the chapters together. I devoted considerable mental effort to the search for such common elements but must admit to failure. As a second approach, I tried to place the chapters in the two-dimensional space used by Snow (Chapter 11, Vol. 1) for the same purpose. I found that I was forced to increase the dimensionality of the space to five in order to accommodate the new points. So, as a last resort, I decided to treat each of the chapters separately, giving the reader three separate discussions.

I take up the chapters in reverse order of their presentation, discussing Donchin and Isreal first, Rohwer second, and Greeno last. This reversal is purely for rhetorical reasons, since it seems easier to me to consider the chapters in increasing order of the complexity of the psychological processes discussed in each.

DONCHIN AND ISREAL'S ELECTROPSYCHOPHYSIOLOGY

There are two aspects of the work presented in Donchin and Isreal's chapter. First, we need to consider the status of electrophysiological work as a whole. Second, we need to consider the implications of the particular work done at Illinois and presented in this chapter. The two topics are not identical, for as I mention later, Donchin and Isreal's work is in many respects well beyond that of the rest of the field.

In speaking of the field in general, a first question concerns its relationship to physiological psychology. It may be mainly a quirk of history that this type of

work is practiced by those calling themselves physiological psychologists. Part of the quirk might be the failure of the early Gestalt psychologists (e.g., Kohler, 1947) to develop the field fruitfully, or it may be that neurophysiologists have traditionally felt more comfortable with electricity than have psychologists. But certainly the type of data that defines the field today is no more related to neurophysiology than are eye-movement data or other observations of nonintentional behavior.

The substance of the work coming out of electrophysiological laboratories seems to be more related to perceptual and cognitive psychology, and the methods used to investigate these topics are drawn from such areas as signal processing, which have their heritage in physics and engineering. The main implications of these points are for scholarship and education in the fields of electrophysiology and cognitive psychology, but related issues may also be of particular concern for the subjects considered at this conference—namely, instruction and the role of individual differences.

As Donchin and Isreal have pointed out, in instructional settings, event-related potentials give us an added channel into the mind of the learner. Thus, it might be possible to attach electrodes to the heads of our students and monitor their progress in learning. Such an application would require extensive theoretical and engineering development, and even if it were possible, the costs associated with the procedure, the time required for calibration to the individual student, and the uncertainty about what to do with information from the extra channel might well make such an application useless or infeasible.

A more promising enterprise might lie in the use of evoked potentials for exploring individual differences and, in particular, for psychological assessment. Some work (Lewis, Rimland, & Callaway, 1976) has already been done in this area, and many of the techniques described by Donchin and Isreal could be tried out as instruments of psychological assessment. In fact, there is much to be said for using evoked potentials as mental tests. They do not, for example, rely inherently on linguistic competence and might therefore not be subject to the cultural biases that may be present in such instruments as vocabulary tests. Evoked potentials may also not be subject to response biases, strategic manipulations, problems with guessing, and other weaknesses of conventional assessment techniques.

There is, however, one real and important danger in the use of evoked potentials for mental assessment. Psychometricians would call the problem a lack of face validity; it arises because there are no performance criteria that can be made clear to those being assessed and because we have, at this point, no notion of the process variations that might underlie individual differences in evoked potentials. It is one thing to tell people that they are being evaluated on the proportion correct or the difficulty of the tasks they are able to do and quite another to tell them that some linear function of the potentials from their scalp is the basis for classification. The latter criterion might be defensible only if there is a process-

based rationale that relates that linear evaluation function to its use. But a crite-
rion derived from blind application of actuarial techniques such as discriminant
analysis is not, to my mind, ethically acceptable.

It should be made clear that I do not view Donchin and Isreal's work, taken as
a whole, as being subject to this kind of criticism. Indeed, they have made
considerable progress in supplying us with the kinds of process explanations that
might serve as a basis for using evoked potentials in psychological assessment. In
doing so, however, they have provided the community with the mathematical
and physical tools that could also allow for considerable abuse of evoked poten-
tials in such an application.

Let me turn now to a discussion of the particular findings and theoretical
developments of the research program described in Donchin and Isreal's chapter.
The physiological heritage of evoked-potential research is evident in this paper
when they propose that their work is concerned with the "receptive field of the
P300," that event-related potential that constitutes the focus of the paper. In fact,
they are no more studying the receptive field of the P300 than Frederiksen, Hunt,
and Rose (Chapters 5, 4, and 3, Vol. 1) are studying the receptive field of the
"same"-button press. As Donchin and Isreal later point out, the real subject of
the paper is not the P300 itself but the cognitive processes responsible for its
elicitation. It is therefore important that we discuss these types of phenomena,
not in terms of the response of either the scalp or the finger, but in terms of the
paradigms themselves, the information-processing requirements of the tasks, and
so on. In one sense, Donchin and Isreal have been acutely sensitive to this need,
and their careful analyses present an impressive case for their theory of the P300.
But in another sense, they seem to have ignored much of the literature relevant to
the paradigms they use in their work. Thus, before discussing the implications of
the work for instruction, I make a few comments on this neglected literature.

The most basic paradigm used in Donchin's lab is that of choice reaction time.
One prime use of the paradigm was to demonstrate how the P300 depends on the
subjects' expectations about task-related stimuli. In particular, he shows that the
P300 varies with the sequence of previous stimuli in the same way that one might
surmise the subject's expectations would vary. The model presented by Donchin
and Isreal is offered as an account of these expectations, but I wonder why they
feel that this model offered advantages over current theories of sequential effects
in choice reaction time. Such models date back at least to Falmagne's (1965)
classic effort, and at least one refinement of that model (Falmagne & Theios,
1969) is quite sophisticated. Models such as these not only give good accounts of
existing reaction-time data but also provide a more process-oriented account of
the task. A similar point could be made about the probability-matching paradigm
used in some of Donchin's experiments, but there is probably less of a substan-
tive nature to be gained from an examination of that literature.

The speed-accuracy trade-off in choice reaction time is also one of the main
topics in the Donchin and Isreal paper. Again, there are many existing theories of

this trade-off, and Donchin seems to have chosen a form of the fast-guess model (Ollman, 1966; Yellott, 1967, 1971). Two-stage models of the form proposed by Atkinson and Juola (1973) might also be applicable to Donchin and Isreal's case, and at least passing consideration should be given to random-walk models such as Link's relative judgment theory (Link, 1975; Link & Heath, 1975).

Finally, and most relevant to the purposes of this conference, is Donchin and Isreal's discussion of the work on list learning. Here they propose that the P300 could be used to examine a subject's reaction to each item. In particular, Donchin argues that large P300s occur whenever subjects are told they are wrong about an item they think they know or that they are correct about an item they think they do not know. This interpretation is, in general, consistent with Donchin's notion that the P300 is an indicator of surprise, but any thorough analysis of the situation would require a theory of the trial-by-trial, item-by-item events in the learning situation. Such theories are not common in the literature on list learning, but a few explicit models can be found (e.g., Greeno, James, DaPolito, & Polson, 1978; Halff, 1977) that might be useful in this context.

If we look at the paper as a whole, however, it seems that the work on learning plays a small part in the overall research program. Here again, I think that Donchin and Isreal may have done a disservice to their own good ideas, and this disservice may only be compounded by their choice of the term *surprise* as a label for the event giving rise to the P300. As Donchin himself recognizes, the term is not strictly accurate, especially in view of the dependence of the P300 on task-relevant events. A more accurate interpretation of the research presented here may be found in another of Donchin's works (Donchin, Ritter, & McCallum, 1978):

It is useful to distinguish between "tactical" information processing, which is concerned on any experimental trial with specific responses to the actually presented stimuli, and "strategic" information processing, which is concerned with laying down of the rules that ultimately determine the tactics chosen on each trial. While subjects are tactically responding to stimuli, they are also processing all available information so that strategies can be adjusted to cope with task demands in the future. Adjusting the templates against which incoming stimuli are compared, choosing between a response set or a stimulus set, allocating resources to one or another input channel—all of these are aspects of strategic information processing. The crucial point is that on any trial much information processing is concerned less with what the response should be on that trial than it is with the rules under which tactics will be determined on future trials. The processes that are manifested by P300 represent such strategic information processing. The P300 represents the activity of a neural system that is engaged in evaluating ongoing strategies in the light of existing feedback information. The amplitude of P300 would then depend on the extent to which any given event forces a change in the subject's strategies (possibly, without conscious meditation). The timing (latency) of P300 depends on the time it takes the subject to identify the strategically relevant

information. Many of the puzzling aspects of the data, in particular the reported lack of direct correlation between P300 and the responses on specific trials, are clarified once it is realized that the processing they represent is in preparation for future trials rather than with the responses on present trials [pp. 393-394].

This quote presents the view that the P300 is not a surprise indicator but rather is a learning indicator, an event that occurs whenever subjects must change their minds or alter their view of the world and their situation. The research presented in this chapter has been mainly concerned with the conditions under which this event will occur. But it is equally important to proceed in the other direction and ask about the consequences of the cognitive processes underlying the P300, for any full understanding of the foregoing quote and useful application of the ideas therein will depend on our understanding of these consequences.

ROHWER'S ELABORATION THEORY

My search for the common elements in these three papers led me at one point to entertain the theory that my complete unfamiliarity with the three subjects discussed was the only common element. Rohwer's paper, however, gives the lie to this idea, for I, in a past life, shared Rohwer's concern both for list-learning strategies (Halff, 1977) and for the role of interitem associations (Weigel, Schendel, & Halff, 1978). Although my thinking has been consistent with Rohwer's, he addressed the issues at a level considerably above my own theoretical efforts, and his more general approach deserves some comment at the outset. The event families that form the basic unit of Rohwer's theory are what others have called schemata (Norman, Gentner, & Stevens, 1976), frames (Minsky, 1975), and scripts (Schank & Abelson, 1977). If these other authors are correct, then these cognitive structures are responsible for guiding and giving meaning to virtually every aspect of mental life. It may therefore be a mistake to give the impression, as Rohwer does, that event families are specialized entities that come into play only in particular circumstances. To be more specific, consider a subject examining a particular pair or tetrad of items, and recall that an event, according to Rohwer, consists of "a beginning state and one or more entities involved in some action that changes that state." Rohwer argues that such events, or event families, only become available under certain instructions or propensity conditions. But one event must always be available—namely, the appearance of the entities in question on the face of the memory drum or other presentation device. It is on this event that uninstructed subjects most rely. The effect of instructions is, not to encourage people to *create* events, but rather to push existing elaborative activity in directions that will better support retrieval. Viewed in this way, it is difficult to distinguish between the concepts of elaborative propensity and event repertoire.

I also found a few interesting points in the data from Rohwer's main experiment. Rohwer's three main interests in this study were in the destructive effects of *repetition* instructions (i.e., instructions to repeat items continually without elaborating), the effects of *prompt* instructions that encouraged the subjects to elaborate, and the *persistence* of the prompt instructions' effects over the course of a day. The main results were:

1. Repetition instructions caused younger subjects to lose more items on the average than older subjects.
2. Prompt instructions, at least for tetrads, resulted in more items being added to the recall of younger subjects than to the recall of older subjects.
3. For tetrads, but not for pairs, the persistence of prompt instructions was greater for older children than for younger children in that the difference in average words recalled was higher for the former subjects.

The results appear to be what Rohwer expected (although I would hesitate to call such expectations "predictions" as Rohwer does). I do, however, have some concern for interpreting these interactions with age when the baseline performance is so different between ages. To be more concrete, suppose we examined:

1. repetition as the proportion r of recallable items rendered unrecallable by repetition instructions;
2. prompt effects as the proportion p of unrecallable items rendered recallable by prompt instructions; and
3. persistence as the proportion s of such gains still manifest on the day after prompt instructions were given.

Estimates of these proportions, derived in the natural way from Rohwer's Table 15.2, are presented in Table 17.1. As can be seen, virtually all of the interesting differences between ages are found in the tetrads data, and even here, the prompt effect seems to be the same across ages. That is, prompt instructions are more effective for younger children only because they have a larger pool of items upon which to work.

Rohwer also discusses three results that have to do with individual differences as measured by performance on Day 1. The interpretation of these results is more difficult than that of the main effects, and Rohwer himself admits to some confusion in this regard. The reasoning that led me to Table 17.1 would at least demand a look at scatterplots and some alternative regression models. The puzzling effects in both the main experiment and its follow-up (e.g., the differences between pair and tetrad learning and the lack of individual-difference effects) may be bound up in the structure of the particular learning task, the precise nature of prompt operations, the management of short-term memory, and other specifics not dealt with in the theory. I suspect that a precise process model of learning,

TABLE 17.1
Repetition (r), Prompt (p), and Persistence (s) Proportions
Taken from Rohwer's Table 15.2[a]

| | List Type | | | | | |
| | Pairs | | | Tetrads | | |
Grade	r	p	s	r	p	s
5	.736	.213	.917	.661	.328	.667
11	.587	.313	1.060	.358	.293	1.222

[a]The proportion r is the repetition score on Day 2 divided by the baseline score on Day 2; p is the difference between the prompt and baseline scores on Day 3 divided by the difference between the maximum score (36) and the baseline score on Day 3; and s is the difference between prompt and baseline scores on Day 4 divided by the corresponding difference on Day 3.

with parameters reflecting the more interesting effects, would be of considerable utility in explicating Rohwer's ideas. A simple interpretation of Table 17.1 and Rohwer's Figs. 15.4 and 15.5 suggests that the effects of using elaboration or repetition are essentially the same across ages and individuals, but younger children are less adept at the use of either method and tend to stop using at least elaboration before older children do.

I have little to note about the implications of this work for instruction, because most of the relevant points have been made by Rohwer. The most important lesson to be learned from the data themselves is that younger children suffer, not from their *ability* to use elaborative techniques effectively, but from their *propensity* to use such techniques. The lack of a proper conceptual and experiential base (i.e., an adequate *event repertoire* in Rohwer's terms) seems to be less of a concern in view of these data.

More of a problem is the task of determining when to use elaborative techniques in a curriculum, and my discussion at the first of this section seems relevant here. My main point there was that elaboration is a particular form of comprehension. Thus, in considering the use of elaborative techniques, we must always weigh their mnemonic value against the value of the understanding itself, the latter often giving the student a power over the material that goes far beyond the ability to remember. The issue, of course, is not new. I recall my father developing an understanding of techniques of integration (for calculus) that allowed him to work any problem based on a few simple forms. Unfortunately, his understanding was not powerful enough to provide the necessary speed to do the scores of problems his teacher requested that he memorize for the exam. My father may have lost that battle, but judging from contemporary textbooks in calculus, he won the war. That such a war may be worth fighting across the full range of learning can even be seen in the examples of brute memory cited in

Rohwer's first paragraph. Vocabularies (foreign and English) have an ortho-graphic structure that enables one to know far more words than he or she has learned. All things considered, better roast chickens will come from those who assemble the ingredients based on an understanding of cooking than from those who memorize the list of ingredients. Like most of us, I have seldom even contemplated the replication of a published experiment. But when I have consid-ered such an act, I usually found that missing from the published report were those crucial details of procedure that I would have memorized had they been present. To replicate the experiment, I would have to have supplied those details on the basis of my understanding of the procedures.

GREENO'S COGNITIVE TASK ANALYSIS

I save Greeno's work for last because it is closer to real-world instructional situations than those described in the other two chapters. To get so close to application, Greeno has made use of a relatively new methodology in cognitive psychology, one that still has some grave shortcomings, in my opinion. I point out some of these shortcomings in my discussion of Greeno's work, and it is important that the reader not take my criticisms as applying to Greeno alone, but rather to the state of the art in general. Points about methodology are always more meaningful when embedded in substance, so let me turn directly to the substance of Greeno's work. I first consider Perdix, discussing the psychology of the theory and then the instructional implications of this psychology. I then develop a parallel discussion of the work on arithmetic word problems.

Perdix

Perdix, you will recall, is a theory of competent geometry problem solving represented as a computer program. Based on thinking-aloud protocols from a group of six high-school students, the theory is written as a production system that includes "procedures and structures of knowledge that enable the model to solve the problems that these students were able to solve, in the same general ways that the students solved the problems." The substance of the theory lies in productions for making inferences, for perceptual organization, for pattern rec-ognition, and—most interestingly—for planning and strategy in problem solv-ing.

The planning mechanism in Greeno's theory is its most important psychologi-cal contribution of a general nature. But if this large, essential, but unwritten body of knowledge underlies problem-solving competence, I think that some evidence for Greeno's particular characterization needs to be introduced. The *objective* relationship of protocol data to theory is neglected in many investiga-tions of this kind, and Greeno's is no exception. I am not so much worried that

the theory is an invalid explanation of the data as I am that the relationship between the two is not public, explicit, or perhaps even replicable. If these theories are to be of any use, they must be something that the scientific community as a whole can examine, test under various conditions, and map into a wider variety of situations than those leading to their original formulation. How could I decide that Greeno's theory did not hold or had to be changed in interesting ways under certain conditions?

These kinds of questions are closely related to a second topic of discussion— namely, the expansion of this theory to a performance model. Such a model would have parameters to account for the effects of variables that might degrade or alter performance, and one point in the parameter space would be the competence model presented in Greeno's paper. Interesting possibilities for these variables include time pressure and other conditions that degrade overall performance. Such variables might be particularly informative about the relative strengths of conditions on some of the productions. Also of interest might be priming effects, which could be useful in determining the implicit relationships between components of the theory.

A third class of variables, which are of more concern to Greeno and to the aims of this volume, are those related to learning. There appear to be two ways of approaching learning. One method involves determining the possible states of partial knowledge from a careful examination of less-than-competent students. Such a cross-sectional analysis might result in what Gagné (1962, 1970) calls a learning hierarchy In taking this approach, one is tempted to rely heavily on the competence theory and denote states of partial learning in terms of absent productions.

That such an approach might be fruitless or misleading can be seen by examining the other approach to the issue of learning—namely, the prediction of various stages of learning based on a process model of the learning process itself. Many such process theories might be consistent with the absent-production approach to incomplete learning. Examples that spring to mind are early pattern conditioning theories (Estes, 1959) and their more recent cousins (Cotton, Gallagher, Marshall, & Varnhagen, in press). But theories such as Anderson, Kline, and Beasley (Chapter 21, this volume) would admit no such consistency. Anderson et al. treat learning much as an evolutionary system, where incomplete states are characterized not so much by missing elements as by an abundance of inappropriate elements that have not been weeded out by experience. On such an account, learning consists, not of the development of correct productions, but of the strengthening of correct productions relative to incorrect ones.

This discussion leads naturally to the potential uses of Perdix for instruction. Greeno's main suggestion is to examine the possibility of explicitly teaching students to use Perdix's strategic principles. My small acquaintance with the literature in the area of instruction in problem solving gives me the impression that direct instruction in problem-solving strategies is generally unsuccessful.

Shoenfeld (1978) has pointed out that the reason for such failures may lie in our failure to equip students with a control structure for using heuristics and other problem-solving strategies. Perdix does have a control structure that could perhaps be taught to students, but such aspects of computer simulation programs are far from unique. A rule-based system is a convenient way of instantiating the theory in the computers of today but is not necessarily isomorphic with the structures found in the minds of Greeno's subjects. There are, no doubt, higher-level organizing principles that structure the knowledge in these subjects' heads over and above the structure dictated by the production system itself. In fact, I gather that Greeno's distinction between strategic, perceptual, and formal rules may be one of his own making, not formally represented in the simulation. It is just these higher-level distinctions that might be most useful in teaching the control of problem-solving processes.

But even if Perdix itself is not teachable, it could serve several useful instructional functions. It might, for example, tell us something about problem difficulty. A complete account of this issue is, of course, far beyond Perdix at this point for we have no performance model. But Perdix should at least be able to order some problems in terms of difficulty or tell us how to make a particular problem into one that is more or less difficult. Such an application might be begun by looking at the relative difficulties of two problems that exercise the same subset of rules or cases where the trace of one problem is found as part of the trace of another problem.

A more exciting application would be to use Perdix to tell us something about the particular skills required to solve problems. A teacher, for example, might want to know about the strategic knowledge or perceptual skills required to solve the problems that he or she might assign. The relative proportions of the three types of skills might be altered systematically throughout the course.

A related but more systematic use of Perdix would be to incorporate the theory into an automated problem assigner. We have heard of some of the techniques for automatically assigning students problems with an effective mix of old, mastered skills and new skills yet to be learned (Snow, Chapter 2, Volume 1; for a more complete description, see Wescourt, Beard, Gould, & Barr 1977). But to support such techniques, one needs an infrastructure that can define the set of skills and determine which members of the set apply to any one problem. Perdix supplies this infrastructure and hence might be useful in such an application.

Finally, let me make brief mention of one application that is probably infeasible at present. It might be possible at some point to use Perdix as a problem generator. Even now, Perdix could probably be made to generate problems conforming to specifications posed in terms of the *competence* model. But for effective teaching, one would also need to control the anticipated behavior of incompetents to the problems thus generated, and for this, a performance model is needed.

Arithmetic Word Problems

Turning now to the arithmetic work, I again must question the basis of Greeno's conclusions in data. Recall that Greeno's hypothesis is that any problem of the type under consideration is mapped into one of the structures i' istrated in Figs. 14.2, 14.3, and 14.4. Each of these schemata is a different way of representing the meaning of the problem. As with the work on Perdix, I fail to see what in his subjects' data led Greeno to divide the semantic domain into these particular three structures. Indeed, I am not quite sure of the distinctions between the three structures. Examination of his figures reveals that the schemata are at least approximately isomorphic, and Greeno notes that some problems may map into more than one schema. What, then, might be the objection to suggesting a single schema? In making this suggestion, I do not mean to imply that children do not know the difference between, say, changes and combinations, but that "the inferences made in the context of arithmetic word problems" (Greeno, Chapter 14, Vol. 2) lead subjects to think of these operations as equivalent.

Greeno's overall research strategy in dealing with the arithmetic problems is also somewhat puzzling to me. In contrast to his work on Perdix, Greeno has chosen to develop a competence theory of problem solving in this case on the data produced by incompetents, small children who can only solve slightly over 50% of the problems presented. A more obvious research strategy might indicate a detailed look at competent performance. Some reaction-time or eye-mov^ment studies of adults might lead fairly naturally to a process model of the task that could then be used to understand the errors made by children.

As with Perdix, I cannot resist making some suggestions about the implications of this work for instruction. I have mentioned my difficulties in distinguishing among the three schemata; yet I feel at least moderately competent to solve the problems discussed by Greeno. I therefore have my doubts about the value of "training in representing problem situations as one of the three general schemata." The assignment of arithmetic val'es to nodes might be somewhat more critical, and instruction on this aspect of the process might be beneficial if children appear to have difficulty in making such assignments. Certainly Greeno's suggestion that we could use these problems as pedagogical tools in teaching more abstract concepts seems to be a more fruitful path.

REFERENCES

Atkinson, R. C., & Juola, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Fourth international symposium on attention and performance*. London: Academic Press, 1973.

Cotton, J. W., Gallagher, J. P., Marshall, S. P., & Varnhagen, S. Theory of learning mathematics by computer-assisted instruction. *Behavior Research Methods & Instrumentation*, in press.

Donchin, E., Ritter, W., & McCallum, W. C. Cognitive psychophysiology: The endogenous components of the ERP. In E. Calloway, P. Tueting, & S. Koslow (Eds.), *Brain event-related potentials in man*. New York: Academic Press, 1978.

Estes, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford, Calif.: Stanford University Press, 1959.

Falmagne, J. C. Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology*, 1965, 2, 77-124.

Falmagne, J. C., & Theios, J. On attention and memory in reaction time experiments. *Acta Psychologica*, 1969, 30, 316-323.

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

Gagné, R. M. *The conditions of learning* (2nd ed.). New York: Holt, Rinehart & Winston, 1970.

Greeno, J. G., James, C. T., DaPolito, F., & Polson, P. G. *Associative learning: A cognitive analysis*. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Halff, H. M. The role of opportunities for recall in learning to retrieve. *American Journal of Psychology*, 1977, 90, 383-406.

Kohler, W. *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright, 1947.

Lewis, G. W., Rimland, B., & Callaway, E. *Psycho-biological predictors of success in a Navy remedial reading program* (TR 77-13). San Diego, Calif.: Navy Personnel R & D Center, 1976. (NTIS No. AD A037 339)

Link, S. W. The relative judgment theory of two choice reaction times. *Journal of Mathematical Psychology*, 1975, 12, 114-135.

Link, S. W., & Heath, R. A. A sequential theory of psychological discrimination. *Psychometrika*, 1975, 40, 77-105.

Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Norman, D. A., Gentner, D. R., & Stevens, A. L. Comments on learning schemata and memory representation. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Ollman, R. Fast guesses in choice reaction time. *Psychonomic Science*, 1966, 6, 155-156.

Schank, R., & Abelson, R. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Shoenfeld, A. H. *Problem solving strategies in college level mathematics*. Berkeley, Calif.: The Group in Science and Mathematics Education, University of California, 1978.

Weigel, G. A., Schendel, J. D., & Halff, H. M. *Cued recall for words presented in separate pairs*. Paper presented at the 18th annual meeting of the Psychonomic Society, November 1978.

Wescourt, K. T., Beard, M., Gould, L., & Barr, A. *Knowledge-based CAI: CINS for individualized curriculum sequencing* (Tech. Rep. No. 290). Stanford, Calif.: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1977.

Yellott, J. Correction for guessing in choice reaction time. *Psychonomic Science*, 1967, 8, 321-322.

Yellott, J. Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, 1971, 8, 159-199.

18

Planning Nets: A Representation for Formalizing Analogies and Semantic Models of Procedural Skills

Kurt VanLehn
John Seely Brown
Xerox Palo Alto Science Center

INTRODUCTION

At some time in our lives, we have all been forced to learn the procedural skills that supposedly comprise mathematical literacy (e.g., place-value addition) through the process of rote memorization, perhaps, enhanced by the use of "models" (e.g., the abacus). These models were intended to provide an intuitive basis for a given procedure. But what really is a "model" of a procedural skill? How does it help in learning? How faithful can it be made to be? And, more generally, how can it help a procedure take on "meaning"?

This chapter is directed at understanding how procedures can take on "meaning." It is intended to provide a small step in that direction by discussing a particular kind of "semantics" for procedural skills, which we call *teleologic* semantics, in the context of the unambiguous and totally specifiable procedural skills of elementary mathematics.

The teleologic semantics of a procedure is knowledge about the purposes of each of its parts and how they fit together. Such knowledge is the province of true masters of the procedure. Its value is extolled by the proverb, "To really understand something, one must build it." Teleologic semantics is the meaning possessed by one who knows not only the surface structure of a procedure but also the details of its design.

This chapter has two arguments. First, we motivate the particular representation that we use for teleologic semantics, which we call *planning nets*, by showing how it can capture analogies between procedures as seen by an expert at those procedures. Second, we show that teleologic semantics, as formalized by planning nets, is useful by describing several potential applications in the field of

education. In particular, some consideration is given to how teleologic semantics can be explained and to how it provides a useful framework for developing "optimal" sequences of "model" procedures (or microworlds) for guided discovery learning.

Analogy Between Procedures

Before we delve into a technical discussion of procedural analogies, let us consider a simple example of an analogy between the procedure for adding two multidigit numbers and a "model" procedure for addition that manipulates physical objects that represent numbers. The model procedure is a physical procedure in that it manipulates physical objects that stand for numbers. Before we can describe the procedure, we briefly describe the objects that it manipulates, namely, Dienes Blocks.

The Dienes Blocks Representation of Numbers.   Dienes Blocks provide an explicit representation of base-10 numbers—namely, a set of *unit* blocks for representing the units; a set of *long* blocks consisting of 10 unit blocks molded into a long stick for representing the 10s; a set of *flat* blocks consisting of 10 long blocks laid next to each other, thus forming a 10 × 10 square for representing the 100s; and finally a set of *cubes* in the form of 10 × 10 × 10 units for representing the 1000s. A number (of four or less digits) can be physically represented by selecting the number of unit blocks to correspond to the units digit, the number of long blocks to correspond to the 10s digit, and so on. Hence a particular multidigit number is represented by piles of units, longs, flats, and cubes. Here, for example, is 123 represented in Dienes Blocks:
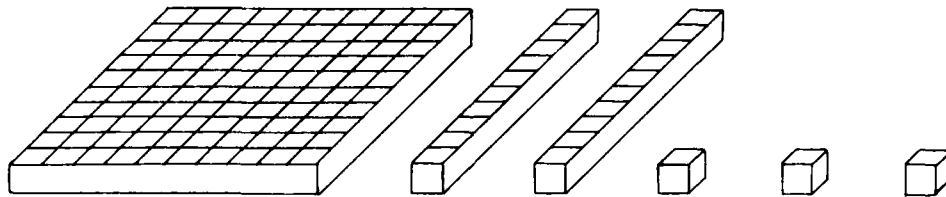
The base-10 nature of the symbolic place-value scheme for representing numbers is then made explicit, since one can see the direct translation of a number represented as piles of Dienes Blocks into a base-1 system (i.e., the total number of units comprising all the blocks in all the piles).

Dienes Block Addition.   Addition of two multidigit numbers represented as concrete Dienes Blocks involves forming set unions and "trading." The units pile for each of the two numbers is first unioned together. This corresponds to adding the units column. Next, the resulting set is examined. If it contains more than 10 unit blocks, then 10 blocks are removed from this set and traded for a long

block (consisting of 10 units), which is then placed in a pile of long blocks of the top number. This corresponds to carrying from the units to the 10s column in standard addition. The procedure now repeats, unioning the longs piles, then the flats, and so forth.

A theory of analogy between procedures, applied to this case, should be able to capture not only the fact that Dienes Block addition and standard addition produce the same answers given the same inputs, but that their *internal struc- tures* correspond as well. Set unions match with column sums, trading matches carrying, and so on.

Two-Pass Addition Illustrates Differences in Closeness.   We were recently struck by the way Dienes Blocks were being used in a school. In particular, the Dienes Blocks procedure being taught was not as described earlier but instead had the students combining all the piles of blocks together and then returning to the units pile and trading up and so on. Thus, in standard multidigit addition, a carry is (potentially) performed after each column operation, whereas in this version of Dienes Block addition, the "trading" (or carrying) operation was being deferred until all the columns has been initially processed. One intuitively feels that this second, two-pass procedure is not as closely analogous to standard addition as the previous, one-pass Dienes Block procedure.

A theory of analogy should have some formal measure that can predict how close an analogy is. The theory that follows has such a formal mechanism, called a *closeness metric*. The degree of correlation between the predictions of the closeness metric and subject's intuitive judgments of closeness is one verification condition for the theory.[1]

Why Arithmetic?   The examples in the chapter are all drawn from the compu- tational procedures of arithmetic, even though the techniques we have developed have wide applicability. We limited our examples to arithmetic for several rea- sons. Everyone knows how to add and subtract, so lack of familiarity with the example domain will not hinder comprehension of these admittedly rather abstract formalisms. Arithmetic is a highly evolved, complex system of proce- dures. It has iteration, recursion, tables of facts, and, of course, a rather nontri- vial data representation—namely place-value numbers. Lastly, arithmetic is taught in school. This means our theories are more likely to accrue the benefits of thoughtful, experience-based criticism from those with a sincere interest in put- ting the theories to work.

---

[1] It is safe to assume that individuals will differ in their judgments of the closeness of analogies. We take the position that this is due to the different deep structures that they assign to procedures. For example, someone who is just learning addition may not find the analogy between one-pass and two-pass addition particularly close. This might be due to a lack of distinct concepts for "carrying" and "column addition." So how one understands a procedure affects the data against which the theory of analogy will be verified. Because we are interested in teleologic semantics and because teleologic understanding is a mark of expertise, it is important to use experts as subjects.

Organizational Overview.   The chapter is divided into three parts. The first part is an exposition of some of the basic concepts of formal theories of analogy. We assume that an analogy can be represented as a *mapping* between a *deep structure* representation of each procedure that is expressed as a *maximal partial isomorphism* between the two deep structures. Thus, after an analogy has been comprehended, we would expect to find cognitive structures that could be modeled by three components; two of which represent the abstraction or deep structure of the two procedures, and the third representing the structure-preserving map (i.e., analogy) between these two structures.

The second part of the chapter motivates the planning-net representation of teleologic semantics by using it as the *deep structure* component of a theory of analogies between procedures. The third part is an examination of some of the applications of this theory to education. In particular, we discuss a paradigm for explaining the teleologic semantics that involves using a sequence of analogies such that each analogy illustrates exactly one concept underlying the synthesis of the given "target" procedure (e.g., place-value subtraction). This paradigm is then augmented with a set of "naturalness" principles for structuring a sequence of analogies, thereby addressing the problem of how to design an optimal sequence of "microworlds" or models for enhancing discovery learning.

We caution the reader that our style of arguing with examples has led to the incorporation of a great deal of detail into the subsequent pages. However, if artificial intelligence has contributed anything to cognitive psychology, it is an appreciation that ignoring trivial detail often leads to overlooking nontrivial problems.

A GENERAL THEORY OF ANALOGY

This section presents a theory of analogy so general that it is almost vacuous. It appears that virtually any theory of analogy, including the theory of procedural analogies that is presented later, can be recast as a special case of this general theory. Thus, this general theory is apparently immune to refutation. Nonetheless, it allows discussion of some concepts common to all analogies, such as "closeness," before becoming immersed in the details of procedures and their representations.

Mapping Between "Deep Structures"

We view an analogy as a comparison of two "things" that can be broken down into three parts: (1) an analysis of the first thing into some abstract description (or deep structure); (2) an analysis of the second thing into another abstract description; and (3) a mapping between the two descriptions. This tripartite breakdown is the foundation of the general theory of analogy. Exactly this breakdown is also

found in Tversky's work on similarity, a domain that illustrates the general theory more clearly because of the simpler "deep structures" that are used (Tversky, 1977).

Much research on similarity has used pairs of geometric figures or letters. A typical task is to rate the similarity of o to c. Tversky's analysis of this task is to assume a feature space, describe each figure as a set of features, then predict the similarity judgments with some "metric" on the overlap of the feature set of o with the feature set of c. The correlation of the judgments with the predictions serves as a verification condition on the feature space and the metric. Often, the features are not very abstract; o might be mapped into the description {curved, circular, closed}, and c would become {curved, circular, open}.

Much of the research on analogy has studied a task one often finds on intelligence tests—namely, to fill in X in a statement of the form: "A is to B as C is to X." Most commonly, the four elements A, B, C, and X are either words or geometric figures. A simple example of a word analogy problem is: "Red is to Stop as Green is to (a. Go; b. Halt)." Superficially, this appears to be a different sort of task than the similarity task, since there are four things rather than two. But the two tasks become very much the same when one considers the analogy task to be a comparison of *relationships* rather than directly apprehendable things. This is a widely held view of analogy. Indeed, the instructions to one analogy test, as quoted by Evans (1968), read: "Find the rule by which Figure A has been changed to make Figure B. Apply the rule to Figure C. Select the resulting figure from Figures 1 to 5 [p. 272]."

Actually, these instructions represent just one strategy for answering analogy problems. Evans' ANALOGY program, for example, used a different strategy, whereby it extracted an AB rule, then found five rules for pairs c_1, c_2, c_3, c_4, and c_5, then finally chose one rule of the five as being the most similar to the AB rule. The existence of many different strategies for solving analogy problems also obscures the parallels of this task to the similarity and metaphor tasks. Yet when one is done finding the analogy, one possesses the same three maps; an abstraction from AB, an abstraction from CX where X is the chosen answer, and the partial match (or mapping) between the two resulting abstract descriptions.

In short, if one ignores the strategic differences between solving an analogy and evaluating a similarity, and if one puts relationships on an equal footing with letters and geometric figures, then there is very little difference between the analogy task and the similarity task. After either task is completed, the cognitive structures can be modeled by three components: the two abstract descriptions and the mapping (in the form of a match) between them.

Basic Definitions

In this subsection, several basic concepts are discussed. They all follow rather immediately from the three-task view of analogy already described. As earlier,

they are motivated and illustrated with examples from Tversky's theory of similarity.

Intersection and Difference Sets.   A good way to summarize the outcome of the matching map is in terms of one *intersection set* and two *difference sets*. As an example, take the similarity task mentioned earlier to evaluate the similarity of o and c. Their descriptions, let's say, are the feature sets {round, curved, closed} and {round, curved, open} respectively. Call these sets A and B, the abstract descriptions of o and c. Then, the intersection and difference sets are:

$A \cap B$ = {round, curved}

$A - B$ = {closed}

$B - A$ = {open}

This is not particularly startling, to be sure. Indeed, there are stereotypical ways of referring to these sets in English similes: "A is like B in that $A \cap B$," or "A is like B except that $A - B$ instead of $B - A$."

Maximal   Partial   Graph   Morphisms   Generalize   the   Notion   of "Match."   With more complex languages than feature spaces for expressing abstract descriptions, one must of course give a new definition of "match." For example, consider the analogy (from Sternberg, 1977): "Washington is to 1 as Lincoln is to 5." Suppose semantic nets are the representation language. The abstract description of the relationship Washington:1 is a certain chain of semantic links from the node "Washington" to the node "1." The description of Lindoln:5 is a different chain. However, when one finally finds the correct way to view the two relationships (which is rather nontrivial for this example), then the two chains end up bearing the same sequence of link names—namely: Last-name, image-of, portrait-on, dollar-amount. That is, "Washington" is the *last name* of the man $NODE_{31}$; the *image of* $NODE_{31}$ appears in the picture $NODE_7$; $NODE_7$ is the *portrait on* the kind of dollar bill $NODE_{68}$; and the *dollar amount* of $NODE_{68}$ is "1." The chain for the Lincoln:5 relationship is a completely distinct chain, but it has exactly the same sequence of link labels. In this sense, the analogy is perfect.

To make these two chains match, the definition would have to be sensitive to: (1) the order of the links; and (2) the labels on the links. A definition in terms of intersection of sets of links would be inappropriate because none of the links are identical and because such a definition would ignore the topology of the descriptions. A definition of "match" that is appropriate for semantic nets (or any other representation with the topology of a labeled directed graph, including planning nets) can be defined in terms of a graph isomorphism:

Adjacency: Two links of a graph are adjacent if they are incident with a common node.

Isomorphism: An isomorphism of labeled directed graphs is a one-to-one corre-
spondence on the links that preserves the adjacency, direction, and label of the
links.

The "match" of the two semantic-net chains X and Y can now be defined to be
the *maximal* graph isomorphism from a subgraph (subsequence) of X to the
subgraph of Y. By "maximal," we mean the isomorphism that pairs the largest
number of links correctly. Unfortunately, use of maximality precludes any
mathematical guarantee of the uniqueness of the resulting isomorphism. How-
ever, in practice, we have yet to be plagued by a nonunique maximal isomorph-
ism.

Note that we have defined "match" as a map that is an isomorphism between
*sub*graphs of the two deep structures. The map between deep structures is not
necessarily total (i.e., onto) in either direction (we are in the process of inves-
tigating a revision of this aspect of the definition as well as the interesting
situation where it is many-to-one and hence would have the properties of a
homomorphism). In other words, the analogy is a mapping that is a *maximal
partial graph isomorphism*. However, we abbreviate our terminology somewhat
and say that the analogy from A to B is formalized by the mp-morphism from A
to B (i.e., we speak of the analogy as being this structure-preserving map).

To replace the terms *intersection set* and *difference set*, we simply use *inter-
section subgraph* and *difference subgraph*. There are, of course, two difference
subgraphs for an mp-morphism—namely the residue portions of each of the deep
structures being compared. Throughout this chapter, we continue to use the
symbology of sets for these concepts, even though the designated entities are not
sets, but subgraphs.

Closeness Metrics.   Both the similarity task and the analogy task involve the
ranking of the match between two things or, rather, between their abstract de-
scriptions. The subject is asked to rank the degree of similarity or choose the
closest analogy. We assert that both kinds of judgments can be modeled by a
function over the intersection set (or subgraph) and two difference sets (or sub-
graphs). In similarity research, this three-argument function is often called a
"similarity metric," even though there are cases when the function is not a
proper mathematical metric (see Tversky, 1977). With the same sloppiness, we
call the function that ranks the closeness of analogies a *closeness metric*.

These metrics can be rather complex. Certain features might be more salient
than others, and one might model this difference by giving the former more
weight in a summation over the various sets. These metrics might even be
asymmetric,[2] which means they are not proper "metrics" in the strict mathemat-

[2]Tversky (1977) weighted the features in the set $A-B$ more heavily than the features in the set
$B-A$ in order to account for certain experimental data—for example, that "Red China is similar to
North Korea" has a lower degree of intuitive similarity than "North Korea is similar to Red China."

ical sense. In short, determining the intersection set and the two difference sets is not the end of the story for predicting similarity judgments; the metric can play a decisive role.

Monotonicity, etc.   We take the position that a precise statement of the closeness metric for procedural analogies can only be determined from detailed empirical studies. However, Tversky has shown that if certain formal conditions on the metric can be guaranteed, such as its monotonicity over subsumption of the intersection and difference sets, then the metric can have a simple, linear form (Tversky, 1977). (One of us—VanLehn—has investigated some of the conditions for procedural analogies and will discuss them in a later report.)

Individual Differences and Learning.   We have been speaking of *the* abstract description (or deep structure) of a thing as if this object were the same for all people. In some tasks, such as assessing the similarity of letters, it seems reasonable for literate individuals to have roughly the same representation language and the same abstraction functions for extracting descriptions from the letters. But this assumption is rather implausible in many other cases. In these cases, individual differences in conceptions of the things being compared is likely to influence judgments of the closeness of analogy. This would make verification of a theory significantly more difficult.

Individual differences affect analogy, but analogy also affects the individual differences. That is, one can learn from analogies. More specifically, when an individual understands an analogy, he or she may become aware of descriptive features that were previously overlooked. So a complete theory of analogy must allow for an evolution of an individual's conception of the things being compared over the course of testing.

In this research, we ignore these difficult methodological problems by assuming that the subjects who are judging the closeness of the analogies are *experts*. That is, they all have a complete representation of the things being compared and, hence, can be assumed to have roughly the same representations. Secondly, they already know all there is to know about the things being compared and therefore learn very little over the course of the testing.

FINDING THE RIGHT REPRESENTATION
FOR PROCEDURAL ANALOGIES

In this section, several candidate representations for procedures are examined as a basis for a theory that predicts the closeness of procedural analogies.[3] Possible

[3]The judgments of closeness are those of experts on arithmetic and so can be taken to reflect the teleologic semantics of arithmetic.

representations range from a very superficial one—namely, a simple chronological list of actions—on up to a very abstract representation that involves goals, constraints, and other planning knowledge—namely, planning nets. Our research has shown that planning nets are the only serious contender, so the discussion of the others is quite brief. However, the more superficial representations are mentioned in this section for a reason—namely, to show how a human (or machine) can construct a very abstract representation of a procedure by ascending through several levels of representation. We do not claim that the structure of this section models the abstraction process that a person executes when assimilating a procedural analogy, but it does provide an indication of the complexity that such a process would have to have.

Traces

The *trace* of a procedure is simply a chronological list of the actions it performed during one particular execution. This representation of a procedure can be constructed directly from observation of the execution of the procedure (although there are the usual problems in choosing the "grain size" of primitives).[4] However, traces are a highly inappropriate representation for procedures, as the following example indicates.

Consider an analogy between Dienes Block addition and written addition. These two traces would probably have few, if any, action labels that match. The action "write '4' " would have to be matched against a group of four actions labeled "place one block in the pile," whereas the action "write '8' " would have to be matched against a group of eight block-placing actions. Such sophisticated matching could not be represented by an mp-morphism. Indeed, the match seems to require the *concept* of "write n" and the *concept* of "repeat single block placement n times." These are abstractions over action sequences and so should be part of the representation rather than the matching mechanism. Incorporating such concepts into the representation lifts us to the next level of abstraction.

Flowcharts

By generalizing over a large collection of traces, one could derive a notion of the observed procedure that could be represented with a programming language,

---

[4]The folklore about protocol taking, supported by a few experiments (Card, 1978), is: When in doubt, use a finer grain size. If the grain size is too large, one might miss distinctions. If one errs the other way and makes the grain size too fine, then one creates more work for oneself; yet if one is tenacious, the relevant distinctions will ultimately appear, probably as relations between *groups* of actions instead of single, individual actions. So it appears that the grain-size issue (and a very similar issue—the choice of primitive actions) is more of a practical trade-off than an insurmountable source of uncertainty in the theory.
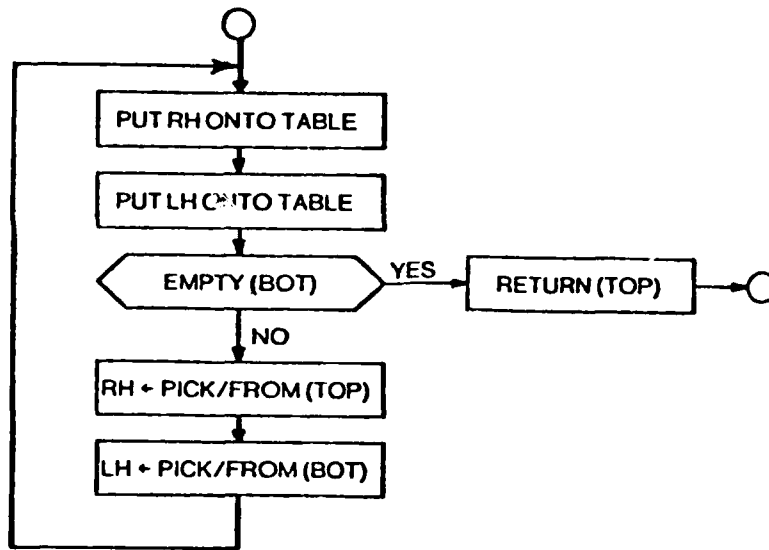
**FLOW CHART FOR A BASE-1 BLOCK SUBTRACTION
PROCEDURE USING TWO HANDS**
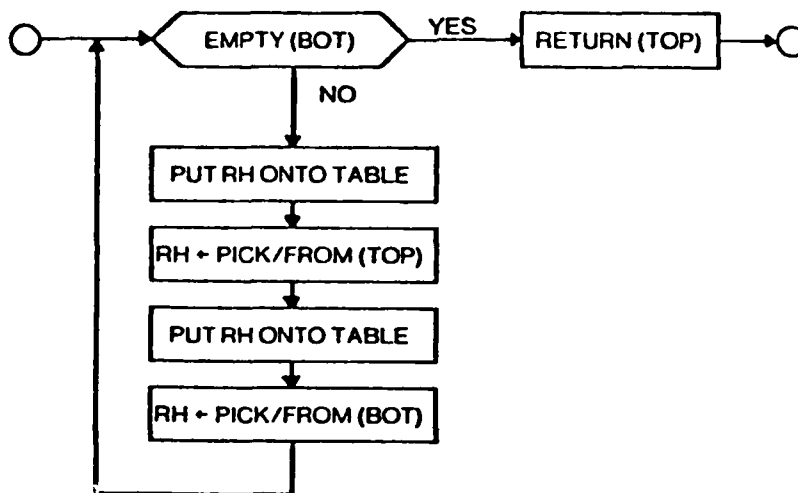
**FLOW CHART FOR A BASE-1 BLOCK SUBTRACTION
PROCEDURE USING ONE'S LEFT HAND**

such as flowcharts. Granted, this generalization would be nontrivial: Repetitious sequences of actions would become loops; objects that are manipulated similarly become the contents of variables, and so on. Nonetheless, constructing a program from examples is well within human ability.

However, flowcharts would also be a poor representation for analogy. Consider a simple subtraction procedure for numbers represented as base-1 blocks as illustrated by the lower flowchart on pg. 104. The primitive terms used in this flowchart are as follows: LH stands for someone's left hand. TOP and BOT stand for place mats on the TABLE. The BOT set of base-1 blocks is subtracted from the TOP set of blocks by pairing off a block from each, using the primitive actions PICK/FROM and PUT/ONTO and tossing them onto the table. When the bottom "number" is "zero" (i.e., empty), whatever is left in the top "number" is the answer. However, notice that by merely shuffling the order of the steps somewhat and using two hands instead of one, a new procedure can be constructed that is intuitively very similar to the old procedure; yet its flowchart (see pg. 104) shares virtually no isomorphic subgraph with the old procedure's flowchart. Because the intersection graph is so small relative to the difference subgraphs, a reasonable closeness metric would have to report that the two procedures are not very close—a false prediction. So for this and other reasons, flowcharts also seem to be a poor representation or level of abstraction for procedural analogies.

Procedural Nets

On the basis of the foregoing example, it might appear that flowcharts are too committed to a set order of performing steps, since the two base-1 flowcharts have the same steps but order them slightly differently. Also, these charts lack the typical hierarchy of subprocedures that is used in computer programs to modularize and organize the procedure. This suggests using a structure that emphasizes the subprocedure hierarchy and deemphasizes the temporal sequence of subprocedures.

Just such a structure has been developed for modeling children's bugs in arithmetic procedures—namely, BUGGY's *procedural-net* representation (Brown & Burton, 1978). Although we do not pause here to explain this representation, a procedural net for a very familiar procedure—namely, standard subtraction—is included as Fig. 18.1. However, procedural nets also fail as a basis for a theory of analogy, as illustrated in the following example.

Consider two Dienes Block subtraction procedures: (1) in "big-pile" Dienes Block subtraction, a number is represented by one big pile of Dienes Blocks; (2) in "sorted" Dienes Block subtraction, all the blocks are kept sorted into little piles according to their shape. Intuitively, these two procedures are quite closely analogous. But when the procedural nets are formed and the matching is done, we find the following statistics:
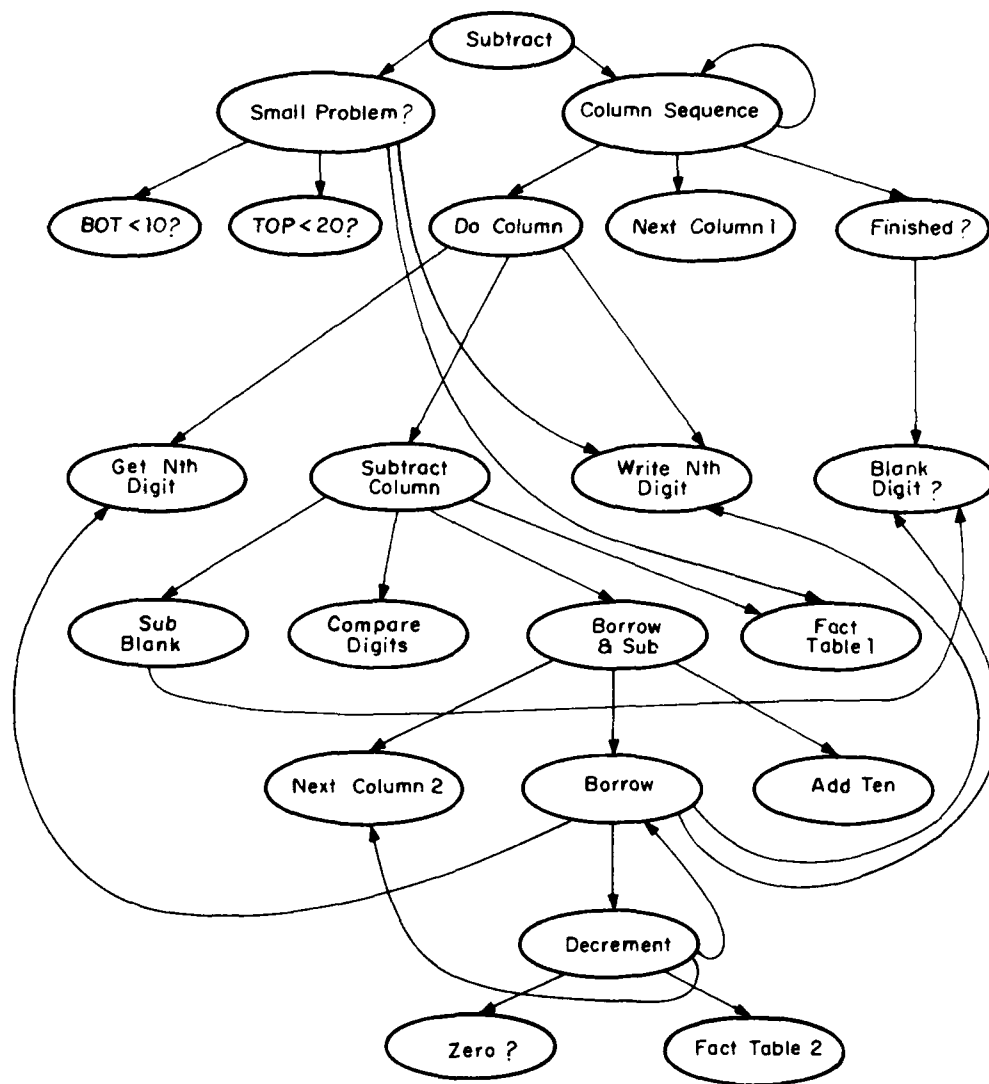
FIG. 18.1.

$A \cap B$ contains 6 nodes.

$A - B$ contains 10 nodes.

$B - A$ contains 16 nodes.

The intersection subgraph is far too small compared to the difference subgraph for this analogy to be rated "close" by any reasonable metric. So again, we must abandon a representation and look for a higher level of abstraction.

Planning Knowledge Seems Necessary

Both flowcharts and procedural nets are at the "program" level of abstraction. That is, they both are close to the sorts of languages one sees for computer programs. The problem with this level of abstraction seems to be that some design decisions that do not seem so consequential to the intuition have an enormously large effect on the "program." The framework that analogy seems to require is something that extracts these sorts of choices out of their final manifestation, makes them explicit, and relates them in a reasonable way to other, more important elements of the design. In short, what seems necessary is a representation of the design process behind a procedure—this allows one to say which choices are important and which are relatively minor. The process of creating a procedure from a set of constraints is traditionally called "planning" by the artificial intelligence community. So the abstract representation that analogy seems to require appears to involve planning knowledge and planning inferencing.

Planning knowledge includes not only the functional decomposition of the surface structure of the procedure but also the reasoning that was used to transform the goals and constraints that define the *intent* of the procedure into its actual surface structure. The formalism we use to represent this knowledge, we call *planning nets*. These planning nets are an extension of Sacerdoti's pioneering work on representing procedural knowledge for robotics (Sacerdoti, 1977). Before presenting the formalism (which lies at the heart of the remaining parts of the chapter), it is best to get some idea of what this "planning knowledge" is that is going to be incorporated into the representation. To this end, we plan out a very simple subtraction procedure, called "base-1 blocks subtraction," that represents a number as a pile of unit blocks. Later, we show how planning nets capture this knowledge in a summary form.

Constraints and Planning Heuristics

The basic idea of formal planning is to take a declarative, rulelike presentation of the goals of the procedure and the world in which it is to be implemented, and transform them into a surface structure that achieves the goals while remaining inside the constraints imposed by the world. There is always an element of common sense in planning, and as this is formal planning, use of common sense must also be recorded.

These two knowledge sources are called *constraints* and *heuristics*. Both can be represented as pattern-action rules in some suitable formal language, but for our purposes, English will suffice.

The constraints that characterize base-1 blocks subtraction are listed next:

1. Goal: If EMPTY (BOT) then return TOP as the answer (i.e., $n - 0 = n$).
2. The decrease in TOP must EQUAL the decrease in BOT (i.e., a recursive definition of subtraction).
3. a is EQUAL to b (i.e., all blocks are equal).
4. Over the action (Y ← PICK/FROM(X)), the decrease in X is EQUAL to the increase in Y (i.e., blocks are conserved over the picking-up action).
5. Over the action (PUT Y ONTO X), the increase in X is EQUAL to the decrease in Y (i.e., blocks are conserved over the putting-down action).
6. The action (Y ← PICK/FROM(X)) requires EMPTY (Y) beforehand (i.e., the hand must be empty before picking up a block).
7. The action (PUT Y ONTO X) entails EMPTY (Y) afterwards (i.e., the putting-down action always empties the hand completely).
8. ~ EMPTY (X) before the action (Y ← PICK/FROM(X)) entails that afterward, there exists a such that a is the contents of Y (i.e., the hand picks up exactly one block).

The meaning of the primitives is as follows. TOP and BOT are place mats on the TABLE. The subtraction problem $n - m$ would begin with n base-1 blocks on TOP and m on BOT (N.B., this is not the way base-1 block subtraction is ordinarily posed in the classroom).[5] There are two hands, LH and RH, which can perform two kinds of actions—namely, picking up one block (PICK/FROM) or putting down a block being held (PUT/ONTO). The primitive predicate EQUAL takes two piles of blocks and says whether they designate the same number. EQUAL is not executable and cannot appear in the final plan.

The foregoing constraints describe the mathematical goals of the procedure, the objects it works with, and the physical manifold within which it operates. The mathematical content of subtraction is expressed in constraints 1 and 2: TOP minus BOT is TOP whenever BOT is empty of blocks, but any changes in the number of blocks on BOT must be echoed by an equal change in the contents of TOP. The objects the procedure manipulates are base-1 blocks. Because these are very simple, constraint 3 suffices to describe them. (By convention, a lowercase letter stands for an arbitrary block, whereas an uppercase letter stands for an arbitrary place mat or hand.) The remaining constraints define the physical manifold within which the procedure will operate. Constraints 4 and 5 ensure that blocks are *conserved* by the actions PICK/FROM and PUT/ONTO. Constraints 6, 7, and 8 describe how the hands that manipulate the blocks work. A complete descrip-

[5] Dienes Block subtraction and other block subtraction procedures are usually taught using oral or written presentations of the problems. Thus, to solve $n - m$, the first step is to translate n into blocks, using some "bank" as a source of blocks. Next, one translates m into blocks, but *uses the first pile as the source.* When one is finished translating, the first pile contains $n - m$ blocks. This procedure for doing block subtraction is so dissimilar to written subtraction that we have avoided using it in this paper.

tion of the workspace would require several more constraints, but these will do for purposes of illustration.[6]

The constraints describe *domain-dependent* knowledge. If the procedure's goals or implementation environment change, then the constraints must be changed to reflect this. For example, if one used Dienes Blocks instead of base-1 blocks, then constraint 3 would be replaced by a new constraint, namely:

3'. *a* is EQUAL to *b* if and only if SHAPE(*a*) = SHAPE(*b*).

If one wished to plan an addition procedure instead of a subtraction procedure, then constraint 2 would become:

2'. The increase in TOP must EQUAL the decrease in BOT.

Heuristics are presupposed to be *domain-independent* knowledge. They represent commonsense planning knowledge, such as: "When you need to accomplish two things, and it doesn't matter which comes first, then pick one arbitrarily, do it first, then the other." We include this distinction between constraints and heuristics only because it is traditional; nothing in our theory turns on this distinction.

Planning a Base-1 Subtraction Procedure

The planning of the base-1 subtraction procedure involved 12 steps. Each step is an application of a constraint or a planning heuristic. The planning begins with a flowchart initialized to the constraint that is marked as the "goal" of subtraction.
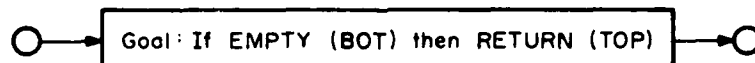
Planning proceeds by progressive refinement of goals to subgoals, or by checking the current plan against the constraints. (N.B., Because we are only interested in having a correct planning net for a procedure, not in *finding* one, we are going to ignore a few of the subtle issues.)[7]

---

[6]In formulating constraints, it is very important to put as little into each constraint as possible. For example, we could have replaced constraint 2 by "decrementing BOT by 1 must be echoed by decrementing TOP by 1." Although adequate for base-1 subtraction, this is not the most general statement of the constraint, and, indeed, this constraint would have to be replaced to handle Dienes Block subtraction. The basic idea is to split the declarative description of the world and the goal as finely as possible, so that small variations on the procedure can be modeled by replacement of one constraint among many small ones, rather than modification of one clause of a large, special-purpose constraint.

[7]We will gloss over a number of very difficult issues in the presentation of the planning steps. For instance, why was the TABLE chosen in Step 5 as the location for emptying the LH? How did we

Step 1: At the outset, the "implication-reduction" planning heuristic that reduces an implication ($A \supset B$) to a sequence of subgoals (A, B) can be applied. The second subgoal in this case is a primitive of the workspace. So the output of Step 1 is a plan with just one subgoal:

```
O──▶│Goal: EMPTY (BOT)│──────▶│RETURN (TOP)│──▶O
```

Step 2: A venerable planning heuristic, traditionally called "hill climbing" (Newell & Simon, 1972), reduces the goal to a loop. The loop test sees if the goal has been achieved, and if not, it takes a step "up the hill," so to speak.

```
O──▶⟨EMPTY (BOT)⟩──YES──▶│RETURN (TOP)│────▶O
           │
           NO
           │
           ▼
     │Goal: Reduce BOT│
```
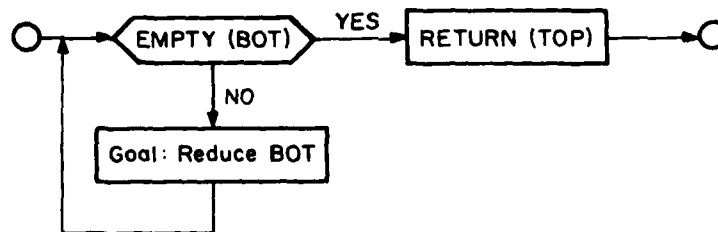
Step 3: The goal matches part of constraint 4—the definition of PICK/FROM. So the constraint is applied, and the plan is now fully reduced to primitive actions:

```
O──▶⟨EMPTY (BOT)⟩──YES──▶│RETURN (TOP)│────▶O
           │
           NO
           │
           ▼
     │LH ◀── PICK/FROM (BOT)│
```
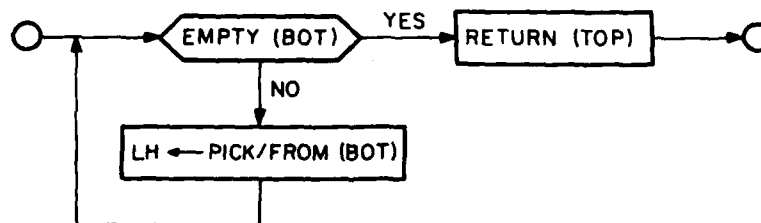
Step 4: Execution of this plan reveals a violation of constraint 6: The left hand must be empty before one can pick something up. So a new goal is created:

---

know not to empty it on TOP or BOT? Only the successful reasoning will be presented—the alternatives that didn't work aren't mentioned. Most of the research in planning for robotics has gone into improving the search for correct plans by recognizing unworkable alternatives and recovering from them gracefully. All these difficult questions involving search can be ignored, because we are not interested in automating the discovery of planning nets.
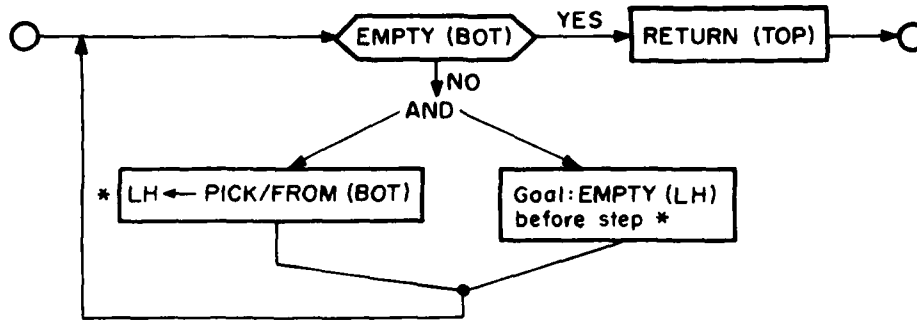
Step 5: This goal is quickly dismissed by applying constraint 7—part of the definition of PUT/ONTO. The left hand is now emptied before use.
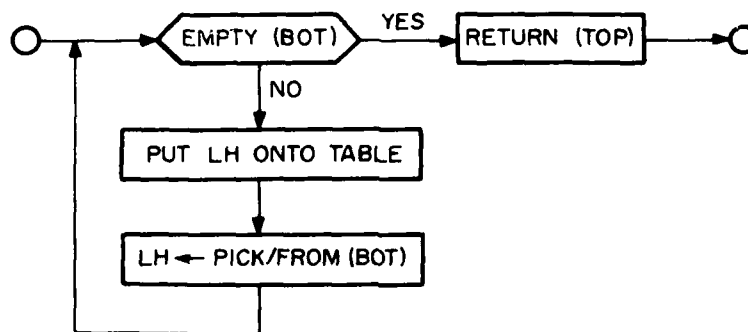
Step 6: Execution of this plan uncovers a violation of constraint 2. Because the bottom place mat is not empty when PICK/FROM is executed, one knows from constraint 8 that the left hand comes to hold exactly one block. Via constraint 4, one infers that the bottom place mat has its contents decreased by PICK/FROM. But there is no way to show that the TOP place mat undergoes an equal change. So constraint 2 is violated, and a new goal must be created. The goal says that there must be a change in TOP to equal the change in BOT.
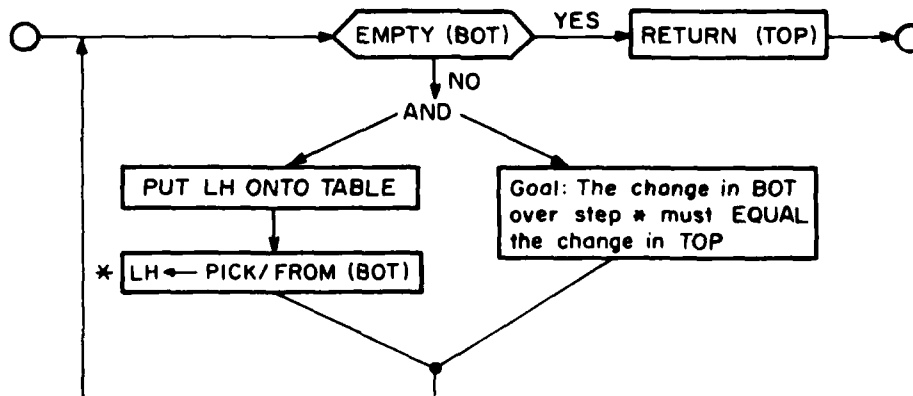
Step 7: Part of this goal matches constraint 4, the definition of PICK/FROM. A new picking-up action is instantiated for the top place mat. This reduces the goal of equal changes to the goal of equal contents of the left and right hands.
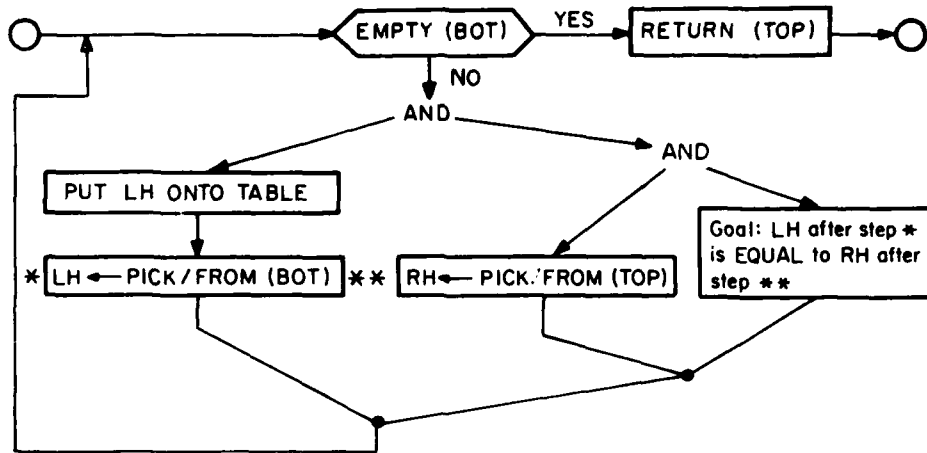
Step 8: *Constraint 8 can apply twice now, once per hand. It says that only one block is picked up by* PICK/FROM. *Thus, the goal of* EQUAL *contents is replaced by equality of two arbitrarily chosen blocks.*
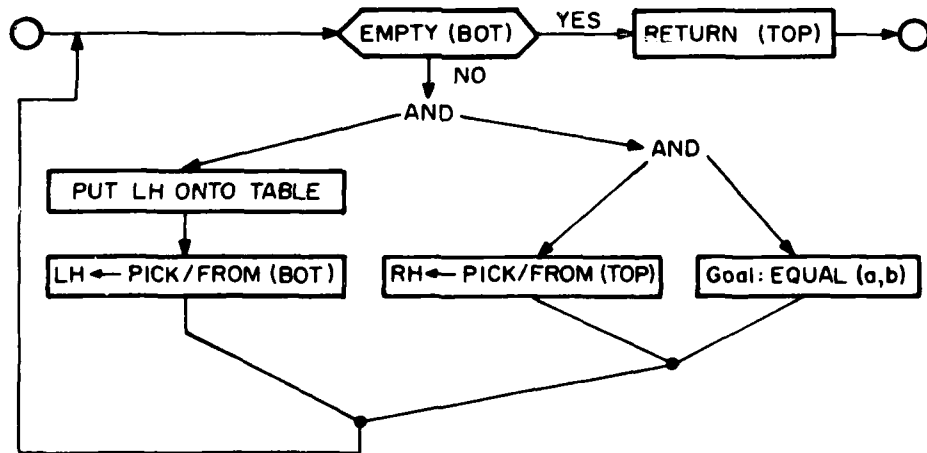
Step 9: Of course, this new goal is trivially satisfied by constraint 3—all blocks count the same in the base-1 number system. So the goal is simply removed from the plan.

```
  O─────────────────────⟨EMPTY (BOT)⟩──YES──▶│ RETURN (TOP) │──▶O
                                 │ NO
                                 ▼
                             AND
                       ┌──────────┴──────────┐
                       ▼                      │
              ┌────────────────────┐         │
              │ PUT LH ONTO TABLE  │         ▼
              └────────────────────┘   ┌──────────────────────┐
                       │               │ RH◀─PICK/FROM (TOP) │
                       ▼               └──────────────────────┘
              ┌────────────────────┐         │
              │ LH◀─PICK/FROM (BOT)│         │
              └────────────────────┘         │
                       └──────────┬──────────┘
                                  ●
```
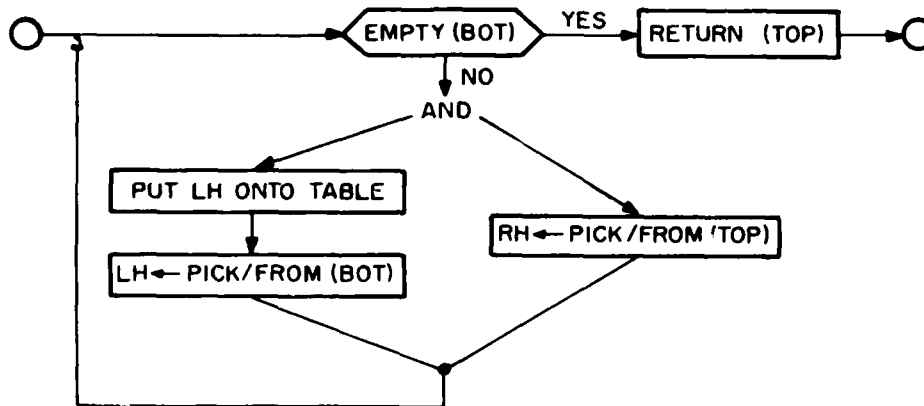
Steps 10 and 11: Execution reveals that constraint 6 is violated again, this time by the right hand. So it must be emptied before use as well, in the same two-step fashion as Steps 4 and 5.

```
  O─────────────────────⟨EMPTY (BOT)⟩──YES──▶│ RETURN (TOP) │──▶O
                                 │ NO
                                 ▼
                             AND
                       ┌──────────┴──────────┐
                       ▼                     ▼
              ┌────────────────────┐  ┌────────────────────┐
              │ PUT LH ONTO TABLE  │  │ PUT RH ONTO TABLE  │
              └────────────────────┘  └────────────────────┘
                       │                     │
                       ▼                     ▼
              ┌────────────────────┐  ┌────────────────────┐
              │ LH◀─PICK/FROM (BOT)│  │ RH◀─PICK/FROM (TOP)│
              └────────────────────┘  └────────────────────┘
                       └──────────┬──────────┘
                                  ●
```
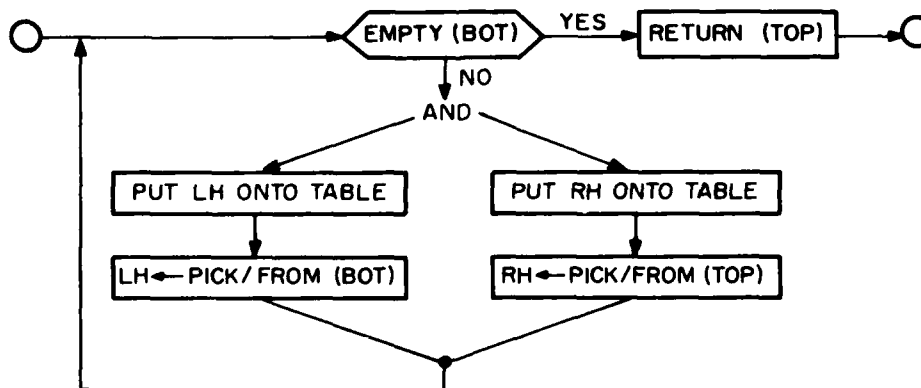
Step 12: A planning heuristic, call it "conjunction reduction," removes the conjunction AND. The AND node is for conjoining subgoals. It makes no statements about which subgoal to achieve first. In this case, it doesn't matter how the subgoals are ordered since they turn out to be independent. So the rule arbitrarily chooses the following ordering:
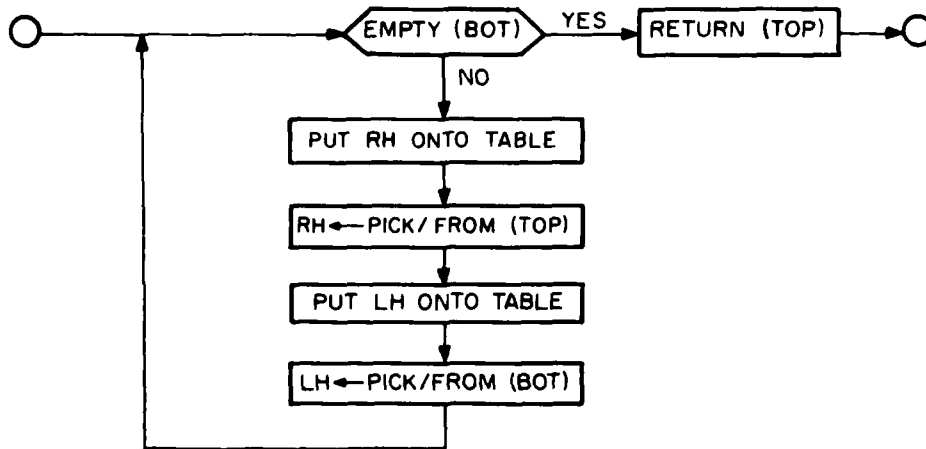
This is the final plan. Every step is a primitive, and all the constraints check out. The planning for base-1 subtraction is complete. The final plan is exactly the flowchart representation of the surface structure of the procedure.

Planning Nets

Planning nets are directed graphs. The nodes of the net represent plans, and the links represent planning inferences. That is, each node of the net stands for a flowchart containing a mixture of primitive actions and subgoals to be expanded. Two nodes are linked only if the application of some constraint or heuristic to one plan results in the other plan. The link is labeled with the planning rule that causes the change.

Sacerdoti developed a very similar structure to aid in automated task planning and monitoring in robotics. It is remarkable that we have found it useful for our research on procedural semantics, as has Greeno for his research on modeling the counting behavior of children (Greeno, Gelman, & Riley, 1978). However, we are faced with a clash in nomenclature. Sacerdoti calls these sorts of structures "procedural nets." We prefer to call them "planning nets," because their content has more to do with the planning of a procedure than with the procedure itself.

Planning Nets Are Partial Orders.   In fact, planning nets are generally not sequences as the chronological presentation of the previous subsection might lead one to believe. Often, two planning inferences can be applied in either order. For example, step 6 could have preceded steps 4 and 5. To represent this independence, we allow the net to be a *partial order*.

Figure 18.2 shows the planning net for base-1 subtraction. In addition to the names of the planning rules, the steps have been labeled with the step numbers used in the previous subsection. The split at steps 4 and 6 occurs because

FIG. 18.2.

115
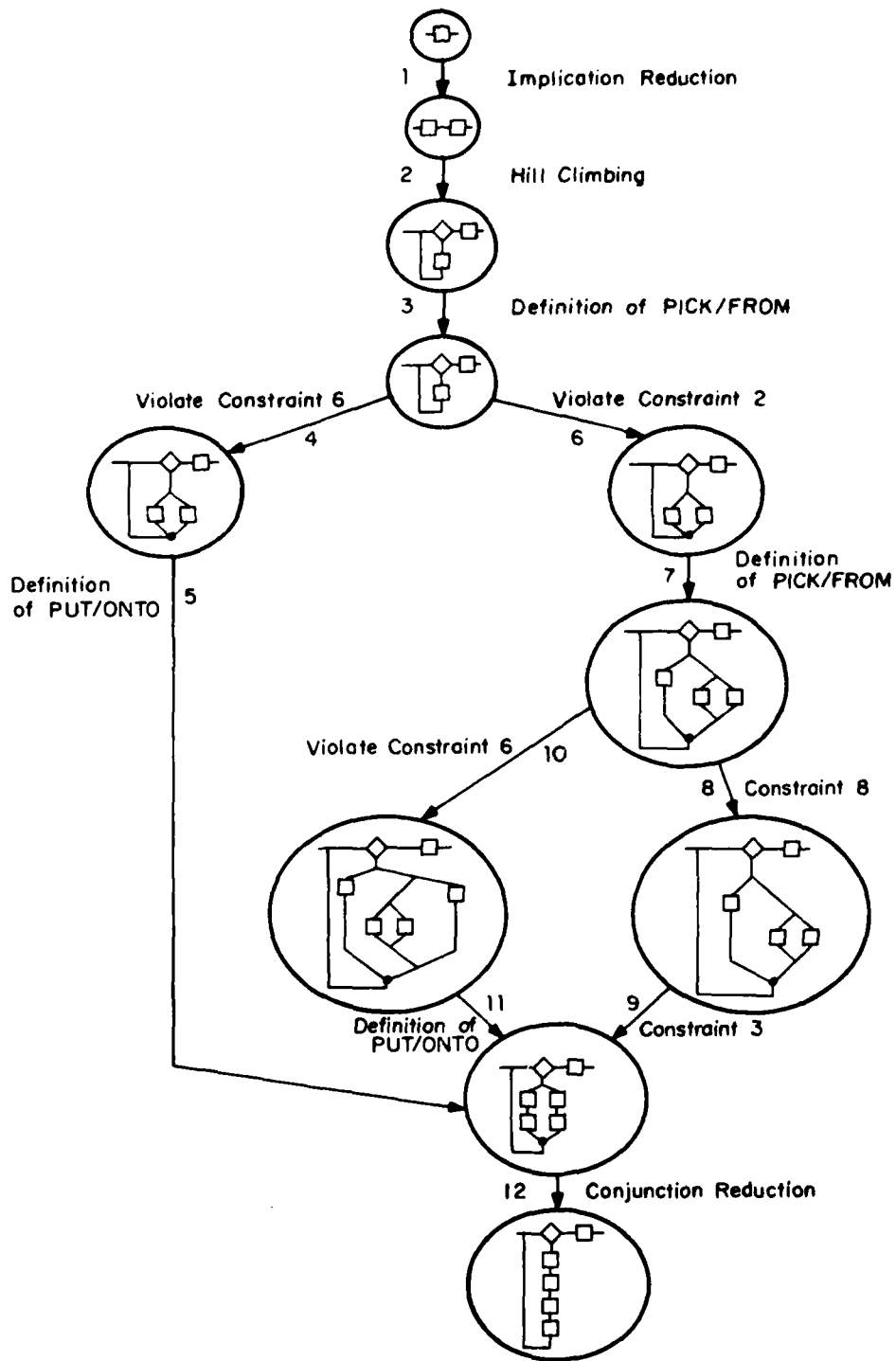
constraints 2 and 6 can be fixed independently. The other split shows that constraint 6, applied this time to the right hand, can be fixed independently of the subgoal reduction due to constraint 8.

Planning Nets Are a Complete Representation.   The previous section may have left the impression that planning knowledge must be represented in three parts: the constraints, the planning steps, and the ultimate surface structure; and that planning serves as a transformation of the constraints into the surface structure. Although this is not a bad way to think of planning, it is unnecessarily redundant. The planning nets alone capture all three kinds of information. The constraints that are relevant to the procedure are exactly those constraints that appear as edge labels. Similarly for the heuristics. The surface structure is the contents of the bottom node, the final plan. So, planning nets are a complete representation of the design of a procedure.

Planning Net mp-Morphisms Formalize Procedural Analogies

To formalize procedural analogies, one merely applies the definition of "match" for directed graphs that was given in a previous section. That is, a procedural analogy is formalized as a graph-theoretic mp-morphism between the planning nets of the two procedures. We illustrate this definition with an example.

Figure 18.3 shows the planning net for a "big-pile" Dienes Block subtraction procedure. This procedure has the same sort of pairing-off action as the base-1 procedure discussed earlier, but it represents a number as a big pile of Dienes Blocks. Although space does not permit labeling the links in the planning net with their planning inferences, the step numbers should be sufficient to describe the match with the planning net of base-1 subtraction, which appears in Fig. 18.2. Step 9 of Fig. 18.2 is replaced in Fig. 18.3 by a subgraph consisting of steps 9.0 through 9.7. So all the links of Fig. 18.2 match the correspondingly numbered links in Fig. 18.3 except for link 9. The reason why link 9 can't be matched is simple: It is the application of the constraint that makes base-1 blocks all count the same—namely, constraint 3. In Dienes Blocks, all blocks do not count the same. Only if they are the same size do they designate the same number. What the subgraph of steps 9.0 through 9.7 is doing is planning out a way to get blocks that aren't the same size to be the same size by doing the appropriate trading. In fact, the planning leads off in step 9.0 by noticing a violation of the constraint 3', which says: "Only blocks that are the same size count the same."

The mp-morphism of the two planning nets results in the following intersection and difference subgraphs (calling the Dienes Block procedure A, and the base-1 procedure B):
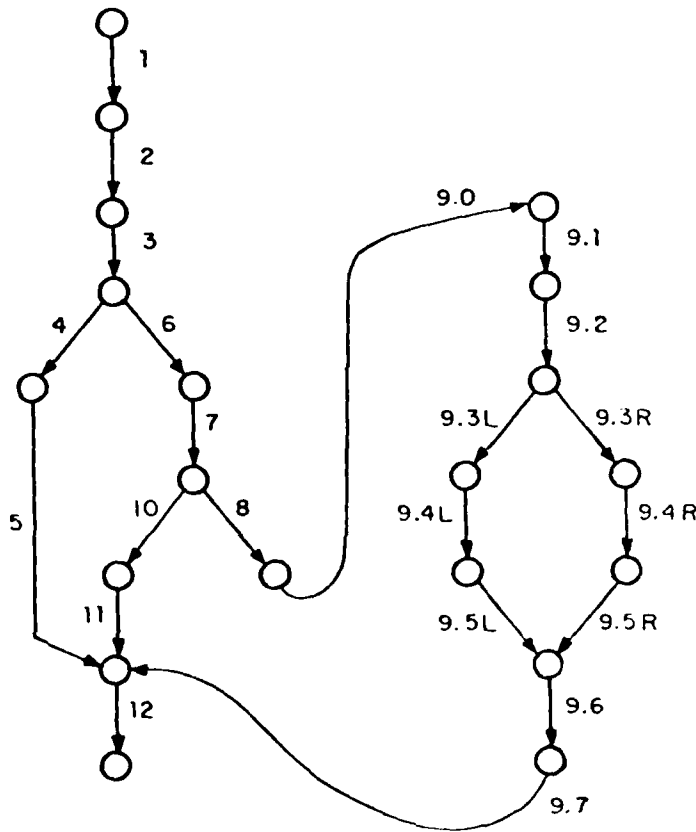
FIG. 18.3.

$A \cap B$ is almost the whole planning net for base-1 subtraction except the link for step 9.

$A - B$ is the subgraph that replaces step 9, whose steps are labeled 9.0, 9.1, and so on.

$B - A$ is just step 9 of the base-1 planning net.

The $A - B$ subgraph is almost the same size as the intersection subgraph, indicating that the closeness metric would probably give the analogy a rating of "moderate," which corresponds with the intuition nicely.

Difference Generators Are Used To Predict Closeness

As we hinted earlier, it is not always the case that the predictions based on the relative sizes of the intersection and difference subgraphs correspond so nicely

with the intuition. However, in those cases, the problem has been immediately apparent and was fixed, utilizing the fact that planning nets are *partial orders*.

To illustrate the problem, a new analogy is introduced and compared to the one described in the previous subsection. Whereas the earlier example was, intuitively, a moderately close analogy, this new analogy is quite a bit closer still. However, the simple view of the closeness metric as corresponding to the relative sizes of the intersection and difference subgraphs leads to the false prediction that the old analogy is actually closer than the new one.

Suppose we compare big-pile Dienes Block subtraction to sorted Dienes Block subtraction, an analogy that earlier provided a counterexample. For convenience, let us attach some letters to these procedures and the ones used in the earlier analogy:

A: base-1 subtraction

B: big-pile Dienes Block subtraction

C: sorted Dienes Block subtraction

The BC analogy is intuitively rather close. However, when the planning nets are compared, we find a huge subgraph of C that isn't matched—namely, all the design that has to do with maintaining the sort. Indeed, this difference subgraph, $C - B$, is much larger than $B - A$ and $A - B$ together. Subgraph $B - C$ is also quite large. Hence, even though $B \cap A$ is somewhat smaller than $B \cap C$, any reasonable metric would predict that analogy AB should be closer than analogy BC, contrary to the intuition that big-pile Dienes Block subtraction is more similar to sorted Dienes Block subtraction than to base-1 block subtraction. There is a mismatch between predictions of the theory and judgments of closeness.

But closer examination of subgraph $C - B$ reveals it has only one entering link, just like link 9.0 of Fig. 18.2. This link is labeled "Violates Constraint 11: keep blocks sorted by size." In other words, it appears that one plan inference is *causing* all the others. We can capture this notion of causation by utilizing the *topology* of planning nets.

As already discussed, planning nets are partial orders. Any subgraph of a partial order is also a partial order. In particular, the difference subgraphs are always partial orders. Any partial order has a unique set of *minimal elements*. This set is the smallest set of links that dominate all the other links in the subgraph. These mathematical facts ensure that the following terms are well defined:

Where X and Y are any two planning nets, let $d(X - Y)$ be the links that are the minimal elements of the difference subgraph $X - Y$, and let $d(Y - X)$ be the links that are the minimal elements of $Y - X$. Call these two sets the *difference generators* of mp-morphism XY.

Difference generators are a formal representation of what is causing the difference between two procedures. Intuitively, what the difference gener-,,ors of mp-morphism represent are the *crucial ideas* that separate the two procedures. All the other differences between the two procedures stem from these few crucial ones.

To illustrate this notion of "crucial ideas," take the analogy between base-1 and big-pile Dienes Blocks, which we were calling analogy AB in the previous section. $d(B - A)$ is a graph with just one link, labeled "Step 9: Constraint 3—all blocks are EQUAL." $d(A - B)$ is a link labeled "Step 9.0: Constraint 3'—two blocks are EQUAL if and only if they have the same SHAPE." Replacing constraint 3 by constraint 3' is about as clear a statement of the difference between base-1 blocks and Dienes Blocks as one can hope to make.

Because difference generators capture the distinctions between procedures so succinctly, they seem highly appropriate as the inputs (or arguments) to the closeness metric. They are decoupled from the unimportant details that fill flow-charts, procedural nets, and planning nets—details that obscure the essence of analogy by inflating difference subgraphs with derived, less meaningful structure. Indeed, the comparison of analogy AB to analogy BC (i.e., the big-pile vs. sorted analogy) now agrees with intuition: All four difference generators— namely, $d(A - B)$, $d(B - A)$, $d(B - C)$, and $d(C - B)$—are about one link big. On the other hand, the intersection subgraphs are as before, with $A \cap B$ being smaller than $B \cap C$. Because the difference generators are about the same size, the intersection sets are more important in the closeness metric. Hence, a reasonable metric would report that BC is closer than AB, which corresponds with the intuition that big-pile Dienes Blocks subtraction is closer to sorted Dienes Block subtraction than to base-1 blocks subtraction. At last, we seem to have found a level of abstraction for procedures where intuitions of closeness correspond to the relative sizes of the inputs to the closeness metric.

Discussion

The main point of this section has been that planning nets provide a basis for a theory of analogy that can predict the judgments of experts on the closeness of analogies between procedures. Moreover, all the aspects of the theory have very natural, almost elegant sources. The deep structure used came naturally from Sacerdoti's work in robotics; mp-morphisms are a general-purpose concept; and the notion of difference generators came naturally from the topology of planning nets.

We have always been struck by how much of the design of a procedure like subtraction is governed by the design of the representation of the objects manipulated by the procedure (e.g., the place-value number system). In fact, many of the actions in any of the elementary arithmetic procedures concern not the mathematical operation per se but rather *how the object representations are*

manipulated. This impression is reinforced by experience in computer programming, which is often a constant interplay between the design of the object (i.e., data) representation and the code, even at the highest levels. Anyone who has tried to understand a program that he or she did not write can vouch for the importance of understanding the data representation. In the process of judging the closeness of an analogy, a popular strategy is first to look at each procedure's object representation, and then to build the understanding of the overall analogy on the basis of the analogy between object representations. In short, it appears to us that a large portion of the "understanding" of a procedure consists of an understanding of the implications of the procedure's object representation.

This view of procedural understanding is entirely consistent with the planning-net formalism. The constraints and heuristics that appear in the net represent, in some sense, the *essence* of the procedure. If object representations were unimportant, then none of the planning inferences would be "about" the object representation. But, in fact, many planning inferences do deal with the object representation. Even in the foregoing base-1 blocks procedure, with its extremely simple object representation, we find constraint 3 addressed solely to the object representation. In more complex procedures, using Dienes Blocks or written numerals, an even larger portion of the constraints concern the object representation. In short, although planning nets abstract out the less important aspects of a procedure, they leave behind the design of the object representation, which is quite compatible with the view that as a representation of "understanding" of procedures, a fair portion of the design should model the "understanding" of the object representation.

We have not discussed the exact definition of the closeness metric, even though some definition would be necessary to verify methodically the correlations we have claimed. There are many difficulties and fine points involved in determining such a definition. In particular, it is plausible that the weight of some planning inferences is quite close to zero. We have in mind the commonsense heuristics, such as implication reduction, that play an almost invisible role in the planning. Also, some planning rules are applied more than once in a planning net; one may perhaps wish to avoid giving such rules an inappropriate prominence by only counting their first occurrence in the difference generators or the intersection sets. These are just two of the many points one would have to consider in defining a closeness metric.

The reader has no doubt noticed the incredible amount of work that goes into analyzing a procedure in terms of its planning. First one constructs the flowchart, then the constraints and a sequential plan for the flowchart, and last calculates the planning net by noting which planning inferences are not ordered with respect to each other. This large amount of work leaves much room for error on the part of the theorists. However, each level of abstraction is well defined and can be checked for consistency by a computer. Thus, one next step is to build a com-

puter system of utilities to aid in the analysis of procedures. However, there is a certain amount of intuition that goes into some parts of the analysis, notably the formulation of a set of constraints, that we doubt could ever be successfully mechanized.

ANALOGIES AND TELEOLOGIC SEMANTICS IN EDUCATIONAL RESEARCH

In this section we consider some of the issues involved in explaining (or teaching) the knowledge we discussed in the first previous section—teleologic semantics. Briefly, teleologic semantics is the kind of knowledge that concerns the purpose of each part of the procedure, as well as the motivation behind the set of constraints that defines the particular representation for the objects. In particular, we consider how an individual piece of teleology can be explained, and how such individual explanations can be combined into an integrated explanation.

The section closes with a discussion of some issues involved in microworld-based curricula. These issues turn out to be intimately related to those involved in teaching teleologic semantics.

Local Explanations: Manifestation and Motivation

An important property of the planning-net formalization is that there is a *natural* notion of how to explain a small piece of a procedure's teleologic semantics. By "piece" we mean a constraint (or a small set of constraints) that is used in the planning net. To "explain" it, one uses a *minimal contrasting pair* of procedures—one with the constraint, and one without it—that compute the same "operation" as the given target procedure. In other words, we use analogies to illustrate constraints. We believe that using a concrete surface structure illustration for each deep structure concept is a very important explanatory technique that naturally falls out of this development. For example, this method frees us from having to explain the planning formalism to the student—a task potentially more difficult than teaching the procedure itself.

More formally, to illustrate some given constraint(s), one uses *two analogous procedures such that one of the difference generators of the mp-morphism between them is exactly the given constraint(s)*. If the pair of procedures forms a minimal contrasting pair, then the mp-morphism constituting the analogy is *elementary*.

Of course, this technique works just as well for explaining heuristics. However, heuristics are often such commonsense knowledge that an explanation of them is unnecessary. So we call the planning inferences to be explained "constraints," avoiding the cumbersome phrase "constraints or heuristics." Also,

our terminology reflects the fact that it is often possible to provide a minimal contrasting pair for each constraint individually (this observation is discussed later). So we use "constraint" in place of "a small set of constraints."

An important realization is that minimal contrasting pairs can be used in two different ways in an explanation. They can be used to show how the constraint is *manifested* on the surface, and they can also be used to *motivate* the inclusion of the constraint in the ultimate design of the procedure. Probably the best way to illustrate the differences between these two uses is with an example.

Explaining the Canonicity Constraint.   The particular constraint that is used in this example is one of the most subtle and influential in arithmetic—namely, the *canonicity constraint*. To show how the planning-net representation can aid in explaining procedures, the constraint is presented as the "answer" to a nontrivial teleologic question.

What is the purpose of carrying? More specifically, if the problem is 52 + 49, why bother to carry 10? Why not leave 11 in the units place? It is not because there is no symbol for the "digit" 11—we could invent one if we wanted. In Dienes Block addition, the question is even clearer. Why not leave the answer in the form of 9 longs and 11 units? Why bother carrying?

The answer is that carrying maintains the *canonicity* of the representation of numbers. A canonical representation puts the representational objects in one-to-one correspondence with the real objects they represent. The Hindu-Arabic representation of numbers is canonical because there is a unique, distinct numeral for each number. Dienes Blocks are not necessarily a canonical representation, since most numbers can be represented several ways. For instance, 11 can be represented as a long and a unit, or as 11 units. The purpose of carrying is to canonicalize the sum by making sure that there are no more than nine blocks of any given shape. In other words, carrying is the *manifestation* of the canonicity constraint.

But suppose that the questioner rejoins by asking what the purpose of the canonicity constraint is. The answer involves another arithmetic subprocedure—comparison.

It is much more efficient to find out which numeral represents a given large number if the representation is canonical. Let us use a Dienes Blocks comparison procedure to illustrate the gain in efficiency. In a noncanonical representation, the comparison procedure must compare all the piles, because a very large pile of small blocks can make up for a deficit of larger blocks. In a canonical representation, the comparison procedure needn't check all the piles. If it finds that one numeral has more flats than the other numeral, then it needn't compare the longs or units: even if the other numeral has the maximum number of longs and units allowed—namely, nine each—the first numeral will still represent the larger number. Imposing the canonicity constraint makes the comparison procedure much more efficient, because it allows the procedure to stop earlier. But the

canonicity constraint is a constraint on the representation of numbers, and so all arithmetic procedures must obey it. Even though the constraint makes part of the addition procedure somewhat less efficient, it makes comparison so much more efficient that it is worth having. This appeal to efficiency is the ultimate end point in the explanation of the *motivation* for carrying and the canonicity constraint.

In this miniexplanation of carrying, we have seen two important facets of teleologic knowledge. In the addition procedure, the canonicity constraint was *manifested* as a carry subprocedure. But the *motivation* for adopting the constraint lay in another procedure, comparison. Each of these two facets, which we now call *local* explanations because they explain just one constraint, was illustrated with a minimal contrasting pair of procedures. One member of the pair was a fully operational version of the procedure that lacked the constraint being discussed, whereas the other member adopted the constraint. But the manifestation part of the explanation involved a minimal contrasting pair that was different from the pair used to motivate the constraint (i.e., addition vs. comparison). As discussed later, it is preferable to have a pair of analogous procedures that illustrate both the manifestation and the motivation of teleologic concepts, but this is not always possible.

It is our belief that the concreteness of this minimal contrasting-pair paradigm of explanation is of crucial importance in making teleologic semantics clear. The learner can *see* in very concrete terms how adopting a constraint affects the procedure. Winston showed that a similar example-based paradigm was sufficient to teach the abstract concepts necessary to recognize toy block constructions, such as an arch (Winston, 1975, 1978).

In fact, many minimal contrasting pairs that manifest the given constraint are available, depending on which of the remaining constraints are adopted. If all the constraints of a given target procedure are adopted, then one member of the pair is the target procedure itself. Otherwise, the contrast is exhibited across a pair of *model* procedures that still satisfy the mathematical constraints of the target procedure. Using model procedures often highlights the contrast, making it much easier to see the constraint under discussion. Such was the case with the canonicity constraint, where Dienes Blocks allowed us to use noncanonical numbers without inventing new digit symbols.

However, model procedures must be used with some care, as the following example illustrates.

The Impact of Efficiency Metrics on "Loop Jamming."   Consider the difference between the standard carry subprocedure and the two-pass version described in the introduction, where carrying was deferred while all the columns were added, then performed on a second pass over the columns. This difference is a constraint that was called *loop jamming*, after the compiler optimization technique of the same name that weaves two loops into one (Allen & Cocke, 1972).

One cannot use Dienes Blocks procedures to motivate loop jamming, because exactly the same number of hand motions, fact-table lookups, and so on are required by each procedure. So, Dienes Blocks are an inappropriate model domain for discussing this constraint.

However, when implemented with written numerals, loop jamming does create a difference in efficiency.[8] The two-pass implementation of carrying requires more writing than the standard implementation. Thus written arithmetic turns out to be an appropriate domain for discussing the loop-jamming constraint.

The important point to notice about this example is that the choice of the model has some impact on the local explanation. In particular, a model that clearly displays the *manifestation* of the constraint in the procedure may not be able to demonstrate the *motivation* for the constraint. For example, because one doesn't have to worry about how to write the intermediate column sums that may be greater than 9 with Dienes Blocks, we can use them to implement both the one- and two-pass addition procedures and thus use them to illustrate the manifestation of loop jamming. Unfortunately, however, they cannot be used to motivate loop jamming, because the resulting procedure is no more efficient.

Another point to notice about the preceding example is the use of *efficiency metrics* in motivating design choices. An efficiency metric is some weighted sum of hand motions, fact-table lookups, table size, amount of paper used, and the like. The weighting of efficiency metrics is very important. For example, if reducing memory load is more desirable than decreasing the number of write operations, then the discussion of loop jamming ends with the opposite conclusion—that two-pass carrying is better than the standard subprocedure.[9] The two-pass version uses less short-term memory but more pencil lead. So exactly what efficiency metric is used greatly affects the local explanation. We do not look upon efficiency metrics as a regrettable new variable that must be tied down and parameterized with careful experimentation, but rather as a source of flexibility that can be used to tailor the teaching paradigm to the needs of particular students.

Principles for Sequencing Local Explanations

For moderately complex procedures, such as subtraction, the number of constraints can be high enough to cause problems of presentation. Our current best

---

[8] In the standard version of subtraction, where the carry loop is jammed together with the add-column loop, one must write $n + m$ digits, where n is the length of the longest addend and m is the number of carries required (it is assumed that one writes a 1 above the columns one carries into). In the two-pass version, one must write $n + 2m$ digits: One must remember from the first pass which columns are overflowing, and this requires m notes to oneself—say, in the form of writing a 1 above the overflowing column. The second m operations come from rewriting the answer digit of the columns that are carried into. There may be even more rewriting if the answer carried into is a 9.

[9] In the column carried into, the standard subprocedure requires adding three digits, one of which

estimate of the number of constraints of subtraction is 17. To explain this many constraints, each with its own manifestation and motivation, may seem a difficult task. However, with the planning net formalism, we can investigate how to sequence "optimally" a collection of "model" procedures; the first procedure (or "model") of the sequence would be a very, very simple version of the skill, and the last procedure of the sequence would be the target procedure. For example, in subtraction, the first procedure might be base-1 block subtraction and the last, standard written subtraction. But how should the intermediate models be sequenced?

Using the formalisms developed earlier, principles for sequencing local explanations can be stated precisely. Several such principles are stated next that we believe will lead to sequences that better enable assimilation of the overall teleology of a procedure from the explanations of its parts. Each one of them falls out quite naturally from the planning net formalism.

It is convenient in what follows to say that such sequences run from left to right—the target procedure is the procedure on the far right. This allows us to talk of the left and right procedures of a mp-morphism. Also, we speak of the left and right difference generators of an mp-morphism; if A is left of B, then $d(A - B)$ is the left difference generator.

Introduce Each Constraint.   As we saw in the previous subsection, it is best to illustrate each constraint with a minimal contrasting pair of analogous procedures. This is probably the most important sequencing principle, that each constraint be illustrated individually. However, it is probably also true that it is better to introduce the constraint rather than take it away. This gives the sequence an air of progression toward the target procedure. Putting this principle formally, we have: Each constraint is the sole contents of the right difference generator of some mp-morphism in the sequence. That is,

Principle 1. For each constraint C in the target procedure's planning net, there exists i such that $d(P_i - P_{i-1}) = \{C\}$.

where the procedures are numbered from left to right (first to last).

Starting with a very simple procedure would, hopefully, tap a person's intuitive understanding. Then, since each of the analogies (mp-morphisms) is very close (or at worst, moderate; we are guaranteed only that one of the difference generators is a singleton set—namely, the constraint being introduced), it should be easy to transfer that understanding along, *augmenting* it only *slightly* as each new procedure is presented.

---

is, of course, the carried 1. But adding three digits requires remembering the sum of the first two digits while assessing the third digit. The two-pass subprocedure doesn't load memory this way, because the intermediate sun is written down instead.

Only Introduce Target Procedure Constraints.   Occasionally, it is necessary to "build" a left procedure to illustrate some constraint. This occurs when one cannot adjust the sequence so that the right procedure of some other constraint is this constraint's left procedure. In this case, one ends up with an adjacent pair of procedures that do not illustrate a constraint from the target procedure. Although the person (or computer) doing the explaining can mention that this analogy isn't so important, it would be better if the sequence didn't have such pairs. So another optimization principle to shoot for is:

Principle 2. For each i in the sequence, there exists a constraint C in the target procedure's planning net, such that $d(P_i - P_{i-1}) = \{C\}$.

Minimize Redundancy.   One should not remove a constraint that has been introduced previously or introduce a constraint twice. Although one could argue that the redundancy of seeing the constraint illustrated in several different contexts (i.e., with different model procedures) serves to reinforce the local explanation, we are of the opinion that this would create confusion rather than dispel it, and in addition, it would create the impression that the sequence was meandering.

More formally, we propose that the sequence obey the following conditions:

Principle 3. For any $i \neq j\ d(P_i - P_{i-1}) \cap d(P_j - P_{j-1}) = \phi$

Principle 4. For any $i \neq j\ d(P_{i-1} - P_i) \cap d(P_{j-1} - P_j) = \phi$

Principle 5. For any $ij\ d(P_i - P_{i-1}) \cap d(P_{j-1} - P_j) = \phi$

The first condition advises one not to introduce a constraint twice, and the secor condition advises one to avoid removing a constraint twice. The third condition says that once a constraint is introduced (the first term), it can never be taken out (the second term). Actually, it also says that once a constraint is removed, it shouldn't be reinserted, which is also a plausible condition to impose for aiding the cogency of the sequence.

Efficiency Should Increase Monotonically.   We mentioned earlier that a minimal contrasting pair for a constraint does not necessarily show an increase in efficiency. That is, all ways of manifesting a constraint do not necessarily motivate it as well. One condition on a sequence is that the model procedures be chosen and sequenced so that efficiency always increases as the target constraints are adopted. That is,

Principle 6. For all i, P_i is more efficient than P_{i-1}.

Because there are many minimal contrasting pairs that manifest a constraint, it is usually not difficult to find some pair that motivates it as well, but putting that

pair into a sequence with the other constraint's pairs can be somewhat difficult. We know of only one constraint for addition or subtraction—namely the canonicity constraint, where the motivation pair *must* be distinct from the manifestation pair. This is inevitable because canonicity is basically designed to improve the efficiency of comparison, not the other arithmetic operations. Thus, if one were only interested in a sequence of addition procedures or subtraction procedures, then the pair for the canonicity constraint would necessary violate this sequence principle. However, with this one exception, it has been easy to fine *some* minimal contrasting pair that serves both to manifest and motivate a constraint for subtraction.

However, putting such pairs into a sequence requires some care. Switching the order of two constraints in a sequence often alters the relative efficiency of the minimal contrasting pair of procedures that manifest the unit. Under one ordering, both constraints might improve efficiency. But under the reverse order, adopting one of the units may result in no increase in efficiency or even a decrease in efficiency. This might seem strange, so let us pause a moment for an example.

Consider ordering the canonicity constraint versus the constraint that Dienes Blocks be kept sorted by size. First, suppose that the canonicity constraint precedes the sort-by-size constraint in the sequence. Under this ordering, the efficiency increases between each procedure; imposing the canonicity constraint forces the procedure to search through the big pile of Dienes Blocks to check that there are no more than 10 blocks of any given shape. Hence, adopting the sort-by-size constraint greatly improves efficiency by eliminating rummaging around through the big pile in favor of simply counting up the number of blocks in each of the small piles.

Now suppose the order in the sequence were reversed and sort-by-size were imposed before canonicity. The minimal contrasting pair for sort-by-size consists of: (1) adding two big piles of Dienes Blocks together by simply forming the union versus (2) adding each of the small piles together in a series of separate union operations. The introduction of the constraint actually decreases the efficiency of addition. Because no carrying is required (canonicity not being imposed yet), there is no use in the separation by size. Maintaining the constraint creates extra work with no reward. So modifying the order of two constraints in the sequence can have an impact on the ability to motivate them.

Although it may be a difficult condition to achieve, if a manifestation-based sequence has monotonically increasing efficiency, the viewer can see with no additional examples not only *what* each constraint is but also *why* it exists (i.e., what good it is).

Telescoping Sequences. Occasionally, one finds mp-morphisms that introduce a constraint but don't need to remove any constraints. The canonicity constraint can be illustrated with an mp-morphism whose left difference subgraph is null (for addition, one could use the two-pass addition procedure de-

scribed in the introduction as the right-hand procedure, and the first pass of it for the left procedure). That is, the mp-morphism is *total* with respect to the left planning net. It seems plausible that mp-morphisms that never removed constraints would create a very strong sense of progression toward a target procedure. Such sequences are characterized by the condition:

Principle 7. For any i, $d(P_{i-1} - P_i) = \phi$

A Space of mp-Morphisms

Needless to say, it will rarely be possible for a sequence to satisfy all the sequencing principles we have mentioned. Indeed, we may only be able to satisfy some principles along part of its length and different principles along another part. We need some way to study the relative contributions of the various principles to ease of explanation.

Ultimately, we would like to develop a representation of all principled sequences to a given target procedure. These sequences could be represented in an economical way by a directed graph whose nodes would represent planning nets. There would be a link from node A to node B only if they appeared as an adjacent pair in some sequence that was considered a plausible explanation sequence, perhaps because it met some minimum number of the principles listed earlier. (In particular, one might include all (known) minimal contrasting pairs for the target constraints; this would correspond to using principle number 1 as a threshold for inclusion in the space.) This directed graph has the property that any sequence from a "most primitive version" node to the "target" node would be a possible sequence for explaining the teleology of the target procedure. We tend to think of this graph as a *space* of mp-morphisms.

One clear problem that could be attacked with such a space is improving on the *naturalness* of teleologic explanations. Presenting the 17 or so mp-morphisms (or procedural models) for place-value subtraction is bound to be very confusing unless they can somehow be aligned along the individual's own cognitive structures (see the Appendix for a detailed example of one such chain of models). We have already mentioned seven principles that probably contribute to better comprehension of such explanations. Each of these principles would be incorporated into the space, perhaps as annotations on the basic partial order. Hopefully, experience and experiment will lead to the discovery of other factors that improve the naturalness of teleologic explanations.

Using the mp-Morphisms Space
in Microworld-Based Curricula

In a microworld-based curriculum, the student explores a rich environment, hopefully inventing something analogous to the target skills (Papert, 1978;

Fischer, Brown, & Burton, 1978). For example, a student might be given Dienes Blocks and a puzzle that requires using multidigit arithmetic to solve it. Actually, how students are motivated to do the arithmetic is not an issue here. The point is that students are not given the sequence of actions that implement arithmetic for the given representation of numbers. Instead, they must invent it themselves.

Tracking a Student's Discovery Process.   The mp-morphisms space could be quite useful as a way to "track" a student's discovery process. The basic idea is that an observer (possibly a computer) analyzes the procedures that the student invents in terms of planning nets. The nodes in the space that correspond to the plans of these procedures are marked. The student's progress is then expressed as the shortest sequence along the constraints that connect the marked nodes. This provides a strong hypothesis concerning what the student has learned during the discovery process.

Such a tracking study would provide an empirical way to verify conjectures about "natural" sequences for teleologic explanations. That is, observing that students generally followed sequences that increase the efficiency of the procedure would support the conjecture that monotonically increasing efficiency is important for cogent, natural explanations.

Sequencing Microworlds.   A persistent problem in microworld-based curricula is how to sequence the microworlds so as to maximize the cumulation of intuitions built up while exploring the microworld and enable them to be transferred to the target procedure. One ready answer is provided by the space of mp-morphism sequences, assuming it has been annotated to show which sequences are most natural.

Sequencing microworlds obviously imposes an order on the traversal of the nodes in the mp-morphism space. One can't move from a Dienes Block procedure to an abacus procedure's node until one leaves the Dienes Block microworld and enters the abacus microworld. So the most natural sequence of microworlds is the one that enables traversal of the most natural sequences through the constraint space. Let us illustrate this conjecture with an example.

Suppose one tried to teach addition with the following sequence of microworlds:

base-1 blocks, the abacus, Dienes Blocks, written numbers

One would expect the students to become frustrated when they find that the teleology associated with place-value encoding of numbers, which they laboriously invented for the abacus, is obviated by the shape-value encoding of Dienes Blocks. And when they find they must resurrect this place-value notion to move from Dienes Blocks to written numbers, one would expect them to become disgruntled or, worse yet, to apply "teacher psychology" and guess that place

value couldn't possibly be part of the design because "we already had that." In comparison, reordering the sequence to be

base-1 blocks, Dienes Blocks, the abacus, written numbers

allows invention of the notion of place-value just once, in transition froi i Dienes Blocks to the abacus, and then maintenance of the notion throughout the abacus microworld and on into the written numbers.

These ordering results could be predicted on the basis of one of the naturalness principles mentioned earlier—namely, that constraints ought to *accumulate* along the sequence. They should be added once and never removed. In the first sequence of microworlds, there is no sequence of procedures that can avoid adding the constraints that express place-value encoding during the first transition and dropping some of them during the second transition.

What Is the Closest Possible Procedure in a Given Microworld to the Target Procedure? Just exactly how close to standard arithmetic procedures can procedures built around a particular representation of numbers, say Dienes Blocks, be made to be? Can a Dienes Block procedure be devised that is totally isomorphic to a standard written procedure? This is a question of interest to educators. For example, it bears on the question of just how much a child can learn about standard arithmetic by inventing a good arithmetic procedure in a given microworld, such as Dienes Blocks. This in turn bears on the question of how many microworlds, and which ones, are necessary to allow the student to easily converge upon the target skill. With a formal theory of analogy between procedures, we can now precisely determine how close the best possible procedure defined over a given microworld can be to the target procedure:

Take any procedure that uses the given representation of numbers. Examine the difference generator of the analogy between it and the target procedure (e.g., written addition). If this set contains constraints that cannot be met because of the basic physics of the representation, then one cannot construct a model procedure that is isomorphic to the target procedure. An example should make this a little clearer.

A careful examination of the planning net has shown that it is impossible to construct a Dienes Block addition procedure whose analogy with written addition is perfect (i.e., an isomorphism). One design issue that is always present in Dienes Blocks involves the shape-value encoding that is the hallmark of Dienes Blocks. There is an encoding of the relationship between position and place value that is present in both written addition and sorted Dienes Block addition, but it is redundantly coded by the visual appearance of Dienes Blocks. If one got rid of this redundancy by evening out the sizes of the blocks, then they wouldn't be Dienes Blocks anymore. So the redundancy is inherent in the representation and will be part of the difference generator of the analogy to written addition no matter how clever one is about inventing Dienes Block addition procedures.

As a consequence, certain subtle *shifts in representation* that occur in the standard procedure for adding written numbers cannot be duplicated in any Dienes Block addition procedure.[10] This deficit gives some bite to the inherent incompleteness; the subtlety of these shifts makes them likely candidates for misunderstandings that Dienes Blocks are apparently helpless to prevent. This essential inadequacy can be directly diagnosed, if not predicted, using the theory of analogy between procedures.

In similar fashion, other microworlds can be evaluated. This evaluation is, however, quite constructive. Once the inherent mismatch with the target procedure has been identified, the gap can be filled by modifying the microworld, or by adding another microworld to the curriculum if desired.

In short, many of the same issues appear to be involved in the teaching teleology and discovery-based teaching. Planning nets seem to provide a formal tool for investigating this relationship further.

CONCLUSIONS

The major claim of this chapter is that planning nets provide useful formalisms for capturing the teleologic semantics of procedures. However, probably the most importnat thought to take away from this exposition is the importance and utility of using *planning knowledge* in the deep-structure analysis of procedures.

In contrast to other work on analogy, we have ignored the process of solving an analogy problem. Instead, we have concentrated on an intuitive determination of what representation most closely models the way experts conceive of procedures in order to understand analogies. This methodology has arrived at the same conclusion that was reached by a completely different method. In particular, our planning nets are very similar to Sacerdoti's "procedural nets" (Sacerdoti, 1977). Sacerdoti has shown his procedural nets to be a *sufficient* representation for designing procedures and indeed much better than other known representations. We have tried to show a similar representation to be a *sufficient* representation for judging the closeness of analogy and indeed much better than other known representations. In short, evidence is accumulating that planning net-like representations are good for many purposes. However, we should point out once again that neither Sacerdoti nor ourselves make any claims that the process of building a planning net, either for analogy or design, exactly models the human *process* of building a planning net.

---

[10]When one adds two large digits from a given column, one gets back a nondigit—for example, 14. The first shift in representation is to break this number down into units and 10s. Next, the units must be converted into a digit in the columns being added, whereas the 10s must be converted into an argument to the carry subprocedure. In Dienes Block addition, the second conversion is superfluous, because the result of the column addition is already scaled up to the value of the column, so to speak. That is, an add in the 10s column yields 140 in the form of 14 LONGS, not 14 UNITS.

Because teleologic knowledge is a part of a certain kind of expertise, one naturally wonders how it can be taught. Planning nets provide a precise framework for constructing explanations and curricula to explicate teleology. In particular, the formalism helps answer the question of how to sequence a set of "model" procedures with certain formal properties. Moreover, many of these same formal properties seem useful in discovery learning curricula.

Our last comment should undoubtedly be that this research is just beginning. There are many deficiencies and questions that must be addressed. Reliable empirical measurements of closeness and transferability must be made. The uncertainties in the uniqueness issue must be investigated. The general precision of the theory must be improved, and its inordinate amount of detail must be tamed, hopefully with the aid of a computer. In particular, we would like a complete, precise, mp-morphisms space for all five arithmetic operations. The limitations of the theory should be tested by exercising it on examples from other domains. In other words, this chapter is more a proposal to investigate a promising line of thought than a report on completed research.

APPENDIX:
AN EXPLANATION OF THE TELEOLOGIC
SEMANTICS OF SUBTRACTION

To give a feeling for how an explanation based on paths of minimal contrasting pairs of analogous procedures might go, an example of such a path is presented here. It begins with a base-1 subtraction model, passes through some Dienes Block subtraction procedures, and ends with the standard procedure for subtraction of written numerals. Although reading these rather abbreviated descriptions can have nothing like the impact of actually handling the blocks and doing the procedures, the power of this technique to explain teleologic semantics should nonetheless be apparent.

Throughout the path, there is a certain ambivalence about the particular material that is used in the representation of number. In fact, the primitives and constraints used to describe and implement procedures really can't differentiate real, wooden Dienes Blocks from, say, drawings of Dienes Blocks, as long as they are manipulated the same way. In fact, there is no particular point where adoption of the constraints of the target procedure (written subtraction) forces us off the counting table and onto paper; one can actually implement standard subtraction with cards bearing digits.

However, the material *does* make a difference to the efficiency metrics. In particular, some of the later constraints can only be motivated by assuming that erasing is more work than writing, which is true of paper but hard to emulate with manipulable materials.

We start with base-1 blocks because the mathematical semantics of this subtraction procedure are simple and concrete.

1 *Polynomial.* Base-1 numerals are rather bulky for representing large numbers. One solution to the block management problem is to let some counters stand for a fixed number of the unit counters. This is the *polynomial* constraint (3' in the text). The next procedure of this mp-morphism is a simple version of big-pile Dienes Block subtraction.

2 *Search Instead of Random Choice.* This model adds the notion that searching for two blocks of the same shape is more efficient than picking two blocks at random, then trading to make them the same shape.

3 *Choose Larger to Trade Down.* The idea here is to trade down the larger of the two blocks. If one picks an arbitrary block to trade down but not the unit block, then eventually one will be able to match their shapes, but it will often take more trading than always picking the larger one to trade down. This procedure requires memorization of which of two shapes stands for a larger multiplier.

4 *Search for Next Larger Before Trading.* When one can't find two blocks of equal shape, and instead has two blocks of unequal shape, then before trading down the larger one, replace it with a block that is the *next size larger* than the smaller block. If the search succeeds, one only has to trade down once. This plan step requires memorizing which shape is the next larger one than a given shape.

5 *Choose TOP to Trade Down.* This model is motivated by observing that when the block that is traded down comes from BOT (the bottom numeral), the subtraction as a whole takes more time than it would if the block had come from TOP (the numeral that is being subtracted from). When a block from BOT is traded down, the nine smaller blocks that are left over go back into BOT. So the main loop must run nine times more. If a block comes from TOP, the nine extras go back into TOP. If BOT runs out soon, they may never be touched. So trading down a block from TOP is more efficient than trading down a block from BOT.

The goal of choosing TOP blocks creates a subgoal that the TOP block be larger than the BOT block. This subgoal is satisfied by a subgraph that is already a part of the left planning net—namely, the union of the subgraphs generated by models 2, 3, and 4. So the new part of the planning net underlying this procedure is just the part that satisfies the goal "choose BOT block" exclusive of the part that satisfies the subgoal.

6 *Canonicity.* This constraint was described earlier.

7 *Base Ten.* The canonicity constraint produces a trading pattern that is much easier to remember if all the multipliers are powers of 10 (or some other base). For example, in canonical American money, which is a
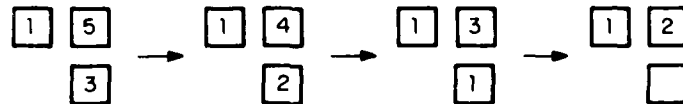
polynomial representation but not a base-10 representation of number, a citizen would canonicalize their pocket change by trading in five pennies for a nickel, two nickels for a dime, three dimes for a quarter and a nickel, and so forth.

8. *Sort by Power.* Canonicalization (= carrying) and decanonicalization (= borrowing) are somewhat easier if numerals are sorted so that all counters of a certain power are accessible at once. Dienes Blocks, as we observed them being used in schools, lacked this constraint. In fact, Dienes Blocks lack the canonical and base-10 constraints as well. However, teachers usually require their students to obey these two.

9. *Power Represented by Location Only.* Numerals must take up space, either on table tops or on paper. Once powers are sorted, location in space redundantly represents the power of a counter. In this mp-morphism, that redundancy is removed by making all coefficient tokens (i.e., "digits") look the same, regardless of the power. The abacus, for example, obeys this constraint. This allows one to represent much larger numbers, since one need not invent new token shapes when one needs to use a new, higher power. That is, one can make an abacus of arbitrary width, but Dienes Blocks, which are inherently unable to obey this constraint, are limited in practice to, at most, four powers.

10. *Zero.* To use location to represent power, a prearranged pattern of locations must be used. But such fixed patterns, like the abacus or columnar ruled paper, can't represent numbers that are larger than they have been designed to represent. Moreover, producing the patterns accurately is difficult to do freehand. A good solution to this problem is to represent power with relative locations, which amounts to using *zero* as a placeholder. A "relative-location abacus" could be built that lays out piles of beads in a line on the table; it would use a clear plastic bead as a placeholder and piles of colored beads as nonzero "digits."

11. *Alignment.* In setting up the subtraction problem, one insists that the numerals be aligned so that digits of the same power are in the same column. This reduces the effort necessary to locate the digits of matching power when subtracting.

12. *Noncountable Coefficients.* It is quicker to arrange counters on a table or write coefficients symbols on paper if the number of counters or strokes is small. This motivates replacing countable coefficients with symbolic ones (e.g., digits). However, with symbolic coefficients, the PICK/FROM operation is radically altered. It is no longer possible to decrement a coefficient by picking up a piece of it (i.e., picking up a block or erasing a hash mark). Instead, a decrementation table must be memorized. That is, one must be able to count backwards from 20.

There is no particular point where the target constraints force us off the counting table and onto paper. Manipulatory systems can be devised that

use noncountable coefficients. One such manipulatory system is simply a set of cards bearing digits, which are laid out in a line on the table.

13. *Memorize Pairing Off.*   The next few minimal contrasting models are designed to minimize the manipulation of the cards in a manipulatory system, or erasing a digit and writing a new one in a written system. In the previous number systems, column subtraction was realized by pairing off decrements of the top and bottom digits. A "movie" of the card procedure doing 15 − 3 would be

$$\boxed{1}\;\boxed{5}\;\xrightarrow{}\;\boxed{1}\;\boxed{4}\;\xrightarrow{}\;\boxed{1}\;\boxed{3}\;\xrightarrow{}\;\boxed{1}\;\boxed{2}$$
$$\boxed{3}\qquad\quad\boxed{2}\qquad\quad\boxed{1}\qquad\quad\square$$

This model replaces this pairing-off loop with a table loopup. A "movie" of the modified card procedure doing 25 − 7 is

$$\boxed{2}\;\boxed{5}\;\xrightarrow{}\;\boxed{2}\;\square\;\xrightarrow{}\;\boxed{1}\;\boxed{10}\;\xrightarrow{}\;\boxed{1}\;\boxed{8}$$
$$\boxed{7}\qquad\quad\boxed{2}\qquad\quad\boxed{2}\qquad\quad\square$$

14. *Memorize Comparison.*   This model procedure replaces the two-step borrowing (see foregoing movie) with a one-step borrow by looking ahead. That is, it looks ahead to see which digit will be zero—the top or the bottom. This amounts to memorizing the greater-than table for digits. Now the movie for 25 − 7 is

$$\boxed{2}\;\boxed{5}\;\xrightarrow{}\;\boxed{1}\;\boxed{10}\;\xrightarrow{}\;\boxed{1}\;\boxed{8}$$
$$\boxed{7}\qquad\qquad\boxed{2}\qquad\qquad\square$$
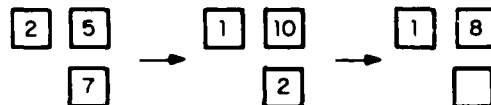
15. *Memorize Teens Facts.*   Two table lookups can be reduced to one, and two digit rewrites can be saved if a new facts table is provided for the teens facts. The new table is 10 by 9 and contains facts like 15 − 7 = 8. The movie reduces to

$$\boxed{2}\;\boxed{5}\;\xrightarrow{}\;\boxed{1}\;\boxed{8}$$
$$\boxed{7}\qquad\quad\square$$
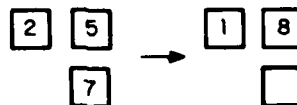
16. *Sequence Columns.*   In the previous systems, columns are processed in random order. However, this necessitates marking the columns that are done by zeroing the bottom digit. This digit rewrite can be saved if the columns are processed in some set order—either left to right or vice

versa. The planning heuristic—that is, the right difference generator of this mp-morphism—could be called "ordering independent operations reduces marking."

17. *Answer Register.*   If a separate place is provided for writing the answer, then erasures of the top digits can be reduced. This is motivated by the fact that writing a digit is easier than erasing—a property peculiar to paper.

18. *Right to Left.*   If the columns are processed right to left, one borrows from the top digit. If the columns are processed left to right, one borrows from the answer. The numeral that gets borrowed from ends up with erasures, whereas the other one has no erasures. If one erases by scratching out the digit and writing the new digit above, then the numeral that's borrowed from can become a real mess. The motivation for this analogy is that there is more need for the answer numeral to be legible than the top numeral. Hence, subtraction is more efficient if one processes the columns from right to left.

At last, we have arrived at the standard subtraction algorithm via a sequence of procedures/models where each model in this sequence has an mp-morphism between it and its immediate successor, thus creating a well-structured sequence of analogous models converging to the desired target procedure.

REFERENCES

Allen, F., & Cocke, J. A catalog of optimizating transformations. In R. Rustin (Ed.), *Design and optimization of compilers.* Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Brown, J. S., & Burton, R. R. Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science,* 1978, *2,* 155-192.

Card, S. K. *Studies in the psychology of computer text editing systems* (Xerox PARC Rep. No. SSL-78-1). Palo Alto, Calif.: Xerox Palo Alto Research Center, 1978.

Evans, T. G. A program for the solution of geometric-analogy intelligence test questions. In M. Minsky (Ed.), *Semantic information processing.* Cambridge, Mass.: MIT Press, 1968.

Fischer, G., Brown, J. S., & Burton, R. R. *Aspects of a theory of simplification, debugging, and coaching* (BBN Rep. No. 3912). Cambridge, Mass.: Bolt Beranek and Newman Inc., 1978.

Greeno, J., Gelman, G. R., & Riley, M. S. *Young children's counting and understanding of principles.* Paper presented at the meeting of the Psychonomic Society, San Antonio, Texas, November 1978.

Newell, A., & Simon, H. A. *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Papert, S. A. *Computer-based microworlds as incubators for powerful ideas* (MIT Logo Lab. Working Paper). Cambridge, Mass.: Massachusetts Institute for Technology, 1978.

Sacerdoti, E. *A structure for plans and behavior.* New York: Elsevier North-Holland, 1977.

Sternberg, R. J. Componential processes in analogical reasoning. *Psychological Review,* 1977, *84,* 353-379.

Tversky, A. Features of similarity. *Psychological Review,* 1977, *84,* 327-353.

Winston, P. H. Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Winston, P. H. Learning by creating and justifying transfer frames. *Artificial Intelligence*, 1978, *10*, 147–173.

19

A Theory-Based Approach to the Study of Individual Differences in Mental Imagery

Stephen M. Kosslyn

Pierre Jolicoeur
Harvard University

The scientific study of individual differences in imagery ability can be traced to the very beginnings of differential psychology. Thus, it may seem somewhat surprising that more progress in the study of individual differences in imagery has not been forthcoming. In fact, there is a striking similarity between contemporary research on the topic (see, for example, Richardson, 1977; White, Sheehan, & Ashton, 1977) and research performed over 75 years ago (see Angell, 1910; Woodworth, 1938). The standard modus operandi has been to devise a test that purports to measure some imagery ability, and then to look for correlations between the test scores and behavior in some domain. This is well and good, in theory. The problem is that it has not worked out very well in practice; we simply have not done a very good job of constructing tests that predict very much. There are at least two reasons for this failure: First, the test items usually are selected solely on the basis of intuition. The experimenter has some hypothesized dimension in mind and selects items that seem to tap this dimension. There is no good evidence even that the construct at hand has validity, let alone that the items measure what they are thought to measure. Second, even if we had good tests, it would be difficult to know this for sure. Only with a theory of imagery can one know what will be the behavioral consequences of individual differences in some aspect of mental imagery (e.g., vividness, mobility, frequency, and so forth). That is, without knowing how images are represented and processed internally, we cannot know what the ramifications of individual differences will be. To illustrate this point, we first critically consider a brief overview of the kinds of assessment procedures that have been used to study individual differences in imagery. Following this, we outline the present theoretical framework, which is discussed in detail in Kosslyn and Shwartz

139

(1977, 1978)   and in Kosslyn (1980). Next, we discuss one particular topic within the context of this model—namely, the question of when imagery is used in answering questions. We present some data arguing that images and "propositional" representations are retrieved simultaneously, and then we consider the loci in our theory where variations among individuals may occur. Having come this far, we are then in a position to discuss how variations in each process, and combinations thereof, could underlie differences in how much one uses imagery in thinking. Finally, we describe a new test we have begun to develop and some preliminary findings obtained using it. This test is intended to measure imagery use, and we hope to develop versions of it that will allow us to measure not only how much a person spontaneously uses imagery but also the underlying causes of this propensity.

INDIVIDUAL-DIFFERENCES RESEARCH

Self-Report Techniques

People probably have always talked about their m. ids and experiences. Thus, it is no surprise that self-report techniques are the oldest method of studying imagery. Because different people said different things, tnis technique became closely intertwined with the study of individual differences. T. e observation that people differed in their abilities to call up sensory experiences from memory was made as early as 1860 by Fechner and later, in more detail with more data, by Galton (1883). Fechner asked people to evoke an image of some named object and discovered that some people were only able to get momentary glimpses, whereas others claimed to experience more detailed, perceptlike experiences. Galton's study made use of the now-famous "breakfast table" questionnaire. This technique involved asking one to image one's breakfast table as it had looked that morning. Subjects were then asked a number of things about the image, such as the brightness (relative to the actual scene), color, amount of detail. Interestingly, slightly over 10% of his subjects claimed not to have any images—and in fact doubted their very existence. Because the nonimagers tended to be successful scholars and scientists, and the women and children usually were imagers, Galton suggested that imagery was the characteristic mode of thought for women, children, and "weak-minded" individuals [!].

The tradition of examining individual differences in imagery was carried on by others, some of whom studied the predominance of different types of imagery (e.g., Stricker, 1880) or studied imagery in famous individuals (e.g., Toulouse, 1897). In addition, Galton's work was refined by Betts (1909), who developed a questionnaire that not only required people to assign a numerical value for vividness of their images but that tested multiple sensory modalities (e.g., auditory and olfactory, in addition to visual) as well. In Betts' initial studies, auditory

and visual imagery were more vivid than the others but only slightly more so. It is interesting to note that in more recent work McKellar (1965) found that virtually all members of Mensa (who score highly on IQ tests) reported some imagery, in contrast to Galton's results, but that these people reported vast differences in the frequency and quality of nonvisual and nonauditory imagery, suggesting that (1) the methodology is flawed, or (2) times change. We cannot assess the validity of the second conjecture, but the validity of the first is clear: Not only does rated vividness often vary depending on incidental details (e.g., the identity of the experimenter; see Sheehan & Neisser, 1969), but it often fails to predict anything once other factors have been partialed out (e.g., see Kosslyn & Alper, 1977). As far as we can tell, only Marks (e.g., 1973, 1977) is having any iuck with self-report measures of image vividness. His VVIQ test (an expanded and improved version of the visual scales of Betts' test) seems to predict who will evince certain kinds of eye movements while remembering parts of pictures: More vivid imagers show *fewer* eye movements than do less vivid imagers (presumably to reduce competing visual input).

The problem with self-rating techniques is clear· There is no way to be sure (1) that everyone knows the referent of the word *image*, or (2) that everyone sets his or her criterion to the same level in assessing images. Further, as Sheehan and Neisser have demonstrated, this technique seems especially susceptible to demand characteristics, response biases, and the like.

"Objective Tests"

Woodworth (1938) describes a number of tests developed in the first decade of the 20th century (see also Angell, 1910; and Fernald, 1912). Because Woodworth's book is not readily available, it seems worthwhile not only to describe the techniques but also to note Woodworth's analyses of them in addition to our own.

Association Method.   A person is given 5 minutes to recall objects having some characteristic colors and 5 minutes to remember things having given sounds. People are judged "visualists" if color is a better cue, and "audiles" if sound is a better cue. Woodworth objects to this method because one "may recall a violin as being a sounding object without any image of the sound [p. 41]." Another obvious objection is that one's experience with different sorts of objects will determine the association strengths—regardless of how one actually uses given modalities in thought.

Word-Type Frequency in Prose.   In this case, the relative frequency of sight-words, sound-words, and so on is tabulated in a person's prose. The relative frequencies again are used in diagnosis. The objections already raised also apply here. Further, Woodworth (1938) reports: "Instances are on record in

which an author remarkable for his vivid descriptions of scenes reports himself *not* a visualist. There is nothing to prevent the non-visualist from seeing what is worth seeing and remembering it so as later to incorporate it in his writing [p. 41]." Yes, indeed.

Learning by Eye or Ear.   These studies examined optimal presentation modality in list learning. People were categorized according to relative memory for visually presented versus auditorily presented lists. Woodworth points out that there is nothing to prevent one from translating modalities internally (most everyone can name a written word!). Thus, presentation modality is not an index of how the material was actually represented internally. Further, one can imagine cases where an "audile" becomes deaf—but this does not disrupt the individual's thinking style. Finally, Woodworth reports that adults, regardless of whether they are visualizers or audiles, learn visually presented words more quickly (but we suspect this depends on presentation rates and other such factors). Woodworth's criticisms of this technique are very important and have yet to be grasped by numerous contemporary researchers: Simply showing someone a picture does not guarantee that he or she remembers it via imagery. Glanzer and Clark (1964) even argue that one remembers visual material solely by describing it (unfortunately, however, more difficult-to-describe pictures are also probably more complicated, so the fact that intricacy of description predicts recall is not surprising for any number of reasons).

Method of Distraction.   The logic of this method rests on the idea that internal processing will be disrupted most by having to process stimuli presented in the same modality. Thus, people were asked to learn lists of words while being subjected to auditory, visual, or kinesthetic distractions. As Woodworth points out, however, the perceptual tasks may simply have been more or less effortful—and hence more or less distracting—independent of mode of internal representation. The fact that different modalities are differentially distracting for different people could simply reflect relative practice or familiarity with stimuli of that sort.

Spelling.   This task has since been named the Hebb test (see Hebb, 1968; Weber & Harnish, 1974). Basically, words are read to a subject, and he or she is asked to spell them backwards. The logic is that if one has a photographic image, one can simply read the words off in reverse. Fernald (1912) found that nobody could do this. Interestingly, however, subjects in these experiments reported that letters (or syllables) seemed to fade in and out—undercutting the assumption that images are like static photographs. In fact, by 1912 Koffka had already published descriptions of subjects' imagery that seemed to show that one could image a coin of no particular denomination, an animal of no given species, and so forth. That is, his subjects reported decidedly nonphotographic images that were vague, contained indeterminate parts, and were missing details (more on this later).

Woodworth (1938) reports: "On the whole, those reporting visual imagery do somewhat better than other subjects in this form of test [p. 42]." But consider possible counterinterpretations of results from a variant of this task, the "letter square."

The Letter Square. In this technique, one is read a series of letters or numbers and is asked to arrange them mentally into a matrix of n rows by k columns. The subject is then asked to read off his or her image, naming the items in the columns, along the diagonals, and the like. The logic was the same as that underlying the spelling task: A photographic image should be able to be read in any direction equally easily. Thus, those able to read off arbitrary portions were to be classified as visualists. Woodworth found two flaws in this logic: (1) Even the most visual of the subjects reported being unable to maintain a rigid, static, photographlike image. Muller (1917) found that his "most competent visual learner" required four times longer to read the columns from top to bottom than to read the rows left to right, and more than seven times longer to read obliques than rows. (2) Perhaps more critically, the assumption that the auditory learner would be inordinately affected by the encoding order, and virtually unable to retrieve in different orders, was faulty. People can impose groupings and other structures on the input, allowing them to repeat the items in other orders later. Presumably, skill in such rearranging will be related to how much one "uses" a given modality, which also may be correlated with reports of imagery. Thus, differences in imagery could be an incidental concomitant of increased practice in organizing visual material.

Description and Memory. Introspective reports of imagery were obtained when a subject described a picture. People who reported more visualization gave more complex descriptions, whereas verbalizers were less likely to elaborate the picture falsely (and hence were more accurate). Davis (1932) repeated some of Fernald's early work and found that those reporting auditory images recalled tones more accurately than those who did not, whereas those reporting visual imagery recalled nonsense forms better than those not reporting visual imagery. The problem here is attention. The preferred mode per se may have nothing to do with performance; perhaps people reporting auditory imagery "listen better," and those reporting visual imagery "look better"; hence, they encode more initially.

In addition to the kind of work described above, there also is a long tradition of attempting to distill imagery or spatial factors that underlie performance across a variety of tasks. Although it has long been claimed in the psychometric literature that visual and spatial abilities are distinct from verbal abilities (see Smith, 1964; Spearman, 1927; Thurstone, 1938), this inference is based on somewhat subjective interpretations of correlational and factor-analytic studies. Typically, performance on a set of tests and/or tasks are correlated and these correlations themselves are analyzed for underlying patterns. The experimenter must interpret

the "meaning" of a dimension or factor by abstracting what seem to be the common elements shared by the tasks that load highly on that dimension or factor (and intuiting what seems missing from those tasks that do not load on the dimension or factor). This approach seems to compound the problems in interpreting the results of a given individual test or task. Now one must worry about multiple interpretations (including some having nothing to do with imagery, as illustrated above) of a dimension or factor, based on multiple interpretations of the results of particular tasks. This approach would be useful if one had some prior reason(s) for believing that given tasks do in fact require imagery, but this is not the case. Given a lack of a priori justification for this assumption, it is difficult to draw inferences about individual differences in imagery per se from the factor-analytic studies.

In summary, then, none of the early work is compellingly "face valid." No reference to functional individual differences in imagery per se is necessary to explain the results. There is always at least one, and often several, equally plausible counter-explanations.

A MODEL OF IMAGERY

Clearly, it is desirable first to have a theory of imagery that will direct one to the important variables. If images are not like photographs, for example, much of the motivation for the earlier work is lacking; if the experience of images is entirely epiphenomenal—worse yet!—then we have no reason at all to expect any quality of the experience of "having an image" to correlate with the functional utility of imagery. That is, there are those who maintain that the images we experience are like the flashing lights on the outside of a computer while it is adding: The lights are merely an incidental concomitant of the genuine functional process. The present section has two main parts: First we provide the major empirical motivation for our model (see Kosslyn & Pomerantz, 1977, for logical, nonempirical arguments against the image-as-epiphenomenon view). This section merely sketches out the kinds of data we considered; detailed reports of the experiments can be found in the papers cited, and a more detailed overview can be found in Kosslyn (1980). Second, we outline the model proper. We present only the aspects of the model that are essential for motivating our work on individual differences in imagery use; other details can be found in the cited sources. Further, the metatheoretical foundation of the research strategy itself is not discussed here but is elaborated in Kosslyn (1980).

Empirical Foundations of the Model.

There are two main kinds of questions that had to be resolved (to some degree of certainty) before we felt comfortable in beginning to construct a model of imag-

ery processing. First we wanted evidence that supported the notion that the experienced images were not merely epiphenomenal. That is, we wanted to collect data that would be difficult to explain if quasi-pictorial, spatial images could not take part in actual information processing. Second, provided that we amassed enough data to motivate modeling a data structure that embodies characteristics of images as we experience them in "active" (short-term) memory, we then wanted to know something about how these images are stored in, and evoked from, long-term memory.

The Ontological Status of the Quasi-Pictorial Image

Four different classes of findings converge in supporting the view that experienced mental images can in fact take part in human information processing.

Experiments on Scanning Visual Images.   Kosslyn, Ball, and Reiser (1978) report a number of experiments that demonstrate that more time is required to scan greater distances across mental images. In one study, people imaged a map containing seven locations and scanned between all possible pairs. Time to scan increased linearly with increasing distance between the 21 possible pairs of locations, each of which was separated by a unique distance. There were no effects of distance in a control condition, where subjects focused on a location in the image but then simply decided whether another location was present without being asked to scan to that location.

In another experiment, people imaged schematic faces on which the eyes were either light or dark and located either 3, 4, or 5 inches above the mouth; in all other respects, the faces were identical. After a given face had been removed, a subject was asked to focus on the mouth and then to image the face as large as possible without it seeming to overflow; or to image it half this size; or to image it so large subjectively that only the mouth was left visible in the image. Following this, the word *light* or *dark* was presented. As soon as either word had occurred, the subject was to "glance up" to the eyes of the imaged face and see whether or not they were appropriately described by the word. Time to judge whether the eyes were light or dark increased linearly with distance from the mouth. Further, overall scanning times were reduced when people were asked to "shrink" an imaged face mentally prior to scanning it, and times were increased when subjects "expanded" a face before scanning. These results are difficult to explain if images are simple "abstract propositional" list structures, but they follow naturally if images are spatial representations that preserve metric distance information.

Measuring the Visual Angle of the Mind's Eye.   The notion that images represent spatial extent suggests that they have spatial boundaries; after all, they do not extend on indefinitely. If images occur in a spatial representational

medium, then their maximal spatial extent may be constrained by the extent of the medium itself. Kosslyn (1978c) used the following paradigm in an attempt to test this idea: People were asked to image an object as if it were being seen from very far away. Then they were asked to imagine walking toward the object and were asked if it appeared to loom larger; all subjects (of those who could do the task at all, which was usually only about 80% of the people tested) reported that it did. Further, these subjects claimed that the image loomed so large at one point that it seemed to "overflow." At this point, the subjects were to "stop" in their mental walk and estimate how far away the object would be if they were actually seeing it at that subjective size. This basic experiment was conducted in a variety of ways, having subjects image various sorts of pictures or animals when given just their names and sizes. In addition, subjects estimated distance by verbally assessing feet and inches or responded by moving a tripod apparatus the appropriate distance from a blank wall.

If images occur in a spatially constrained medium, then the larger the imaged object, the farther away it should seem at the point of overflow. In addition, a constant angle should be subtended by the imaged objects (which ranged in actual size) at the point of overflow. Using simple trigonometry, the "visual angle of the mind's eye" was computed from the estimated distances and longest axis of each imaged object. In all of these experiments, the basic results were the same: First, people claimed that smaller objects seemed to overflow at nearer apparent distances than did larger objects (the correlation between object size and distance was always very high), and distance usually increased linearly with size of the imaged object. Second, the calculated "visual angle" at the point of overflow remained constant for different-sized objects when subjects imaged pictures or objects that had just been presented. The actual size of the angle varied, however, depending on instructions: More stringent definitions of "overflow" resulted in smaller angles. These last findings imply that images do not overflow at a distinct point but seem to fade off gradually toward the periphery. (The best estimate of the maximal angle subtended by an image while still remaining entirely visible seemed to be around 20 degrees.)

In another experiment, people were asked to scan images of lines subtending different amounts of visual arc, and the amount of time required to scan each degree was then calculated. These people also scanned an image of a line they had constructed to be as long as possible without either end overflowing. The visual arc subtended by this "longest possible nonoverflowing line" was inferred from the time required to scan across it. This estimate was very close to one obtained using the technique already described and to one obtained by simply asking people to indicate the subjective size of the possible longest nonoverflowing line by holding their hands apart so as to span the length of that line.

Effects of Subjective Size on Ease of "Seeing" Parts of Mental Images. If asked which is higher off the ground, a horse's knees or the tip of its tail, many people claim to image the beast and to "inspect" the image, evaluating the

queried relation. It is possible that the "inspection" of images makes use of some of the same classificatory procedures used in categorizing perceptual representations. If so, then we might expect constraints that affect ease of classifying parts perceptually also to affect ease of imagery classification. Parts of smaller objects are "harder to see" in perception, for example, and also may be harder to "see" in imagery. This result was, in fact, obtained (see Kosslyn, 1975); parts of subjectively smaller images of objects did require more time to classify mentally than did parts of subjectively larger objects. In addition, simply varying the size of the part per se also affected time to examine an image. In this case, smaller parts—like a cat's claws—required more time to see on an image than did larger parts—like its head. This last result was obtained (Kosslyn, 1976a) even though the smaller parts were more strongly associated with the animal in question, and were more quickly verified as belonging to the animal when imagery was not used (more highly associated properties are typically affirmed as appropriate more quickly than less associated ones in studies of "semantic memory"; see Smith, Shoben, & Rips, 1974). These findings, then, not only are consistent with the notion that images are functional spatial representations, but also serve to distinguish between processing imaginal and nonimaginal representations.

Effects of Subjective Size on Later Memory.   If parts of subjectively smaller images are less distinct, then one might expect that the imaged object itself would be more difficult to identify. Thus, if one actually encodes a subjectively small image into memory, one's ability to recall the object later should be poorer than if the image had been larger—if in fact the image itself is recalled and inspected when one tries to recall the encoded words or objects. Kosslyn and Alper (1977) asked subjects to construct images of the objects named by pairs of words. Sometimes one of the images was to be very small subjectively, and sometimes both images were to be "normal" sizes. When a surprise memory test for the words was later administered, memory was in fact worse if one member of a pair initially had been imaged at a subjectively small size. This result was replicated in several studies, each of which controlled for different possible confoundings (e.g., less "depth of processing" may have occurred when people constructed subjectively smaller images).

These four classes of results converge in supporting the claim that characteristics of the quasi-pictorial images people report experiencing can influence information processing. The most elegant accounts for the described results seem to include the notion that mental imagery is quasi-pictorial and functional, and we treat it as such in our model.

The Origins of Images

The image we experience may arise in any number of ways from any number of different kinds of representations in long-term memory (which are not experi-

enced directly, but are only experienced when activated) There are three related issues concerning the origins of images. First, it could be that experienced images are simply stored intact and merely retrieved *in toto* when later recalled. Alternatively, images may be actively constructed from material in long-term memory. If images are generated via the combination of separate "chunks" rather than retrieved holistically, more effort should be required to generate more complex images. In one experiment Kosslyn, Reiser, Farah, and Fliegel (in preparation) asked people to remember pictures of animals and objects that were drawn either with minimal detail or with many visual details. More time was required to form a visual image of a picture when it was drawn with many details. Kosslyn (1975) also found that subjectively larger images (of animals, in this case) required more time to generate than did smaller ones. If more "detail" is inserted into larger images, these effects of subjective size also indicate that images are in fact constructed and that construction takes time. These results do not make much sense if images are simply turned on like a slide that is projected on a wall.

Alternatively, perhaps images are in fact simply stored intact but are retrieved a little at a time. In this case, the simple amount stored will dictate how much time is required to retrieve the image. Kosslyn et al. tested this possibility by asking people to image geometric forms that could be described as having been formed by combining relatively few, overlapping figures or by combining relatively many adjacent figures. For example, the Star of David could be seen as two overlapping triangles or a hexagon and six triangles. Although the same actual forms were imaged, more time was taken if a form was initially seen as being composed of more units. One could argue, however, that the different descriptions simply led subjects to retrieve parts of encodings differently, and that in both cases the underlying representation was an integral encoding. It is important to discover whether the imagery system has the capacity for storing separate units and combining them into a single image. Thus, another experiment was performed to address this issue. These subjects were asked to memorize drawings that were presented in three different ways: (1) All of the object was drawn on a single page; (2) it was broken into parts arranged in the correct relative locations on two pages; or (3) it was broken into parts arranged on five pages. For any given subject, a particular drawing was presented in one of the three conditions, but three groups of subjects were used so that each drawing occurred equally often in each of the three conditions. Subjects first were shown the drawings, a page at a time if more than a single page was used, and were asked to be able to construct an image of the entire object ("mentally gluing" separate parts together if necessary). Later these people participated in a reaction-time task in which they were asked to image a drawing, push a button when it was present, and then answer questions about the image. Interestingly, even though the area covered and the number of details present were presumably the same in the images (a claim bolstered by the fact that time to "see" probed

properties was the same in the three conditions), more time was required to generate an image when the drawing had been divided into more units (on more pages). In fact, the increase in generation time was linear as a function of number of units.

Finally, given that images can be composed from separate encodings, the main question now concerns the sorts of representations used in this construction process. On one hand, images could be the result of assembling separate perceptual memories, like one assembles a jigsaw puzzle. On the other hand, images could be constructed using both perceptual and conceptual information, like arranging a set of photographs on a table in accordance with a description of the total configuration. It seemed to us that the latter alternative almost had to be true. After all, people apparently can construct images of novel scenes upon being given verbal descriptions of them. For example, nobody we have talked to seems to have trouble imagining Jimmy Carter standing on a surfboard riding a foaming wave, although no one claimed to have ever witnessed such an event. In this case, people seemed to be able to use the conceptual information underlying their understanding of the words to amalgamate various perceptual memories into a single scene. Gomez and Kosslyn (see Kosslyn, 1978b) performed a very simple experiment in order to demonstrate that conceptual information can in fact be used in image construction. People saw a six-by-three matrix of letters, which was then removed and named either the matrix of "three rows of six" or the matrix of "six columns of three." When later asked to image this matrix, more time was required if it had been conceptualized in terms of six columns instead of three rows. Kosslyn, Reiser, Farah, and Fliegel (in preparation) present other evidence that conceptual information clearly is used in storing and later generating visual mental images.

Thus, we know that images are spatial representations in active memory that can be generated from long-term memory in conjunction with conceptual information. Let us now see how our findings outlined thus far motivated the essential features of our model.

THE MODEL

The foregoing set of results led us to propose the following model of image structures and processes. It is most convenient first to discuss the data structures and then to consider the kinds of processes that make use of them.

Data Structures

Images (the quasi-pictorial entities we experience) are treated as surface representations generated from more abstract "deep" level representations.

The Surface Representation.   Images are represented as configurations of points in a matrix. This "display structure" has four properties:

1. Points in the display correspond to points of the represented object such that all interpoint relations are preserved. Thus, distance as such is preserved, as suggested by the scanning experiments.

2. Resolution is highest in the center and tapers off toward the edges. That is, the center region is most sharply in focus, and acuity fades until no cells are available to the interpretive procedures (the "mind's eye"). This property was motivated by our finding that the estimates of the angle of the mind's eye were affected by the criterion of "overflow."

3. The matrix is a short-term memory structure; material within fades and must be continuously regenerated. This property was motivated by data showing that more complex images are more difficult to maintain (see Kosslyn, 1975).

4. The medium has a "grain" such that if an image is too small, parts will be difficult to discern. This property was motivated by our findings on effects of image size on detection times and on subsequent memory for imaged objects.

The Deep Representation.   An image is represented in long-term memory in terms of files addressed by the name of the imaged object. There are two types of deep-image representations: First, the perceptual memory of the appearance (which is not semantically interpreted but corresponds to the products of "seeing that," not the products of "seeing as") is stored in a file containing r, θ coordinates. These polar coordinates specify locations (at some distance at some angle from an origin) where points should be placed in the surface matrix. A polar coordinate representation was chosen because: (1) it allows easy placement of images at different locations in the surface matrix (by shifting the location of the origin); and (2) it allows images to be easily generated at different subjective sizes (i.e., different sizes in the display, by multiplying the r values by a constant). We have data that people, in fact, can easily place images at different "locations" and can easily evoke them at different subjective sizes. The Polar coordinate representation also allows images to be generated at different angular orientations (by multiplying θ values by a constant), a property that has yet to be studied in human imagery. The perceptual memories that underlie the actual surface display may be stored in several files; one file corresponds to the "global" or "central" shape and serves as a skeleton upon which details may be placed.

The second type of deep representation consists of stored facts about an object; these facts are represented in a "propositional" format. Facts include information about: (1) how and where a part (represented as a file containing locations of points) is attached to the global or central image (e.g., a cushion is "flush on" a seat); (2) a description of a part's appearance, which consists of an ordered list of numbers, each of which indexes a procedure that searches for a

pattern of points in the surface matrix; (3) the name of the file that contains the "perceptual memory"; (4) a size-category tag, relative to an absolute standard (e.g., "very large," "small," etc.); and (5) the name of the superordinate category.

Image Processes

There are three sorts of image processes: There are procedures for generating, inspecting, and transforming images. In addition, we have also begun to hypothesize about how imagery representations and propositional representations of general world knowledge are accessed in the course of question answering, which is the central concern of this chapter. Before seeing how we might model the interface between imagery and nonimagery representations in answering questions, however, we must first get some feeling for how imagery per se might be used.

Generating Images. In generating an image of an object, the program first looks for the file that contains the propositional information about the object. If this file is then successfully located, it is searched for the name of the file that contains information about the literal appearance of the object (i.e., a list of polar coordinates that specify where points should be placed in the surface matrix). The first image file looked up is hypothesized to represent a "skeletal" image. The skeletal image is meant to contain "first glance" information. (Whether this is some sort of global shape information or simply information about the most centrally structured part is a question for future research.) This image file is then accessed, and the specified points are turned on in the surface matrix. Before each point is turned on, the r and θ values are set appropriately if a nondefault size, orientation, or origin is specified.

Once the skeletal image is printed out in the surface matrix, the program will stop unless the user has requested that a detailed image be generated (some data suggest that people do not add detail unless it is needed; see Kosslyn, 1980). If detail is requested, the program goes back to the propositional file of the object and checks for an assertion that a particular part belongs on the object (e.g., for a car, HASA REARTIRE). If such an assertion is found, the propositional file for the part is located and searched for the name of the image file (i.e., the file containing information about the literal appearance of the part). If this information is found, the name of the image file is stored (rather than being looked up later when needed, because if it is not there, the program need not continue). Following this, the proper location of the part is looked up. This information consists of a relation and a foundation part (e.g., UNDER REARWHEELBASE, for a car). Next, the propositional file for the foundation part is located (REARWHEELBASE, in this case), and the program searches for a description of the part's appearance located within. This description is a list of numbers, each of

which indexes a procedure that tests for various spatial configurations in the surface matrix. The procedures indexed by the description are then used to search the image in the surface matrix and—if successful—to delineate the boundaries of the foundation part. Once the proper location is found, the additional part is printed at the correct size and location in the surface matrix. After the part is successfully integrated into the image, or if any of the description procedures failed, the program returns to the propositional file of the object and checks for further parts and then attempts to integrate these parts into the image.

Thus, because of limited resolution and difficulty in finding foundation parts on subjectively smaller images, we expect people to be faster to generate subjectively smaller images than larger ones, but smaller images should take longer to inspect. These results were in fact obtained by Kosslyn (1975) and by Kosslyn, Reiser, Farah, and Fliegel (in preparation). We assume that this is in part because fewer details are placed on smaller images: People, like our program, may have difficulty in locating the foundation parts on subjectively smaller images and thus may integrate fewer details into them.

Inspecting Images.   The LOOKFOR procedures allow one to search an image for a given property or part. Let us consider an example: An image of a car is present in active memory, and the program is asked whether it can find the rear tire. When asked to find a given part (or object) in an image, the program first looks for the propositional file for the specified part. If this file is found, the program looks within it for a size tag (relative to an absolute standard). The program will not bother to search an image until the image is at the correct subjective size. Thus, the program then checks to see if the resolution (dot density in the matrix) of the image in active memory is within the range of the optimal resolution (associated with the sought part's size). If not, the image is expanded or shrunk, as appropriate. Following this, the description of the part is looked up, the correct region of the image is centered in the surface matrix, and the image is searched for the part (in the same way as already described when looking for a foundation part during image generation). If procedures indexed by the description of the part's appearance are successfully executed, the program responds that the part is present. If any one of the procedures fails, the program then looks up the names of additional image encodings in the appropriate region and inserts new parts into the image, which is then inspected again. If the part still cannot be found, the program responds in the negative (see Kosslyn, 1980, for details).

Transforming Images.   Images are altered in two ways: A "shift" transformation shifts the points defining a surface representation in some specified fashion. "Scanning" an image consists of moving the points across the surface matrix such that different portions of the depicted object seem to move under the

center, which is most highly resolved and most sharply in focus. "Rotating" an image consists of moving the points around a specified pivot. "Expanding" or "contracting" an image consists of migrating the points away from or toward a specified pivot (usually the geometrical center of the image). Each of these "shift" transformations moves only a part of an image at a time, and the rate of transformation is limited by how far portions may be shifted before the image seems to fragment. The advantage of this type of transformation is that no information needs to be stored in a special buffer, as would be necessary if the locations of all points composing the image were transformed before any were printed out. The disadvantage of this sort of transformation is that the procedure is iterative: Larger transformations require more operations. (Hence, larger distances to be scanned, rotated, or expanded/contracted will require more time than when smaller distances are involved.)

The other sort of image transformation, a "blink" transformation, is not iterative. In this case, an initial image is "erased," and a new one—exhibiting the required alterations—is generated. In general, this sort of transformation requires more effort than a shift, and hence most transformations should be accomplished by shifting portions gradually. Although this erase-and-regenerate process does require considerable effort, the amount of effort does not depend on the size of the transformation to be performed, as it does with iterative shift transformations. Thus, for larger transformations, a blink transformation may be more economical than a shift. (There are special problems with blink rotations, however, due to the fact that the procedures for integrating parts into an image are not orientation invariant. Thus, it is difficult to add details when the image is constructed at a specified angular orientation.)

IMAGERY AND QUESTION ANSWERING

The simulation initially was intended only to deal with image processing per se. However, it became obvious that we had to consider more than the mechanisms of image representation and processing; we also had to consider the role of imagery in cognition. Clearly, one would not need to use imagery if some sought information were represented explicitly in a propositional format. Thus, we began by embodying what seemed to us an intuitively sound model of how image and propositional processes may interact (see Kosslyn, 1980; Kosslyn & Shwartz, 1978). In our first program, if we simply inquired whether an object had a property (i.e., did not insist on image use in answering), the object's propositional file was first looked up in memory. Next, this file was searched for the queried fact. If the sought information was listed in the object's file, the program responded affirmatively. If it was not, the program then looked up the name of the object's superordinate category (for *car, vehicle*); that file was then

looked up and searched for the sought information. If the information was found here, the program responded affirmatively. If it was not listed here, an image was generated and searched for the property (as already described). A few simulation runs quickly showed us a fundamental error in this model: Fewer operations were required to decide that a particular car had a hood ornament (which required imagery in the program) than to decide that it did not have a brain. The literature on human reaction time, in contrast, shows that absurd false properties are rejected very quickly; one does not seem to decide that a given property is false simply by scanning through all of memory and failing to find the appropriate representation. Smith, Shoben, and Rips (1974) account for this "fast no" finding by positing an initial "relatedness check"; if objects share no common properties, search simply is not initiated. We simulated this process by storing a table of relatedness values, which is obviously inadequate as a theory of how relatedness is computed. But lacking any theory of how relatedness is derived, this "kluge" is satisfactory for present purposes. In our model, then, before search is initiated, relatedness is looked up; if an object and property are from completely different domains, the program makes a negative evaluation without actually searching the relevant files. If the property and the object are related closely enough, then the process described earlier is initiated.

The revised model just described makes at least one clear prediction: If the difficulty of answering a query is a reflection of how many underlying operations are necessary, then questions requiring imagery should be judged more difficult than those not requiring imagery. In addition, presumably the frequency with which one accesses some information will in part determine whether that fact is entered explicitly into a propositional file. Thus, more frequently considered facts ought to require imagery less often than less frequently encountered ones. We tested these predictions in a simple experiment. We composed a number of true and false statements (see Kosslyn, 1978a, for details) and printed these out in a random order on a page. Two groups of subjects were asked to decide whether each statement was true or false. One group was asked to rate, on a standard 7-point scale, how difficult it was to make this decision and how frequently they had thought of the predicated noun–property relation. These ratings were to be made for each statement immediately after the subject made his or her truth judgment. The other group of subjects, in contrast, rated on a 7-point scale how much they felt they had used imagery in arriving at their evaluations of the truth of each statement. Interestingly, the correlations between the difficulty ratings and the image-use ratings was $r = .80$, and the correlation between the frequency and image-use ratings was $r = -.64$. Both predictions, then, received support from the results. Unfortunately, these results are also consistent with a whole raft of models, many of which differ substantially from our model. We undertook a more systematic investigation to discriminate between what we took to be the main classes of alternative models, and we discov-

ered that another contender—not the model just outlined—emerged as most successful.

Discriminating Among the Models

The simulation includes two sorts of representations in long-term memory. This defines a "space" of five basic kinds of models of how information is accessed in the course of answering questions: (1) Only information in propositional files might be accessed; (2) only imagery encodings might be accessed; (3) propositional files might be accessed first, and then imagery files if necessary; (4) imagery files could be accessed first, and then propositional files if necessary; or (5) both imagery and propositional files might be accessed but not in any particular order (i.e., at the same time, or in different sequences with some probability, or via an alternating "time-sharing" switching system, and so forth). This last category of models includes those that are formally equivalent to parallel access models, where both sorts of information are retrieved in a "race," and Kosslyn, Murphy, Bemesderfer, and Feinstein (1977) argue that this is the most straightforward and elegant way to conceptualize this class. Thus, we treat this last class as "parallel processing" models, as is described later.

Kosslyn et al. (1977) tried to distinguish between the five classes of models by using a size-comparison task. In this task, people are asked to evaluate two named objects on the basis of size. The usual finding is that the larger the disparity in sizes between the objects, the faster the subject is to respond. The five classes of models just noted make different claims about how this task would be performed and why this "size-disparity effect" occurs. The models were evaluated in a simple variant of this task: Subjects learned to draw six stickmen, each of which was a different size and color. Following this, they learned to categorize the smallest three as "small" and the largest three as "large." The trick of the experiment involved having two groups, which differed only in how much overlearning of category labels was required. One group of subjects learned the size tags to a criterion of 500% overlearning, whereas the other only learned them to a criterion of 200%. Let us now consider how each of the five models accounts for the basic size-disparity effect, and then note the predicted effects of amount of overlearning according to each model. The models presented here are the most basic representatives of each class; see Kosslyn et al. (1977) for a more detailed discussion of variants.

Pure Propositional Models. Subjects retrieve a category-size tag for each of the to-be-compared objects. If the objects are relatively disparate in size (e.g., mouse and elephant), tags will mismatch (i.e., one might be "large" and the other "very small"), and a decision may be reached quickly. If objects are relatively close in size, tags will not mismatch (e.g., mouse and hamster may

both be labeled as "very small"). In this case, a second operation will be required, looking up detailed size information (e.g., in feet and inches) and using this to make the comparison. Thus, the closer in size two objects are, the more likely it is that two operations, instead of one, will be required—resulting in longer times.

According to this model, overlearning the size tags should result in general speeding up if overleaning affects ease of looking up the size information. Further, when stickmen are drawn from different size categories, there should be no effects of size disparity (because tags mismatch), nor should there be effects of size disparity when stimuli are from the same category (because detailed size information will be retrieved in both cases).

Pure Image Models. These models are obviously not viable as a general conception of how people answer questions; one can memorize definitions of abstract words, for example, that presumably cannot be encoded simply via mental images (cf. Wittgenstein, 1953). In the size-comparison task, however, images could be compared via the same interpretive procedures used during perception. In this case, whatever mechanisms are responsible for size-disparity effects in perceptual judgments could also underlie the analogous effects when people make the judgments from memory.

Because category information is not used at all in this model, we expect no effects of amount of overlearning of labels on decision times.

Propositional-Image Sequential Models. The most basic form of this class of models is like the pure propositional model except that images are consulted if tags initially match (i.e., for objects close in size).

This model also predicts only a general increase in speed with more overlearning of categories. Further, no size-disparity effects should arise when objects are drawn from different categories; when they are drawn from the same category, however, we now expect size-disparity effects (because image comparison will be used, producing perception-like effects as noted earlier).

Image-Propositional Sequential Models. These models entail first consulting an image and then, if necessary, consulting propositional information. One account of the size-disparity effect posits that the initial image comparison is cursory, in which case only large disparities will be discriminated. If size disparity is minimal, propositional information will be required to accomplish the evaluation. Alternatively, the image comparison is made as in a pure image model, and propositional information is used only if images are not available.

In this model, we do not necessarily expect any effects of overlearning, either on overall speed or on the size-disparity effect (see Kosslyn et al., 1977, for more details).

Image-Propositional Parallel Race Models.   In these models, one retrieves the size tags at the same time one is using imagery in trying to make a decision. The larger the size disparity, the more quickly the imagery process can operate (for reasons already noted for pure image models) and the more likely it is to "outrace" the propositional processes. However, when objects are disparate in size, it is possible for the propositional tag comparison not only to succeed (because tags mismatch) but to outrace the image-comparison process; when disparity is small, tags may match, in which case decisions must rely on the image processes (which will be slow because of discriminability problems).

On this view, no size-disparity effect will occur if tags are highly overlearned and members of a pair are drawn from different categories. In this case, overlearning should speed up time to locate the size tag in the propositional file and thus increase the probability that a comparison will be made on the basis of comparing tags instead of by comparing images. Imagery processes (i.e., using information stored in the "perceptual" files, either directly or via generating a surface image) should be used when members of a pair are within the same size category (so tag comparison will not produce a decision) and whenever tags are not well overlearned (so imagery processes are likely to "outrace" propositional processes). Thus, we expected the amount of overlearning to produce different effects depending on whether members of a pair were drawn from the same category or not.

The results of the experiment were clear-cut: Amount of overlearning was critical in determining whether size-disparity effects were obtained. Only with large amounts of overlearning of category labels were the effects of size disparity eliminated when to-be-compared items were from different categories (and then only within a limited subset of the data). This result is consistent only with the fifth class of models, as is discussed in detail in Kosslyn et al. (1977).

If, in fact, imagery and propositional information are retrieved at the same time, then it makes sense to ask how the two parallel processes interact. On one hand, they may be completely independent. On the other hand, subproducts of one may influence the processing of the other. Kosslyn et al. (1977) performed another experiment that bears on this question and which provides converging evidence that the image-propositional parallel model is worth taking seriously. This experiment examined the so-called congruity effect. That is, people are faster to say which of two large things is the larger than which is the smaller, but are quicker to say which of two small things is the smaller instead of which is the larger. Kosslyn et al. (1977) describe how different theories of the congruity effect are aligned with different theories of the size-disparity effect. Instead of recapitulating these competing theories here, let us simply describe the interesting prediction of the image-propositional parallel model. One way of accounting for the congruity effect is by appeal to a "sampling range." If one is set to expect two large things, one may have to "recalibrate" some retrieval and/or compari-

son processes in order to compare small things—which would require time. Similarly, if one is set to compare small things, recalibration may be required before one can compare large things. Now, if one retrieves propositional tags while one is performing an image comparison, perhaps the tags can be used to initiate the recalibration procedures *before* the imagery procedures (which are incorrectly calibrated) fail. That is, even if both objects are categorized the same way (e.g., mouse and hamster) and the tags cannot be used to generate a decision, perhaps the information that the objects are in a given size range can be used to calibrate the sampling range of the imagery processes. In this case, the size of the congruity effect should be lessened because recalibration should begin sooner. If this notion holds water, we would expect that because more over-learned size tags are retrieved more quickly (as demonstrated by Kosslyn et al., 1977), there would be less of a congruity effect when tags are highly overlearned than when they are only moderately well overlearned. And, in fact, Kosslyn et al. (1977) found this result, using the stickman paradigm already described (except that subjects were asked "which is larger" and "which is smaller" on half the trials).

But what evidence do we have that surface images per se have anything to do with the nonpropositional comparison processes? That is, perhaps the underlying deep representations of images (the r, θ pairs in the simulation) are accessed directly and images themselves are never constructed. Consider the following experiment: One is given a noun and asked to image the named object mentally. Further, one is asked to image it either at a subjectively normal size or at a seemingly very tiny size. Following this, one receives a second noun and is asked whether the object named is larger than the first object. If images of the two objects must be compared, we expected that more time would be required if a person starts off with a subjectively tiny image of the first object; in this case, one may have to "zoom in" prior to comparing the objects—an operation that is not necessary if one begins with a normal-sized image of the first object. We used this basic technique (also used by Holyoak, 1977) as a diagnostic for whether imagery was used to perform comparisons; if more time was required when people began with a tiny image instead of a normal-sized one, we assumed that this difference was due to manipulating the first image prior to comparison (as was, in fact, reported by subjects when queried after the experiment).

In this experiment, then, we were primarily interested in the time required to compare pairs composed of items of very similar sizes (hereafter referred to as "near pairs"), which presumably fell into the same "natural" size category. Subjects were divided into two groups; one group was taught to classify the items into two categories, "large" and "small," whereas the other group did not categorize the items. The category-learning group greatly overlearned the category labels. As before, we expected that when pairs included items from different categories and category labels were well overlearned, discrete processes would usually outrace imagery processes. If so, then when category labels were

different for two items, the initial subjective size of the first image should be irrelevant; in addition, the size-disparity effect should be attenuated as occurred in the stickman experiment described earlier. When pairs include items from the same category (for the category-learning group), in contrast, comparison of category labels cannot result in a decision, and images presumably must be compared before a decision can be reached. In this case, the subjective size of the initial image should be important, and we now expect to find size-disparity effects in the data.

Although the results of this experiment (actually, two experiments were performed to test these ideas) were somewhat complicated, the basic predictions were confirmed. Importantly, subjectively small initial images slowed down comparisons in the no-category-learning group in general and in the category-learning group when pairs contained items from the same category. This result, coupled with the postsession self-reports, seemed to implicate imagery in the comparison process, as described earlier. Further, in the category-learning group, the size-disparity effect was in fact eliminated when to-be-compared items were drawn from different categories but was found in all other conditions of the experiment.

Before we can take these results as support for the image-propositional parallel model, however, we must demonstrate not simple differences between a group receiving category learning (with large amounts of overlearning, as in the foregoing experiment) and a group receiving no category training; we must demonstrate effects of *amount* of overlearning per se (according to the logic already outlined). Thus, we tested an additional group of subjects. This group received identical instructions to the category-learning group except that they were trained on tag learning to a criterion of only 200% instead of the 1200% given to the other group (see Kosslyn et al., 1977, Exp. 4, for details of instructions and procedure).

The results of this experiment were encouraging: The most interesting findings are presented in Table 19.1. As is evident, size of the initial image influenced time to assess both pairs whose members were from different categories (''across-half'' pairs) and pairs composed of items from the same size-category (''within-half'' pairs) for the 200% category-learning group. For the 1200% category-learning group (reported in detail in Kosslyn et al., 1977), in contrast, there appear to be no overall effects of the size at which the first item of a pair was initially imaged. Initial image size did affect verification time in general, F' $(1, 60) = 6.74, p < .05$. Although the effects of image size seem greater when the first object was in fact the larger, this result was not significant with the quasi-F, F' $(1, 37) = 1.38, p > .1$, although it was significant in a standard ANOVA including only subject variance in the error term, F $(1, 22) = 6.26, p < .05$. Interestingly, we obtained significant effects of initial image size for both the within-half and different-half pairs in the 200% group, F' $(1, 60) = 5.93, p < .05$, but found no effects of initial image size in the across-half pairs of the

TABLE 19.1
Results from the 200% and 1200% Overlearning Experiments[a]

| Image Size: | First Object Actually Larger | | | | First Object Actually Smaller | | | |
| | Normal | | Small | | Normal | | Small | |
Group	W	A	W	A	W	A	W	A
200%	2.016	1.186	2.106	1.504	1.804	1.230	1.910	1.241
	(31.2)	(6.2)	(38.2)	(9.4)	(33.8)	(3.8)	(24.4)	(5.0)
1200%	1.193	.752	1.290	.772	1.194	.713	1.201	.707
	(17.5)	(2.5)	(23.8)	(2.5)	(21.2)	(1.9)	(22.5)	(1.2)

[a] Percent errors (in parentheses) are given below reaction times. A = across-half (different-category) pairs; W = within-half (same-category) pairs.

1200% group, as hoped. We did, however, find some effects of initial image size in the within-half pairs of the 1200% group (see Kosslyn et al., 1977, for details). As is evident in Table 19.1, for pairs wherein the first object was in fact larger, there were virtually no effects of initial image size for pairs with members from different categories for the 1200% group; for the 200% group, in contrast, initial image size had the expected effects, F' (1, 57) = 16.67, for the interaction. For pairs including items from the same category, in contrast, initial image size had equivalent effects for the two groups, $F' < 1$, for the interaction. Further, the 1200% group was faster overall that the 200% one, F' (1, 25) = 12.88, $p < .01$. Finally, across-half pairs generally were faster, F' (1, 26) = 31.79, $p < .01$, and subjects were faster when the first object was in fact smaller (and hence, presumably, less "zooming in" was usually required; see Kosslyn et al., 1977), F (1, 42) = 6.88, $p < .01$. No other effects or interactions were significant. Finally, the error rates for the two groups are listed in Table 19.1 under the reaction times. As is usually the case in these sorts of experiments, reaction times and error rates were positively correlated. The lower error rates in the across-half condition for the 200% gr   suggests that category learning did have some effects; unfortunately. ': .. ·ge size disparity was slightly larger for across-half pairs than for w'·· ,-nalf p   ·, which could underlie the differences in error rates.[1]

We also examined the effects of disparity in the relative sizes of objects composing pairs (not the image size). As in the previous experiment, we tested for the effects of size disparity in the across-half (different-category) and within-half (same-category) pairs for both groups; this analysis was performed like the one reported for the near pairs in Experiment 4 of Kosslyn et al. (1977).

[1] The authors discovered that Table 2 in Kosslyn et al. (1977) contains an error: The across-half and within-half subheadings are reversed. As in the present case, errors there did in fact increase with increases in verification times.

These results are presented in Table 19.2. As is evident, we again replicated the basic result of the first stickmen experiment described earlier. For the 1200% category-learning group (wherein category tags were greatly overlearned), relatively near across-half pairs (containing items from different categories) were evaluated in the same amount of time as were relatively far (i.e., more disparate) pairs; when members of a pair were within-half (and from the same category), size-disparity effects were obtained as before, with more time being required for relatively near pairs. For the 200% group, in contrast, pairs containing names of relatively similar sized objects required more time to evaluate than pairs naming objects differing more widely in size; this was true of both across-half and within-half pairs. For the comparison of data from the overlearning groups, the three-way interaction between group, relative size disparity, and the across-/within-half variable was significant, F (1, 28) = 10.48, $p < .01$. As usually occurs in these experiments, errors again tended to be positively correlated with verification times; there was no obvious evidence of speed-accuracy trade-offs in our results.

In addition to the main experiment just discussed, Kosslyn et al. also performed another experiment, using a different set of items. These people evaluated only pairs composed of items quite disparate in size (e.g., ant-elephant, tomato-dishwasher, teapot-iceberg). These subjects received instructions and procedure exactly like those given to the category-learning groups in the main experiment except that they never learned to categorize the items into "large" and "small" categories. This group was tested to see whether "natural" categories—which probably differed for the largest and smallest items used to compose the far pairs—would be used instead of imagery comparisons when near pairs (which seem to require imagery) were eliminated. The results of this experiment were clear-cut: There were absolutely no effects of the size of the initial image, nor were any interactions with initial image size significant.

In summary, then, the subjective size of initial image had no effects for the 1200% category-learning group when members of a pair were from different categories. The 200% category-learning group, in contrast, apparently used im-

TABLE 19.2
Effects of Size Disparity in the Two Conditions[a]

	Slight Disparity		Large Disparity	
	W	A	W	A
200% group	2.144	1.400	1.979	1.291
	(41.2)	(7.5)	(26.8)	(8.2)
1200% group	1.308	.751	1.183	.776
	(23.8)	(2.5)	(18.1)	(1.9)

[a] Percent errors (in parentheses) are given below reaction times. A = across-half (different-category) pairs; W = within-half (same-category) pairs

agery regularly, as was evinced by strong effects of initial image size: More time was required when these people started out imaging the first item of a pair at a tiny subjective size instead of at a larger, more normal size. If images of the named items are compared, presumably time is required to adjust the tiny image prior to comparison, whereas the normal-sized image needs no (or very little) adjustment prior to comparison.

Category learning also influenced the time needed to evaluate near and far pairs: For the 1200% category-learning group, there were no effects of size disparity (within a small range of sizes) for across-half (i.e., different-category) pairs, whereas for the 200% group, we found the usual effects of size disparity for these items, near pairs requiring more time to assess than relatively far pairs. These results are entirely consistent with the notion that category tags could be used by people in the 1200% overlearning group to reach decisions before analogue imagery comparisons were completed in a parallel "race." When no differentiating category tags were available, or when tags were not highly over-learned, however, the subjects consulted pairs of images and based decisions upon this comparison.

The results from the two groups that did not learn to categorize the items are also entirely consistent with the present theoretical conception: For the group evaluating the near pairs, more time was required to make a comparison when people started out with a subjectively tiny image as opposed to a normal-sized one; for the group evaluating only pairs very disparate in size, in contrast, there were no effects of the subjective size of the initial image. These results support our claim that "natural" (i.e., not learned in a laboratory) categories can be used in evaluating relative sizes. Only when objects fall into the same natural category should imagery necessarily be consulted. Our near pairs were composed of objects that probably do fall into the same natural category, and hence it is gratifying that imagery does seem to be used habitually in evaluating these pairs. The results from the no-category-learning group with the far pairs supports the notion that our model can be generalized to real-world situations, wherein categories are not explicitly taught for use in the present task. If our model only accounted for tasks wherein categories were explicitly taught, it would be of little value.

INDIVIDUAL DIFFERENCES IN IMAGERY USE

The design of a meaningful test of individual differences in imagery requires at least two things: First, we must have a theory of the ways in which people can differ in how they represent and process information. Second, we must isolate which variables will affect various aspects of information processing within the context of the theory. The first step, then, was to begin to assess which aspects of the model were flexible, that is, were open to a range of variation across indi-

viduals. We have begun with an interest primarily in the role of imagery in question answering; in the concluding section of this chapter, we speculate further about individual differences in other aspects of imagery. Our model postulates that the speed and ease of retrieval of properties in an object's propositional file should in part determine whether an image will be consulted when one is answering a question. This is a fundamental characteristic of the image-propositional "race" discussed in the foregoing section. Our basic approach, then, was to ask subjects to answer questions that included nouns and properties that were associated more or less strongly (e.g., "A zebra has stripes" includes a noun and a property that are highly associated, according to normative ratings; whereas "A zebra has knees" contains a noun and a property that are not very highly associated). Our parallel race model posits that if the search for the property in the object's propositional file should fail, then one will attempt to deduce the answer via looking up the superordinate category. In this case, the ease with which one may retrieve the superordinate file for the object and the subsequent ease of finding the property in the superordinate's propositional file should affect the propositional search process. The relative speeds of this sort of processing and imagery processing should determine the outcome of the race between the imagery processes and the propositional search and deduction processes. Thus, this model accounts for our earlier result that more difficult statements more often were evaluated with imagery in the following way: "Less difficult" statements are those requiring relatively few propositional operations that run relatively quickly; "increasingly more difficult" statements tap relatively poorly overlearned propositional information, require deduction, or cannot be answered via propositional search and deduction because the requisite encodings are not available. Each of these factors operates to impair propositional search and hence to increase the likelihood that imagery will be used.

We tried to capture the two characteristics of the deductive aspect of the propositional search procedures by varying: (1) the association strength between an object and its superordinate, and (2) the association strength between a part or property of an object and the object itself. We hypothesized that a low association strength between an object and a property of that object implies that the property is very low in the propositional file of the object, or that the object's file will be unlikely to contain a proposition stating the relationship between the object and the part. In either case, the propositional representation should require more time to locate than when object and part are highly associated, and the imagery system will become more likely to "outrace" the propositional lookup/deductive processes. Similarly, if the part is not listed in the object's file, then the association strength between the object and the superordinate should reflect the ease of looking up the superordinate's name (in the object file) and thus should influence the time to look up the superordinate's file. Thereafter, the association strength between the part and the superordinate would reflect ease of looking up the proper entry. Thus, low association in both cases should lead to

longer processing times and again should increase the probability that imagery processing will produce the answer faster than will propositional processing.

We decided to begin to explore these predictions with a very simple task: We merely asked our subjects to tell us whether a declarative statement was true or false. Each statement had the form "An x has y" or "An x has a y." A list of 152 object-property pairs was composed to be used in the construction of statements for the verification task. Each property was composed of an adjective followed by a noun (e.g., a canary has *yellow feathers*). We first wanted to obtain proper superordinates for each object in our pairs. Thus 17 Harvard Radcliffe undergraduates were asked to write down up to three superordinates for each of the objects. We asked our subjects to write these names in the order in which they had occurred to them. For each subject, the first response to each object was later assigned a score of 3; the second, a score of 2; and the third name, a score of 1. The superordinate with the largest mean score was taken as the "best" superordinate for the object in question and was used in categorizing our actual test items in our analysis of the data.

We now had a list of object-property-superordinate triplets. For each triplet, we now obtained association ratings between each pair of items (words) from a new group of 24 Harvard/Radcliffe undergraduates (15 women, 9 men). These people were asked to rate, on a standard 7-point scale, how highly associated the members of each pair of items were with each other. We also asked these people to rate how easy it was to form a mental image of the object and how easy it was to see the part or property on their images of the object. Both ratings again used a standard 7-point scale. These latter imagery ratings were obtained in an effort to estimate the ease of performing the imagery processes, which also should affect the outcome of the underlying race. Perhaps fortunately, the ratings of ease of forming and inspecting an image of the object showed very little variance. Thus, we postulated that for the present items, imagery processes should be relatively constant for items differing in association strength. Differences in imagery use, then, ought to be determined by differences in speed of the propositional lookup/deductive system.

After obtaining the ratings just described, we realized that the familiarity of an object could have an effect on the processes under investigation if familiarity dictates how many encodings are likely to be entered in an object's propositional file. That is, more familiar objects could have more properties explicitly encoded and, hence, would less often require deductive processes. Thus, we obtained ratings of familiarity with the objects from a new group of 23 Harvard/Radcliffe undergraduates (12 women, 11 men), and at the same time, we collected an additional set of object-part association ratings for each of the 152 triplets from these people. The new association ratings were collected to be compared with the first ratings to establish their reliability. The correlations between the first set of object-part association ratings and the second set was $r = .90$. The second set of

association ratings were used in all subsequent analyses because we had slightly more confidence in them, given that the overall ratings task was easier (because fewer ratings were required) for these subjects; in addition, because we used the familiarity ratings, it seemed a good idea to use the association ratings from the same people.

The object-part-superordinate triplets were used to construct a list of 152 statements, each asserting that the object had the part (e.g., "A blimp has metal propellers"). In addition, an equal number of false statements were constructed to be used as distractors in the experiment. These false statements ranged in difficulty from being quite subtle and nonobvious to being relatively easy (according to our intuitions). A new group of 20 Harvard/Radcliffe undergraduates (17 women, 3 men) decided whether each of the 304 statements was true or false. Immediately after each true/false evaluation, the subjects were asked to rate how much they had consulted an image in arriving at their decision, again using a 7-point scale (in the manner described earlier).

The triplets were divided into eight possible inter-item association strength patterns. The association between the object and property could be either high or low; the association between the part and superordinate could be high or low; and the association between the object and superordinate could be high or low. Thus, we constructed a $2 \times 2 \times 2$ cube to represent every possible combination of interitem association ratings, as is evident in Table 19.3.

Within each cell, 10 triplets were retained for analysis. For four of these cells, we kept the 10 triplets with the highest mean object-part association ratings. These four cells are the top two cells in each of the 2×2 tables shown in Table 19.3. Similarly, the remaining four cells were filled with 10 triplets with the lowest mean object-part association ratings. Within each of these groups of four cells, items were further sorted according to whether the mean ratings were above or below the overall mean for that dimension. When the object-part association for a particular statement was greater than the overall mean object-part association between all statements, the association was considered "high"; if the association was below this mean, it was considered "low." The same procedure was used to sort the statements on the basis of part-superordinate association strength and on the basis of object-superordinate association strength. Thus each statement was assigned to one of the eight cells in Table 19.3. In selecting the items to be rated, we tried to ensure a spread along the three dimensions and that all combinations of values would be likely; fortunately, our intuitions proved reasonably sound and allowed us to assign 10 items to each cell, resulting in 80 statements being used in the analyses.

The mean imagery-use ratings obtained for each cell are shown in Table 19.3. The main hypotheses of this study did in fact receive support: The four cells with high association between the object and the part and the four cells with low association between object and part contain the lowest and highest imagery-use

TABLE 19.3
Mean Imagery Ratings in the Eight Conditions[a]

	OBJECT SUPERORDINATE				
	High			Low	
	PART SUPERORDINATE			PART SUPERORDINATE	
	High	Low		High	Low
OBJECT-PART High	2.775	3.355	OBJECT-PART High	3.215	3.390
OBJECT-PART Low	4.085	3.960	OBJECT-PART Low	3.490	4.315

[a] *High* and *Low* refer to relative association strength along the relevant dimensions.

ratings, as expected. That is, the main effect of object-part association strength was significant, $F(1,72) = 16.96$, $p \leq .001$. Note also that the cell in Table 19.3 with high association between all variables and the cell with low association between all variables contain the lowest and highest imagery-use ratings, respectively. This is, of course, as we expected. Furthermore, a contrast designed to test our overall experimental predictions (i.e., comparing cells in which we predicted relatively high imagery use with those in which imagery use was not expected) was significant, $F(1,72) = 8.16$, $p \leq .006$. A regression analysis revealed that familiarity per se was unrelated to rated imagery use; thus we have not included these ratings in subsequent analyses. We performed a number of additional analyses (e.g., all possible correlations and partial correlations), but none of them adds to the conclusions evident in the results reported above. Thus, our hypotheses received good support from the data; although the magnitude of the differences observed was not overwhelming, the differences discussed so far were in fact in the directions predicted and statistically significant.

One of the purposes of developing a general model, we have claimed, is that it helps to focus one's attention on the junctures where important individual differences may take place. In the present model, the most important processes are: (1) the ease of looking up a listing of the part in the object's propositional file, and should this fail; (2) the ease of looking up the name of the most common superordinate and from there looking up the part in the superordinate's propositional file. We conducted a second analysis in order to discover if the variables that presumably reflect these factors were in fact sensitive to individual differences. That is, if these are the critical variables, then deviations from the mean association values ought to be particularly important, and our mean ratings should not predict everybody's performance equally well. In the first analysis we were interested in how well the data fit the predictions of our model, and thus we

pooled the results over subjects and considered the mean ratings for each statement as the unit of analysis. In the present analysis, subjects were included as a factor. Not surprisingly, we discovered substantial differences among the mean scores for different people, $F(19, 1368) = 11.67$, $p < 1.0 \times 10^{-6}$. More importantly, however, we found that only the two most relevant of the six possible interactions with subjects were significant. First, the effects of object-part association strength determined by our previous ratings were different for different people, $F(19, 1368) = 5.67$, $p < 1.0 \times 10^{-6}$. Second, the three-way interaction between subjects, object-superordinate association strength and part-superordinate association strength also was highly significant, $F(19, 1368) = 2.28$, $p < .005$.[2] These results, then, are exactly as one would expect if our model is correct. We can therefore feel some confidence in looking more closely at individual differences within the context of our model. Not only that, but we can now turn a sow's ear into a silk purse: The low mean differences in the averaged data might be expected if there are substantial amounts of individual variance along the relevant dimensions; large and consistent averaged differences would have cast a shadow on the possibility of significant differences among subjects.

The Imagery-Use Test

Our test is in the very first stages of development and experimental validation. Thus far, we have explored only the crudest ways in which individuals may differ. That is, we have only looked at the general tendency for people to use or not to use imagery in answering questions. Provided we can find evidence of stable individual differences in this task, we will then try to isolate the particular variables underlying particular individuals' preferred strategies. It seems reasonable to suspect that each combination of the eight cells describes some topic domain for any given person, but the particular topic domains described by each cell may vary. Further, although we may all use imagery in answering some kinds of questions, we may do so for different reasons (e.g., lack of propositional encodings). We are betting, however, that above and beyond particular domains,

---

[2]The model leads us to expect relatively little imagery use when the object-part association strength is low but both the object-superordinate and part-superordinate association strengths are high; in this case, the deductive processes still ought to "outrace" the imagery ones on some proportion of the trials. The results, however, did not bear out this prediction; people reported using imagery relatively often in this condition, $\bar{X} = 4.085$ (see Table 19.3). The present result offers an explanation for this failure: The individual differences simply may have been potent enough to overcome effects predicted on the basis of the mean ratings. This is especially likely as there was a wide range of different superordinates offered for each of our objects in the initial ratings task, and the subjects in the present experiment may well have performed deductions via superordinates other than the ones upon which we based our predictions.

some people may consistently use, or not use, imagery. Further, we intend to explore the claim that differences in imagery use may arise for different reasons in different people (e.g., one might be slow in making deductions; another might be especially fast in image generation and inspection). Before lunging ahead, however, we must be certain that the horse is placed firmly in front of the cart.

The Imagery-Use Test consists presently of a total of 32 statements, 16 of which are true and 16 of which are false. These items were selected from the original 304 statements described earlier. The items were chosen such that a wide range of imagery-use ratings was represented. Further, no noun or property is mentioned more than once.

For each statement in our test, the subject is asked first to indicate whether the statement is true or false. Then the subject is asked to make one of three judgments regarding the role of imagery use in assessing the veracity of the statement. The alternatives are: (1) An image was consulted in making the truth judgment; (2) an image was not consulted (even though an image may have been present); or (3) the subject is unsure of whether an image was or was not consulted in making the truth judgment. We switched from the 7-point-scale rating technique used earlier because we feared that we had been confounding an imagery-use judgment with a confidence rating (of the use judgment itself—for the intermediate or "unsure" cases). For each statement rated "used imagery," a score of 2 was added to that person's total score; for each statement rated "uncertain," a score of 1 was added to the total score; and for each statement rated "did not use imagery," a score of 0 was added to the total Thus, the total score on the test presumably reflects how much the subject used imagery in evaluating the veracity of the statements.

The first stage in test validation, it seemed to us, was to compare our test to other available tests of various imagery abilities. The tests most commonly used today are probably the Gordon Test of Visual Imagery and the Vividness of Visual Imagery Questionnaire (which has been used very successfully by Finke, 1980); in addition, we also administered the Visualizer-Verbalizer Questionnaire, which has recently been offered as an improvement on Paivio's much lengthier Ways of Thinking test (see Richardson, 1977). We administered the four tests to 35 Harvard/Radcliffe undergraduates (21 women, 14 men). A Latin square design was used to ensure that each test was presented in each order as often as was possible.

The Gordon test (see Richardson, 1969) purports to measure how well one can manipulate one's images. It consists of a series of questions about one's control over a mental image of a car. These questions are presented in order of (assumed) increasing difficulty. The test requires that one closes one's eyes and then tries to image a described scene (e.g., a car standing in the road in front of a house). The more items one reports being able to image, the higher one's score is. In our sample, we found little variance in scores on the Gordon test; most people seemed to have no trouble in imaging virtually all of the scenes. Thus, it is not

surprising that we found essentially no correlation between scores on our Imagery-Use Test and scores on the Gordon test, as is evident in Table 19.4. The Gordon test did, however, correlate with the ''eyes open'' and total score on the VVIQ (the more control one had over one's images, the more vivid they were, or vice versa), as is evident in Table 19.4.

The Vividness of Visual Imagery Questionnaire (see Marks. 1973, 1977) attempts to measure the vividness of a person's visual imagery. This test is an elaboration of the visual scale on the Betts test (published in Richardson, 1969). A person is asked to image each item (e.g., a rainbow appearing around a rising sun) and to rate the vividness of the image on a 1- to 5-point scale (1 indicating a very vivid image, 5 a nonexistent one). The subject is asked first to image all the items with eyes open and then to go back and rate the items a second time, now forming the images with eyes closed. The ratings from the VVIQ ''eyes closed'' did, in fact, correlate significantly with our test, as is evident in Table 19.4. Note that low scores on the VVIQ indicate vivid imagery and that more vivid imagery was associated with higher scores on our test. Thus, people with more vivid imagery seemed to have a greater tendency to use imagery while answering questions, or vice versa. This relationship was not very strong, however, and did not prove significant when the ''eyes open'' or total VVIQ score was considered.

Finally, the Visualizer-Verbalizer Questionnaire is designed to assess one's preferred mode of thinking. A verbalizer is expected to use words more often than will a visualizer; furthermore, a verbalizer should feel more comfortable using words while thinking. A visualizer, on the other hand, purportedly should tend to think in terms of mental images and should experience more vivid images and dreams than will a verbalizer. We doubt that there is a necessary relation between visual and verbal abilities (see Kosslyn, 1980, for a discussion) but felt it would be worthwhile to compare our test to this one. The VVQ is composed of 15 statements. The subject is asked to indicate whether or not each

TABLE 19.4
Correlations Among Scores on the Four Tests and Sex

	Sex	IUT	VVQ	VVIQ₁[b]	VVIQ₂	VVIQ₁	Gordon
Sex	X	.33[a]	.09	.09	.06	.01	.04
IUT		X	-.05	-.26	-.34[a]	-.31	.09
VVQ			X	.14	.19	-.17	.23
VVIQ₁				X	.88[a]	.97[a]	-.38[a]
VVIQ₂					X	.97[a]	.29
VVIQ₁						X	-.35[a]
Gordon							X

[a] $p < .05$.
[b] VVIQ₁ is with eyes open; VVIQ₂ is with eyes closed; and VVIQ₁ is the total score.

statement is an appropriate description of himself or herself. A high score on the test places one on the visualizer end of the continuum; a low score indicates verbalizer tendencies. Not only did this test fail to correlate with ours; it also did not correlate with any of the tests we administered, as is evident in Table 19.4.

Table 19.4 also presents correlations between the four tests and the sex of the subjects. Interestingly, our test was the best indicator of male-female differences in imagery processes and, in fact, was the only one to indicate significant sex differences, r (33) = .331, $p \leq .05$. Males, on the average, seemed to have a slightly greater tendency to use imagery than did females. This finding would seem to contradict Galton's (1883) conclusion that imagery was the characteristic mode of thought for women and children. Perhaps our finding is not surprising given the consistent result that men score higher on tests of spatial abilities than do women (see Tyler, 1965). We should be cautious, however, in equating spatial abilities with imagery abilities, or in equating verbal abilities with facility in formal reasoning. Surely, spatial reasoning can be accomplished without imagery (e.g., see Boden, 1977, for a discussion of relevant computer programs), and formal reasoning can be facilitated by image use (cf. Shepard, 1978). In summary, then, our test does seem to be measuring something different from what is measured by the other tests, although image use may be related (weakly) to image vividness.

Further Development of the Imagery-Use Test.   Until now we have been considering only the crudest ways in which people may differ in imagery ability. Variables such as vividness of, or control over, one's imagery have not proven to be very instructive to date, possibly because their study has occurred in an almost atheoretical context. Our test, however, is evolving from a rich theoretical background. Future refinements of the test will be directly relevant to the processing assumptions we have entertained in our model. Ideally, a refined version will allow us to localize the particular processes that underlie the ways in which a given individual deviates from the mean. In this section we discuss some of the ways in which we are trying to extend our test.

As a first set of refinements, consider the following: People reporting using imagery relatively often could be thought of as having a relatively slow propositional lookup time (perhaps because of impoverished propositional representation, as may occur in children; see Kosslyn, 1978a) and/or a relatively fast image generation and inspection time. A person reporting less than average use of imagery would differ in the opposite direction, having either a very fast propositional file search and/or a slow imagery generation-verification process. We could differentiate among these possibilities by adding subscales to our test. First, we would need to add another dimension to the 2 × 2 × 2 taxonomy presented in Table 19.3. Ideally, in each cell, half of the items would be rated to be easily imaged and seen, on the average, and half would be generally agreed to

be difficult to image and inspect (perhaps this could be induced by systematically varying the size of the properties). Then, by looking at relative scores in each cell (subscale), we could isolate why a given individual used imagery. For example, we could localize the basis for imagery use if we found no difference for items rated more or less easily imaged (indicating that image speed was not crucial) but found high ratings of imagery use for all statements including low object-part associations (indicating that part-superordinate and object-superordinate associations were irrelevant and, hence, did not reflect differences in the corresponding processes—which presumably were very slow or nonfunctional). More fine-grained analyses can be drawn from detailed comparison of items from different cells.

Validating the Imagery-Use Test. The reader should note that at present, we have no guarantee that our test does in fact measure what we designed it to measure. We currently are conducting some reaction-time experiments to validate our test and will validate each subscale in similar ways. For example, we are using the task reported in Kosslyn (1976a), where subjects decided whether animals had various properties either by referring to an image or simply by answering as quickly as possible. In this experiment, there were two kinds of "true" properties—those that were small but highly associated (e.g., for a mouse, whiskers) and those that were large but not highly associated (e.g., back). With imagery instructions, we found that the large/unassociated properties were verified more quickly than the small/associated ones, but the reverse was true when imagery instructions were not used. In a related experiment, Kosslyn (1976b) found that even when no imagery instructions were given, first graders who reported spontaneously using imagery when questioned after the task were faster with larger unassociated properties than with smaller/associated ones, but children who claimed not to have used imagery showed the reverse pattern. This result suggests that if our test is valid, we should be able to use test scores to predict the direction of the difference in verification times for the two kinds of properties: People scoring high on the test should tend to be faster in accordance with increasing size, whereas people scoring low on the test should be faster with more associated properties. This basic technique could be used to validate the different subscales by varying association between object, part, and superordinate in the appropriate ways and looking for differences in the predicted directions in each of the cells of our design. Further, we feel it is important to validate the test using a variety of independent tasks. Thus, we also plan to ask people to take part in an experiment wherein they image an object and then mentally focus on one end. The end they focus on may contain a property about to be probed, or the other end may contain the probed property. In this task, we would simply ask the subject to answer whether a query was true or false, without necessarily referring to his or her image. If imagery is used, we expect that more time will be

taken when the subject has to scan to see the property than when the individual initially is focusing on the appropriate portion of the object. If imagery is not used, no such scanning effects should be apparent.

In addition to validating the test per se, we also wish to validate the theoretical assumptions of the model underlying it. One of our basic predictions is that lower object-part association values will lead to imagery use, on the average, because the imagery processes will tend to outrace the propositional ones. In addition, if a person can form and inspect images very easily, we expect imagery processes to outrace propositional ones. If so, then we do not expect any effects of object, part, or superordinate associations. The appropriate sort of experiment would be exactly analogous to those described earlier, but now using association strength as the index of speed of propositional processing instead of amount of overlearning. As before, effects of initial image size should occur only when one "zooms in" to "see" a queried part on the image. We can also perform separate tasks to validate the subscales, some of which would measure how quickly subjects can generate and inspect images and how much differences in association strength affect relative verification times. These measures should predict which subscales will engender imagery use and which will not, if our general imagery-propositional parallel race model is correct.

In addition to these relatively subtle analyses into sets of subscales, we can go even deeper. Say someone is slow to inspect images. This could reflect poor resolution of "the mind's eye." We can measure this by asking subjects to image a grating and pretend that they are walking back from the grating in their image. If subjects can reliably estimate how "far away" a grating is when it seems to blur in their image, we can use this to estimate the resolving power of their mind's eye. If slow inspection time is not due to poor resolution, it may be due to poor "interpretive procedures," poor tests that classify spatial patterns into semantic categories. If so, then we expect similar deficits in detecting parts of actual pictures (if the same interpretive procedures are used in classifying parts of images and percepts, as was suggested earlier). There does not seem to be any necessary limit to how subtly one may localize the particular processing components underlying a person's proclivities for using imagery in question answering.

CONCLUSIONS

This chapter has outlined a theoretical approach to the study of individual differences in imagery use. We have concentrated on the question of when one would use imagery in answering questions but have said virtually nothing about the implications of this sort of propensity. The most obvious implications bear on how people may approach problems. Shepard (1978) notes numerous cases where famous scientists and inventors have reported arriving at solutions to problems by imaging the essential elements of the problem in some way. On a

less rarefied level, an ordinary person's tendencies to use imagery might influence the kinds of strategies he or she would take in solving mathematical problems; for example, whether one would try to couch problems in geometrical or algebraic terms. Further, the ease with which one can solve a problem may depend on how that problem is couched; perhaps frequent users of imagery will find it easier to solve spatial/pictorial problems than abstract, "logical" ones. The obvious extension of this idea is to learning: Perhaps the effectiveness of visual aids is different for different sorts of people, for example. It need not be that a picture is worth a thousand words for all of us.

In closing, we are reminded of an ancient Chinese parable (originally related to one of us by Eleanor Rosch, who has forgotten the source and told it much better): Once upon a time the adult animals in the forest got together and lamented the state of the younger generation. They were hanging around clearings, loitering, and not developing their potential. So the adults decided to start a school for their offspring. When the question of a curriculum arose, the bears promptly insisted that digging be included; it is an absolute necessity to dig, they pointed out. And the birds chirped in that flying was definitely not to be overlooked; nor climbing, said the squirrels. So, soon there were young birds with broken wing tips from trying to dig, baby bears with broken backs from trying to fly, and so on. . . . The Moral of this story should *not* be that some people are best fitted for some kinds of jobs or tasks, that some of us are birds and others bears, that some ought to fly and some to dig. Rather, once one knows what sort of animal one is, one then knows how to approach a particular task, whether to dig with wing tip or foot, to fly with flapping arms or in an airplane. Hopefully, the systematic study of individual differences within more general theories of information processing will make it possible to learn who is a bird and who is a bear, and how to take advantage of one's proclivities in learning (or to know which aspects to strengthen before attempting to learn), making it easier, more efficient, and more enjoyable for one to acquire a new skill, a new body of information, or a new way of thinking about things.

ACKNOWLEDGMENTS

REFERENCES

Angell, J. R. Methods for determination of mental imagery. *Psychological Monographs*, 1910, *13*, 61-107

Betts, G. H. *The distribution and functions of mental imagery*. New York: Columbia University, 1909.

Boden, M. *Artificial intelligence and natural man*. New York: Basic Books, 1977.

Davis, F. C. The functional significance of imagery differences. *Journal of Experimental Psychology*, 1932, *15*, 630-661.

Fechner, G. T. *Elemente der Psycholphysik*. New York: Holt, Rinehart & Winston, 1966. (Originally published, 1860.)

Fernald, M. R. The diagnosis of mental imagery. *Psychological Monographs*, *58*, 1912.

Finke, R. A. Levels of equivalence in imagery and perception. *Psychological Review*, 1980, *87*, 113-132.

Galton, F. *Inquiries into human faculty and its development*. London: Macmillan, 1883.

Glanzer, M., & Clark, W. H. The verbal-loop hypothesis: Conventional figures. *American Journal of Psychology*, 1964, *77*, 621-626.

Hebb, D. O. Concerning imagery. *Psychological Review*, 1968, *75*, 466-477.

Holyoak, K. J. The form of analogue size information in memory. *Cognitive Psychology*, 1977, *9*, 31-51.

Koffka, K. *Zur Analyse der Vorstellungen und ihrer Gesetze*. (Cited in Woodworth, 1938), 1912.

Kosslyn, S. M. Information representation in visual images. *Cognitive Psychology*, 1975, *7*, 341-370.

Kosslyn, S. M. Can imagery be distinguished from other forms of internal representation? Evidence from studies of information retrieval time. *Memory & Cognition*, 1976, *4*, 291-297. (a)

Kosslyn, S. M. Using imagery to retrieve semantic information: A developmental study. *Child Development*, 1976, *47*, 434-444. (b)

Kosslyn, S. M. Imagery and cognitive development: A teleological approach. In R. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978. (a)

Kosslyn, S. M. Imagery and internal representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978. (b)

Kosslyn, S. M. Measuring the visual angle of the mind's eye. *Cognitive Psychology*, 1978, *10*, 356-389. (c)

Kosslyn, S. M. *Image and mind*. Cambridge, Mass.: Harvard University Press, 1980.

Kosslyn, S. M., & Alper, S. N. On the pictorial properties of visual images: Effects of image size on memory for words. *Canadian Journal of Psychology*, 1977, *31*, 32-40.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 47-60.

Kosslyn, S. M., Reiser, B. J., Farah, M. J., & Fliegel, S. L. *Generating visual images*. Manuscript in preparation, 1980.

Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. Category and continuum in mental comparisons. *Journal of Experimental Psychology: General*, 1977, *106*, 341-375.

Kosslyn, S. M., & Pomerantz, J. R. Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 1977, *9*, 52-76.

Kosslyn, S. M., & Shwartz, S. P. A simulation of visual imagery. *Cognitive Science*, 1977, *1*, 265-295.

Kosslyn, S. M., & Shwartz, S. P. Visual images as spatial representations in active memory. In E. M. Riseman & A. R. Hanson (Eds.), *Machine vision*. New York: Academic Press, 1978.

Marks, D. F. Visual imagery differences and eye movements in the recall of pictures. *Perception & Psychophysics*, 1973, *14*, 407-412.

Marks, D. F. Imagery and consciousness: A theoretical review from an individual differences perspective. *Journal of Mental Imagery*, 1977, *1*, 275-290.

McKellar, P. The investigation of mental images. In S. A. Barnett & A. McLaren (Eds.), *Penguin science survey*. Harmondsworth, England: Penguin Books, 1965.

Muller, G. E. *Zeitschrift Psychologie Erganzbd.* 1917, 9. (Cited in Woodworth, 1938.)

Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.

Richardson, A. *Mental imagery.* New York: Springer, 1969.

Richardson, A. Verbalizer-visualizer: A cognitive style dimension. *Journal of Mental Imagery,* 1977, *1,* 109-126.

Sheehan, P. W., & Neisser, U. Some variables affecting the vividness of imagery in recall. *British Journal of Psychology,* 1969, *60,* 71-80.

Shepard, R. N. The mental image. *American Psychologist,* 1978, *33,* 125-137.

Smith, I. M. *Spatial ability, its educational and social significance.* San Diego, Calif.: Knapp, 1964.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A feature model of semantic decisions. *Psychological Review,* 1974, *81,* 214-241.

Spearman, C. *The abilities of man: Their nature and measurement.* New York: Macmillan, 1927.

Stricker, S. *Studien uber die Sprachvorstellungen.* 1880. (Cited in Woodworth, 1938.)

Thurstone, L. L. Primary mental abilities. *Psychometric Monographs.* 1938, *1.*

Toulouse, E. *Review Paris,* 1897, *6,* 88-126. (Cited in Woodworth 1938.)

Tyler, L. E. *The psychology of human differences.* New York: Apleton-Century-Crofts, 1965.

White, K., Sheehan, P. W., & Ashton, R. Imagery assessment: A survey of self-report measures. *Journal of Mental Imagery,* 1977, *1,* 145-170.

Wittgenstein, L. *Philosophical investigations.* New York: Macmillan, 1953.

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

20 Multiple Conceptual Models of a Complex System

Albert L. Stevens

Allan Collins
Bolt Beranek & Newman Inc.

INTRODUCTION

Designers of intelligent computer-aided instructional systems have spent much effort developing techniques for representing knowledge. interpreting student inputs. presenting clever displays. and providing ways to motivate students to interact with the systems. These efforts all reflect important aspects of the problem of providing an environment that facilitates learning. Nevertheless, we believe that most efforts to date have neglected one of the most important aspects of the problem: a deep and thorough analysis of the strategies and knowledge that a skilled teacher uses to communicate a subject matter effectively.

Our efforts have been directed at analyzing the strategies and skills necessary to teach complex topics such as geography, climate, and meteorology. We have found that the skills necessary are indeed complex. A teaching dialogue, rather than following some a priori knowledge structure, is best characterized as a mixture of diagnosis and correction strategies where the tutor probes the student's understanding and uses the surface errors as clues about the deeper misconceptions that they manifest. These diagnosis and correction strategies require knowledge about common errors and their relationship to misconceptions, an understanding of the types of real-world experiential knowledge that students bring to bear on comprehending new problems, and an understanding of the ways that this real-world knowledge can be applied.

In this chapter, we present some of our current analyses and ideas about the teaching process. We briefly review our analyses of the teaching strategies we have observed in dialogues and the goal structure necessary to support them. We present some of the errors that we have observed students make about the causes

of rainfall and show how these can be characterized as arising from deeper misconceptions. In the final section, we describe our ideas about some of the conceptual models we believe are necessary to deal with students learning about rainfall and suggest how they interact to produce understanding.

Teaching Strategies

One of our first goals was to characterize the set of strategies that teachers use in dealing with students' questions and responses. We examined tutorial dialogues that used a Socratic or case method. Based on analyses of dialogues covering several different topic areas, we were able to derive a set of pattern-action rules that account for many of the specific teaching strategies used by the tutors (Collins, 1977). The rules assume a simple knowledge structure that represents the functional dependencies of the domain being taught. For the purposes of the analysis, we assumed that functional knowledge was represented as an and/or graph. The and/or formalism serves basically to differentiate between necessary and sufficient conditions for the various factors taught. For example, rice growing requires three necessary factors: a flooded flat area, fertile soil, and warm temperatures. A flat area is the result of either of two sufficient factors: flat terrain or terracing.

The teaching rules were formulated in terms of a conditional test paired with an action to perform if the test is true. We can illustrate this analysis with two sample rules:

1.   If the student gives as an explanation a factor that is not an immediate cause in the causal chain,
    then ask for the intermediate steps.
2.   If the student gives as an explanation one or more factors that are not necessary,
    then formulate a general rule by asserting that the factor is necessary, and ask the student if the rule is true.

The analysis in Collins (1977) consists of 24 rules. This set captures much of the local structure of teaching dialogues but fails to deal with global structure. As we pointed out in that paper, characterizing the structure of the global interactions requires additional layers of theory.

Goal Structure

In order to characterize the global structure of teaching dialogues, we have conducted additional dialogues. In these, we attempted to open another channel into the tutor's thinking by isolating the tutor from the student, having them communicate over linked terminals, and taking a verbal protocol from the tutor.

In the protocol, we asked the tutor to comment on two aspects of the process: (1) what he thought the student knew, or didn't know, based on the student's response; and (2) why he responded to the student in the way he did. This technique provides insights into how the tutor organizes the knowledge taught, how the tutor develops a model of the student, and how these two factors influence the tutor's choice of questions and responses to the student.

We developed the outlines of a theory of tutors' goal structures. The goal structure we derived is summarized in Table 20.1. The top-level goals are: (1) Refine the student's causal model, and (2) refine the student's procedures for applying the model. These directly govern the selection of cases. As the student's knowledge becomes more refined, moving from an understanding of first-order factors to higher-order factors, cases are selected that are exemplary of the factors the tutor is trying to teach. As the student's predictive ability becomes refined, cases are selected that are progressively more novel and complex, taxing the student's predictive ability more and more.

The process of achieving these top-level goals involves two types of subgoals: diagnosis and correction. Both of these subgoals govern the selection of basic strategies.

The purpose of diagnosis is to discover gaps and misconceptions in the student's knowledge. This generally requires that the tutor probe the student by

TABLE 20.1
Outline of a Socratic Tutor's Goal Structure[a]

Goals	Manifestations
Refine the student's causal model, moving from first- to nth-order factors.	Case selection rules: Select cases that are exemplary of the relevent factor
Refine the student's procedures for applying the causal model to novel cases	Case selection rules: Select less familiar cases, exemplary of new factors
Subgoals	
Diagnose the student's "bugs" (i.e., the difference between the student's knowledge and the tutor's knowledge)	Ask-for factor rules. Prediction rules Entrapment rules Probe-reasoning-strategy rules
Correct the diagnosed bugs	Inform-student rules Missing-factor rules Forming hypotheses rules Testing hypotheses rules Information collection rules

[a] The manifestations refer to the rules described in Collins (1976) and in Stevens and Collins (1977).

asking for relevant factors, by requiring the student to make predictions about carefully selected cases, and by trying to entrap the student into making incorrect predictions. It is clear from our analysis of human dialogues that diagnosis cannot be completely characterized in terms of a simple mapping between surface errors and underlying misconceptions. Rather, the process involves sophisticated use of a student model and knowledge about common misconceptions in order to simulate the student's reasoning processes and to pinpoint the underlying misconceptions or missing information. In some situations, a single answer may reveal a whole set of misconceptions, whereas in other cases, the tutor must carefully probe the student, testing alternative hypotheses.

Typically, when a misconception is diagnosed, the tutor attempts to correct it. This may require a single statement for simple factual errors or an extended dialogue to correct problems in the student's causal model. In Stevens and Collins (1977), we illustrate the application of this goal structure model by using it to analyze a tutorial dialogue.

Our outline of goal structure is relatively general and probably can be applied to many different knowledge domains and tutorial interactions. However, in order to specify it in detail, we need to know what the misconceptions are, how they can be represented. how they are diagnosed from errors, and how they can be corrected.

Conceptual Bugs

In a sense, the previous two sections describe preliminaries to some of the hardest problems that must be faced. What are the conceptual bugs? What knowledge and knowledge representation are necessary to support the basic teaching strategies and the global goal structure? What knowledge and knowledge representation are necessary to correct diagnosed bugs?

We have recently completed an experiment to examine the misconceptions that occur in understanding rainfall (Stevens, Collins, & Goldin, 1979). We compiled a systematic set of questions by generating an and/or graph representation for the causes of heavy rainfall. For each node in the graph, we generated a question that asked what the prior factors were and a question that asked what the subsequent factors were. This resulted in 32 questions that we assembled into a test booklet and presented to eight students. Some examples are: "How is the moisture content of the air related to heavy rainfall?" "What role does rising air play in causing rainfall?" "What causes evaporation?" At the top of the test, we included a paragraph that described what we meant by heavy rainfall and instructed the students to answer all questions in the context of that paragraph. We asked the students to answer all questions, even if they felt they were just guessing. because in previous work, we have found that students often know a good deal more than they think they do.

To analyze this experiment, we first tabulated all responses that we judged to be errors. We subsequently analyzed these errors by classifying them according

to a basic set of bugs. Development of the set of bug types occurred in combination with the error analysis. Our analysis revealed two points of interest: (1) A particular conceptual bug is often shared by several students; and (2) a particular conceptual bug is often manifested in different ways. For example, one of the most frequent bugs is the "cooling-by-contact" bug that occurs for six of the eight students. Some verbatim examples of manifestations of this bug are:

1. "Cold air masses cool warm air masses when they collide."
2. "Winds cause air to cool."

TABLE 20.2
The Set of Observed Misconceptions

Misconception	Number of Subjects	Example
1. Cooling by contact	6	"Mountains cause condensation because cold land touching air causes condensation."
2. Heating by radiation	6	"The sun warms the air."
3. Small moisture source	5	"A 12-by-12-by-10-foot pond is enough to cause rainfall."
4. Rising causes increased pressure.	3	"Rising air makes the moist air rise, pressure increases...."
5. Absorption by expansion	3	"... decrease in pressure causes water molecules to expand, causes evaporation."
6. Heating by contact	3	"... land warms the air at night."
7. Squeezing causes condensation.	2	"Putting pressure on air masses causes condensation."
8. Temperature of water irrelevant for evaporation	2	"Temperature of water is unrelated to evaporation."
9. Temperature differential causes evaporation.	2	"Air has to be cooler than the body of water for evaporation to occur."
10. Insufficient warming of water	2	"A current can be warm because it comes from a warm source of water—for example, a lake which is warm."
11. Heating causes condensation.	1	"Air warming up causes rainfall."
12. Winds cause pressure increases.	1	"Winds are forceful and cause various air pressures."
13. Cooling causes evaporation.	1	"When a body of water is cold, it evaporates."
14. Rising results in pressure equalization.	1	"Air that is warmer is expanded and has less pressure. It rises until its pressure is equal to surrounding air."
15. Cooling causes air to rise.	1	"Cooling causes air to rise."
16. Evaporation causes air to rise.	1	"Evaporation causes air to rise."

3. "Mountains cause condensation because cold land touching warm air causes condensation."
4. "Cold fronts, wind, snow, and rain cause air to cool."
5. "Cold air masses cool the clouds so the rain falls."

None of the foregoing types of cooling are of any consequence in causing heavy rainfall. The type of cooling necessary occurs when an air mass is forced to rise. The rising results in expansion and energy loss.

We identified 16 different conceptual-bug types from this analysis. Table 20.2 shows these 16 bugs in order of frequency. Using these 16 misconceptions, we were able to account for 58% of the errors. Many of the remaining errors are factual errors—for example, "Heavy rainfall occurs only in warm areas"; or naming errors—for example, "When water evaporates, it turns to steam." (Heavy rainfall occurs in many cool and cold areas; the standard term in meteorology for the product of evaporation is *water vapor*.)

Note that the mapping between the manifestations and the bugs is often not simple. There are sometimes obvious surface clues; for example, the sun as an agent of "warming air" indicates the "heating-by-radiation" bug. Other cases require a more subtle analysis; for example, detecting the "small-moisture-source" bug requires knowledge about relative sizes of bodies of water.

MODELS

We believe that the bugs we have isolated are still rather shallow, reflecting even deeper levels of misconceptions in students' knowledge. The major reason for this is that the bugs themselves seem to form patterns, and the patterns seem best explained as the result of deeper problems in the students' knowledge. In this section, we discuss some of the issues that we see as important in understanding where bugs really come from and what is necessary to characterize adequately the knowledge a student must acquire to understand a complex system. The view we propose is that people maintain multiple, procedural representations, which we call *models* (Collins, Brown, & Larkin, in press; Stevens & Collins, 1977). We often refer to these models as simulations (Brown & Burton, 1975), but the models are not complete simulations of the world. Simulation models only make it possible to represent certain properties of the world. The properties represented may be both incomplete and incorrect, but by knowing how they interact, it is possible to "run" the model under different conditions to examine the consequences. Thus, a simulation model is like a motion picture that preserves selected properties of the world.

There are three themes that run through our discussion of models. These themes have strong implications for the design of expert CAI systems. The first theme is that any model can be more or less sophisticated, and learning is largely

a process of refining models so that they correspond better with the real world. We have observed several kinds of refinement:

1. *Adding parts to a model:* A model might be refined by adding different parts to it. For example, if molecules of water are represented as billiard balls bouncing around in a container, the surface tension of the water might be added to the model by representing it as a partially reflecting mirror.

2. *Replacing parts of a model:* A model might be refined by replacing one part of the model with another part. For example, the partial-mirror model of surface tension might be replaced by a view that surface tension results from the unbalanced forces of molecular attraction at the surface of the water.

3. *Deleting parts of a model:* A model might be refined by removing irrelevant parts. For example, a functional model of evaporation that includes the temperature of the heat source could be refined by deleting this aspect of the model.

4. *Generalizing parts of the model:* A model might be refined by generalizing from particular cases. For example, a model of how the Gulf Stream affects rainfall in Europe might be generalized to how currents flow around a rotating sphere and affect landmasses on the sphere.

5. *Differentiating parts of the model:* A model might be refined by breaking down parts of the model into subcomponents. For example, a simple functional model (see Fig. 20.1) might be further differentiated to specify component subprocesses.

The second theme is that models provide the power to consider alternative possibilities and to derive predictions about novel situations. It is possible to look at alternative situations by running the model with different values assigned to its variables. To make predictions, it is often necessary to choose critical values for particular variables in order to determine what the most likely outcomes are and what are the boundary conditions for which the model holds. Thus, the power of models derives from the ability to run them under different assumptions.

The third theme is that students' underlying misconceptions derive from simplifications or distortions in their models. We show how some of the rainfall misconceptions described earlier come out of incorrect underlying models. We think it is possible to counteract some misconceptions by checking results found in one model against another model. Learning how to use different models and to map between them may, in fact, be one of the most important aspects of understanding complex systems.

Models of the Weather System

We can illustrate these notions with four models that we have observed people use in understanding meteorological processes. Two of these models are con-

cerned specifically with evaporation processes. The first of these we call a simulation model of evaporation and the second, a functional model of evaporation. The other two models include evaporation processes as local aspects of more global processes. The third model we call a water-cycle model and the fourth model, a climate model. We illustrate rudimentary forms of these models that we have observed people use and have seen in textbooks. We also provide more sophisticated versions of each model, though of course the sophisticated versions are not completely correct either.

The notion of simulation of the weather can best be understood in terms of a simulation game, such as the Civil War game marketed by Avalon Hill. In the Civil War game, one player represents the North, and the other player represents the South. It is a game of attack and strategy much like Risk, Diplomacy, or even chess. The game consists of a board, playing pieces, and a 17-page booklet of rules. The board and pieces represent the state of a simulated war at a particular moment in time. The rules embody many of the constraints that existed physically and politically at the time of the American Civil War. For example, the rules allow supplies to be moved rapidly along rivers and railroads. This it is clear why Vicksburg figured in an important battle; it is located where a major railroad line crosses the Mississippi River. Furthermore, it is clear why the North attacked along the east coast and the Mississippi rather than through the Appalachian mountains. The rules allow troops and supplies to move through the Appalachians at only one-tenth the rate allowed through other parts of the region.

Given such a simulation game, it is possible to consider how likely it was that the North would win the Civil War. The answer is given in terms of the frequency with which the North wins any game played under the rules. To evaluate such a frequency, it is necessary to consider a set of critical cases. These critical cases must be constructed by examining what happens when the North and South apply different general strategies—for example, when the North centers its strategies around a naval attack, a western campaign, an Appalachian campaign, or an east-coast campaign. In fact, it turns out that the North usually wins.

It is also possible to understand how people process hypothetical questions. For example, we can consider what would have happened if the North had invaded the South through the Appalachians instead of along the east coast and down the Mississippi. The answer comes from characterizing the set of games that are played when the North invades through the Appalachians. For example, a characterization of those games might be that the South's chances improve dramatically and that the winner depends on certain tactical decisions made in any battles that take place in the Appalachians. In fact, because movement is so difficult through the mountains, the South has ample time to anticipate and counter any move by the North, so the South usually wins those games.

These examples illustrate some of the potential power inherent in the simulation approach. Simulations do not represent every possible situation or all aspects of the world (in the Civil War game, there is no provision for the assassination

of Jefferson Davis or for the invention of the airplane), and a large amount of information is necessary to accomplish the simulation; but simulation models do enable one to test out the potential consequences of varying certain aspects of the real world.

A Simulation Model of Evaporation

One possible model of the evaporation process views air molecules as billiard balls. In the rudimentary version of this model, the water particles are thought of as billiard balls bouncing around, hitting each other, and sometimes flying out of the water into the air. One aspect of this model that one might notice is that particles flying out of the water come from the area of the water nearest the surface. The effect of temperature in this model is to speed up the rate at which the billiard balls move. As the particles bounce around, sometimes those near the surface fly off. As the particles are sped up by increasing their temperatures, the whole process speeds up, and so more fly out of the container in a given period of time. Thus, with these few simple local properties, a person can run the model and derive certain consequences—for example, that water evaporates from the area near the surface and that warmer temperatures result in faster evaporation.

Note that even for this rudimentary model, our description is only a very rough approximation to an actual model. It does not explicate the set of laws, processes, and control structures that enable the model to be run under different conditions. For example, the laws governing movement and collision of particles must be internalized in the model, so that when run, the proper consequences of differences like particle speed can be derived.

A more sophisticated version of the model may incorporate the notion of molecular attraction. Molecular attraction can be seen as a force that pulls the billiard balls closer together. Thus, there is a constant pull between the motion of the billiard balls trying to move them apart and the attractive forces trying to bring them together. When the motion is small, the attractive forces can hold the billiard balls together. This corresponds to the liquid state of water. As the motions increase, they overcome the forces of molecular attraction, and the billiard balls fly apart. This corresponds to the vapor state of water. Note that because the amount of motion is an average across all molecules, some will be moving faster than others. So at any time, some molecules will be moving rapidly enough to break free of their neighbors. However, because these molecules are surrounded by millions of other molecules bouncing around rather slowly, subsequent collisions will slow them down, and they will be captured again. It is only those near the surface that really have a chance to break free from the others, pass into the less densely packed air molecules, and remain nonliquid.

The concepts in this model can be used to infer and understand additional properties of water. For example, molecular attraction explains surface tension as the result of the unbalanced forces of attraction that occur near the boundary

between the water and the air. At the boundary, there is a net pull inward, compressing the molecules closer together.

We can illustrate some of the power of this more sophisticated model by showing how it can be used to deal with three different changes to the basic situation of a standing body of water. The model may not be correct, but at least it allows one to make certain predictions.

The first change is to add a layer of oil to the surface of the water (as the world seems to be doing to its oceans). In the model, the effect of oil is to increase the thickness of the surface barrier, and thus to increase the length of the path for particles passing from the water into the air. Thus, the prediction from the model is that a layer of oil on the water should decrease the evaporation rate.

The second change is to make the water choppy instead of smooth. Choppiness increases the surface area of the body of water and thus increases the surface area for particles to escape through. So choppiness should increase the evaporation rate.

The last change is to add winds. We can add winds to the model in at least two different ways. Because winds increase the choppiness of water, they increase the evaporation rate. They also act to bring new portions of the air mass in contact with the surface of the water. If the air near the surface contains a large number of water molecules, then because they are moving randomly around, a large number will return to the water. If there is a large enough number, there will be as many returning as are leaving, and there will be no net evaporation. Thus, winds blow away the part that is saturated and bring in new parts of the air mass where there is a smaller density of water molecules. Winds thereby again increase the amount of evaporation.

These examples illustrate some of the power of such a model. There is a large amount of knowledge that people must have to construct the model; for example, that temperature is represented as average amount of molecular movement, that average movement is related to individual movements in certain ways, that winds affect both choppiness and mixing of the air mass, and that forces can balance or add together. Any such knowledge that is missing or forgotten is likely to lead to the wrong conclusions. But despite these limitations, such a model gives a person enormous power for making new predictions.

A Functional Model of Evaporation

A rather different perspective on the evaporation process is seen in the functional representation developed by Stevens et al. (1979) to account for people's misconceptions, and in the finite-state-automaton model of Brown, Burton, and Zdybel (1973). This functional perspective describes the input variables and output variables in the functional relationships involved in evaporation.

The upper part of Fig. 20.1 shows a rudimentary form of such a functional model. It is what a person might derive from watching water heating on a stove or evaporating from a dish in the sun. In this rudimentary form of the model, the

amount of evaporation is a function of the amount of heat affecting the water. Thus, if the burner on the stove is turned on high or the sun's rays are particularly hot, more water will evaporate from the container. The person probably would not know the exact functional relationship—just that the rate of evaporation is an increasing function of the amount of heat applied to the water. Different people might construct slightly different versions of this model; for example, they might decide that the amount of evaporation is a function of the temperature of the water. But in any case, they must construct something like the model in Fig. 20.1.

A more sophisticated version of the model might break the process down into different components. The breakdown shown in the bottom part of Fig. 20.1 is approximately what is taught in meteorological texts. In this breakdown, escape rate is seen to be a function of the temperature of the body of water. This is more precise than in the rudimentary version of the model. At the same time, the water-holding capacity of the air is an increasing function of the air temperature. The relative humidity of the air is the ratio of the amount of moisture in the air to its holding capacity. Relative humidity determines the return rate: The higher the humidity, the higher the return rate. The amount of moisture that the air actually absorbs is a simple function of these two output variables—escape rate minus return rate. For the purposes of thinking and talking about evaporation, we can treat these five functional relationships as separate or we can merge them together with the temperature of the water, the temperature of the air, and relative humidity as input variables and the amount of moisture the air absorbs as the output variable.

The differences between the rudimentary version of the model and the more sophisticated version gives some idea of how people can refine this kind of model of a process. In particular, they can learn the controlling variables on different processes; they can learn better the functional dependencies between the input variables and the output variables; they can learn to break the process into its various component subprocesses. Both the texts and the teaching dialogues we have looked at have emphasized these aspects of evaporation. We think this is because the functional viewpoint is critical both for making predictions about the evaporation process and for talking about it.

The mathematical equations for evaporation come from quantizing the functional relationships between the input and output variables of the model, defining the boundary conditions, defining the critical changes of state, and combining these all together. Brown, Burton, and Zdybel (1973) have shown how the cross product of local finite-state automata can be run until equilibrium is reached to determine the effect of any change in an input variable. Stevens et al. (1979) have indicated how large a proportion of teaching dialogues concern the various input variables, output variables, and functional relationships. They further show how many student misconceptions can be represented as perturbations of various parts of such a model.

We should point out that and/or graphs can be derived by instantiating the

RUDIMENTARY VERSION
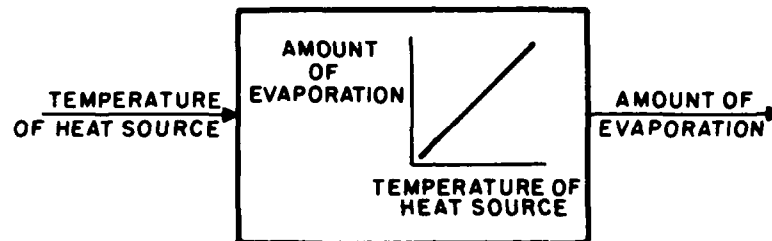
FIG. 20.1.   Two versions of a functional model of evaporation.

variables of such a model. For example, the model in the lower half of Fig. 20.1 can be instantiated to represent a case of high evaporation as shown in Fig. 20.2. In teaching about functional relationships, teachers often talk about input and output variables in these instantiated forms (Collins, 1977).

The simulation models and the functional models give different perspectives on the evaporation process, but it is important to be able to map between the two kinds of models. Undoubtedly, people often have inconsistencies between different models; for example, the rudimentary versions of the two models are inconsistent in that the rate of evaporation is related to the temperature of the water in the simulation model, but to the temperature of the heat source in the functional model. Refinement of models is in part a process of making different models consistent and working out the mappings between them. Thus the more sophisticated versions of each of these models attempt to preserve consistent mappings between them. For example, evaporation rate is treated functionally as an equilibrium process in the functional model, which enables it to map with the process of water molecules entering and leaving the body of water in the simulation model. We suspect that it is important to have models that provide such different perspectives on understanding a process. The simulation model provides an understanding of the mechanism or rationale for the interacting variables described by the functional model. The functional model provides a summary of the physical processes and an indication of the critical boundary conditions for which it is valid.

The Water-Cycle Model

So far we have examined two types of models useful for teaching and understanding evaporation. Evaporation is only one subprocess necessary for rainfall. To understand evaporation in context, there must be other, more global models to tie it in with other processes. One such model, typically taught in high school, is the water-cycle model. It turns out that many of people's misconceptions come from incorrect variants of the water-cycle model (Stevens et al., 1979).

MORE SOPHISTICATED VERSION

FIG. 20.1. *continued*

189
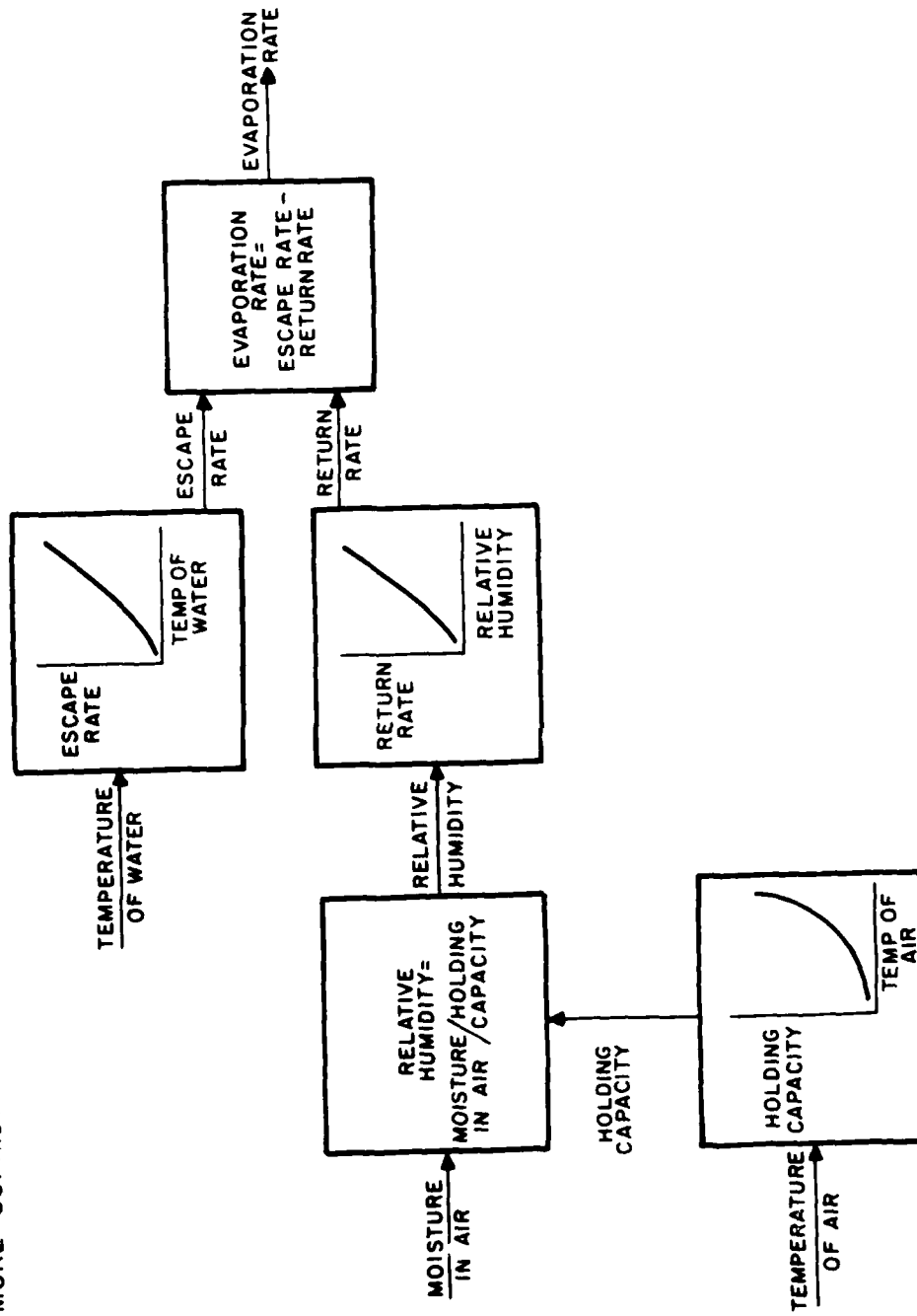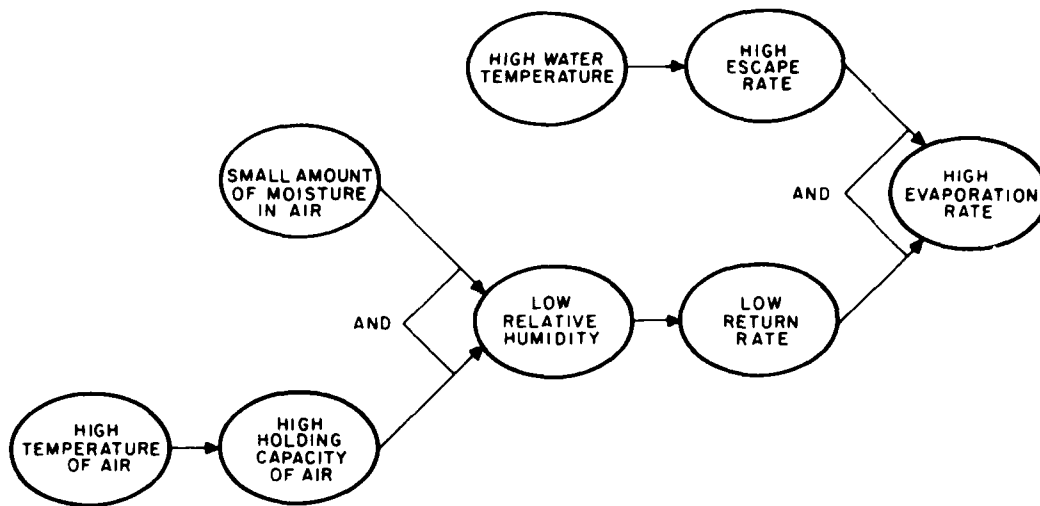
FIG. 20.2. An and/or graph derived from the sophisticated version of the functional model of evaporation.

The top part of Fig. 20.3. illustrates a rudimentary version of the water-cycle model. In this model, moisture evaporates from the ocean, lakes, trees, soil, etc., and rises into the air to form clouds. The clouds move inland, where the moisture falls as precipitation and is carried back to the ocean.

This particular version of the water cycle leads to many students' misconceptions. For example, one student, when asked to name the moisture source for rainfall in the Amazon jungle, answered that it came from the river and the trees. This in part is true, but in fact, the great quantity of moisture comes from the Atlantic Ocean. Another common misconception arising from this rudimentary model is the notion that if a place is close to the ocean, it will have a lot of rainfall. Such a view follows from the proximity of water and land shown in pictures illustrating the model. Another misconception concerns the importance of clouds in the rainfall process. Most meteorology texts treat clouds as a transient step in the condensation process. Novices, however, tend to think of clouds as critical entities in the water cycle. This may follow largely from everyday experience of clouds, but it also relates to the model presented in Fig. 20.3, where clouds are treated as the form moisture assumes in the air. These examples illustrate some of the dangers of teaching oversimplified models.

A more sophisticated version of the water-cycle model is illustrated in the bottom half of Fig. 20.3. Air masses become the critical entities here, rather than clouds. Moisture is seen as evaporating from a large body of water, such as an ocean or large lake. The amount of evaporation depends on air temperature and water temperature. As winds carry the air mass over the body of water, it absorbs more moisture the further it travels. When the air mass moves over land, it can encounter different obstacles. If it encounters a warmer air mass, it tends to go under that air mass. If it encounters a cooler air mass or mountains, it tends to

rise over that obstacle. When an air mass rises, it cools rapidly, leading to precipitation. As the air mass travels over land, it continues to lose moisture as it rises over obstacles and thus has less and less moisture to precipitate. The moisture that is precipitated is carried by rivers back to the bodies of water from which it evaporated.

This particular model enables people to understand many different aspects of the patterns of rainfall in the world. For example, it explains why cold fronts usually bring dry weather; why warm fronts bring rain; why precipitation frequently occurs when two air masses encounter each other; why it tends to be drier farther inland; why mountains have more rainfall than surrounding regions; and so forth. Together with knowledge about geography, this model enables students to make predictions about rainfall patterns in different places. As with the simulation model of evaporation described earlier, the details of a concrete water-cycle simulation are not obvious. The complete model must embody the laws, control structure, and processes in a manner that makes it possible to derive relevant consequences.

As we pointed out earlier, one of the motivations for the notion of models is that students' misconceptions at our level of analysis seem to form patterns. One of the most interesting sets of misconceptions seems to result from a perturbation of the water-cycle model. We call this perturbation the "sponge model" of evaporation and condensation. In it, an air mass is viewed as expanding as it

RUDIMENTARY VERSION
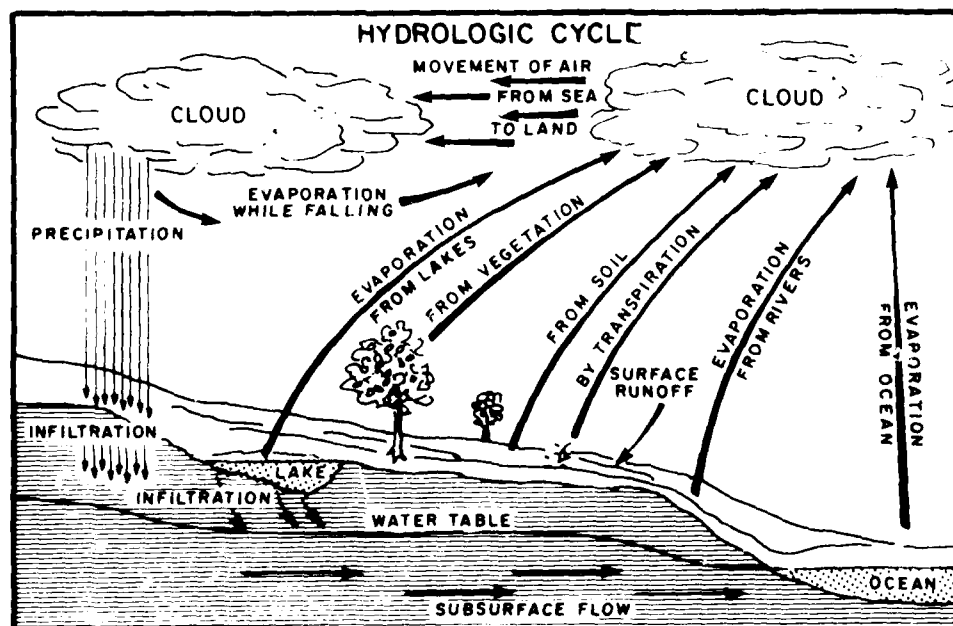
FIG. 20.3.   Two versions of a water-cycle model of rainfall. (Upper figure is redrawn from United States Department of Agriculture, 1950).

MORE SOPHISTICATED VERSION

192

AIR MASS 1

WARM AIR MASS 1

AIR MASS 1

AIR MASS 1

AIR MASS 1

AIR MASS 1

COOL AIR MASS

EVAPORATION

LARGE BODY OF WATER

CONTINENT

RAINFALL
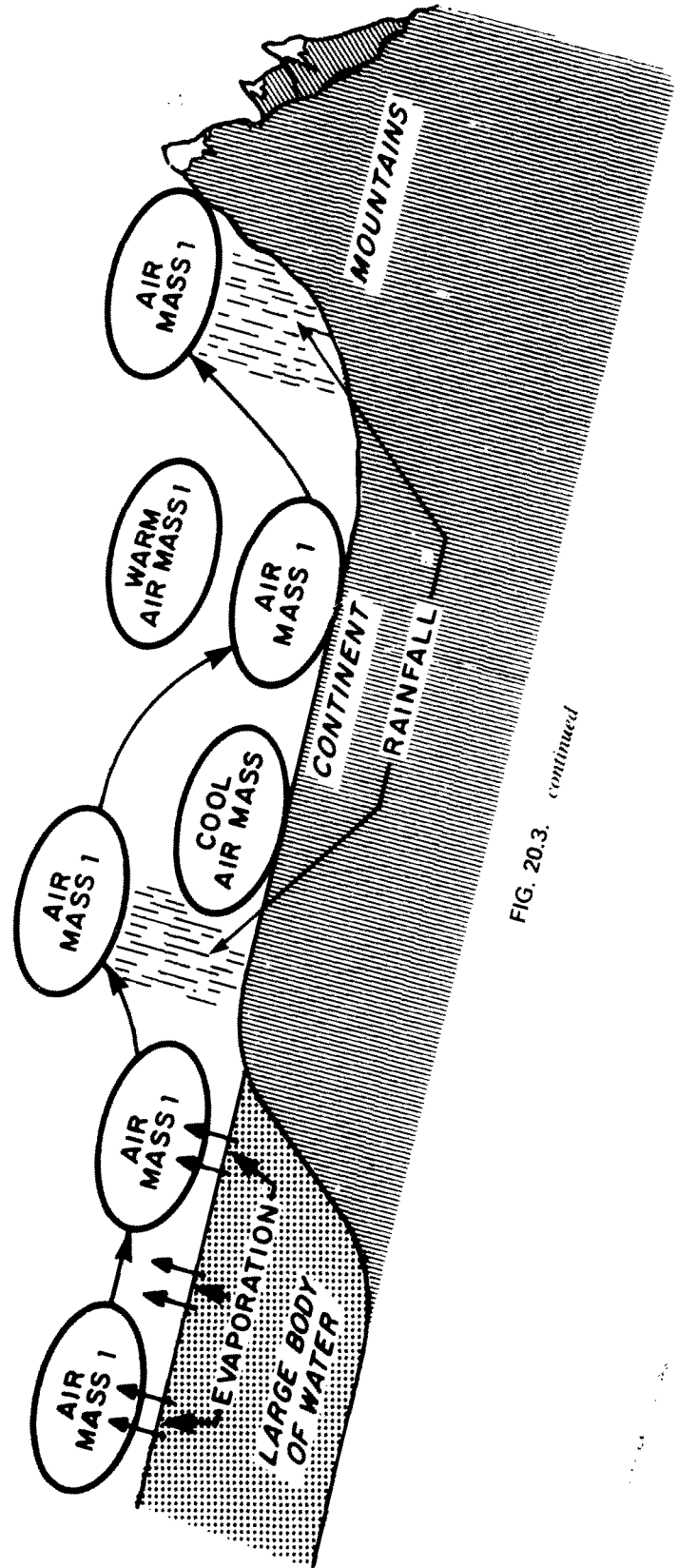
MOUNTAINS

FIG. 20.3. *continued*

absorbs moisture out of the body of water. When it comes in contact with other air masses or mountains, the water is squeezed out of the air mass by the pressure from the collision. The sponge model makes sense of some aspects of the process, but it leads to serious misconceptions such as ignoring the effects of temperature on condensation. One of the important design goals for an adequate teaching system is that it recognize these incorrect models from the patterns of misconceptions that students show.

Finally, we want to point out the relationship of the water-cycle model to the kind of script-structured knowledge that was used in the original Why system (Stevens & Collins, 1977; Stevens et al., 1979) and that is emphasized as important for understanding everyday phenomena, such as going to a restaurant (Schank & Abelson, 1977), a birthday party (Minsky, 1975), or a grocery store (Charniak, 1975). A script represents certain of the critical events that occur in any process. For the water cycle, a script might include the moisture evaporating from the water into the air, being carried over the land, rising, cooling, condensing, and finally being precipitated. A script, then, consists of a set of snapshots taken at different times during the process. Wherever the process can take different paths, depending on events in the world (such as what the opponent does in the Civil War game), a script must break apart into a lattice or tree structure. But scripts inevitably sacrifice much of the inherent power in a simulation model. When people talk about their models, they inevitably describe the critical events that occur in them. Hence, they seem to be talking about scripts they have in their heads. However, we would argue that, in fact, they may be merely describing critical events—for example, events associated with the changes of state that occur when they run their model.

A Model of Climates

The final model we want to describe involves the way water and air currents travel around the world, and what happens when they encounter different landmasses. This model parallels the water-cycle model, but it presents the events from a geographical perspective rather than a meteorological perspective.

Figure 20.4 illustrates a rudimentary version of this model. The Gulf Stream is depicted as following the coast of North America and then crossing the Atlantic toward England and Europe. As the current encounters land, it turns south along the continental border with parts going north of the continent and into the North Sea. The winds carry the moisture-laden air inland over Europe. This rudimentary model is essentially correct, but it contains very little predictive power.

A more sophisticated version is shown in the bottom half of Fig. 20.4. This is the model contained in college geography texts (Hoyt, 1973; James, 1966). It shows the pattern of ocean and air currents as they encounter a hypothetical continent. Driven by the Coriolis effect from the Earth's rotation, currents travel

RUDIMENTARY VERSION
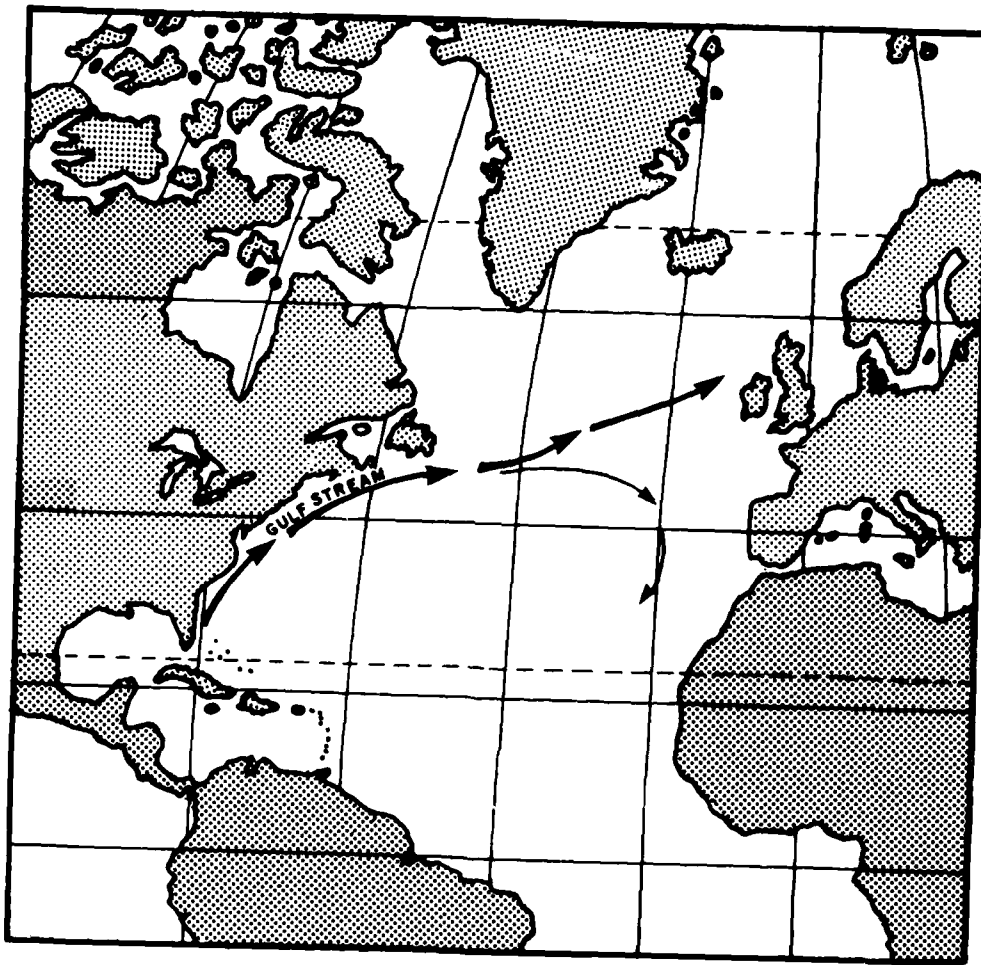
FIG. 20.4.   Two versions of a geographical model of rainfall.

westward along the equator. Ocean currents turn poleward at the eastern edge of any continent and eventually form the prevailing westerlies that occur at approximately 50 to 60 degrees latitude. The circuit is completed by currents running toward the equator along the western edge of a continent. Heavy rain occurs where the ocean currents encounter land. Dry lands occur along the western edge of continents where there is a cold current offshore. The model can be much more complicated than this, involving the movement of high-pressure and low-pressure centers seasonally, but this provides the basic geographical model.

   This basic model can be derived from generalization of specific cases, such as the Gulf Stream model. With the generalization comes genuine predictive power. One could, for example, consider the effects of putting down continents of different sizes and shapes at different places in the South Pacific. The effects on

MORE SOPHISTICATED VERSION

FIG. 20.4. *continued*

195
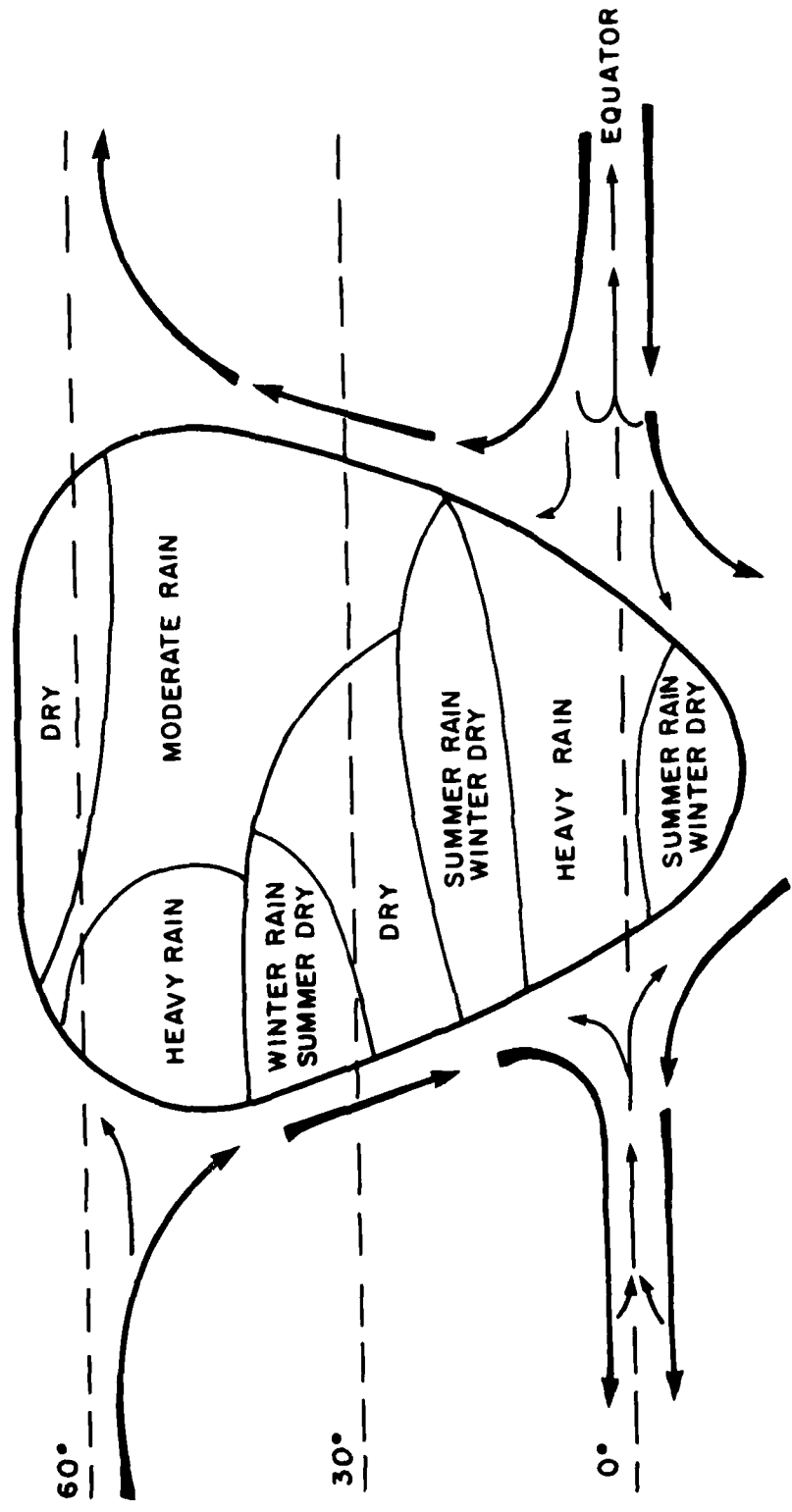
Australia would be minimal, but the effects on South America would be much greater by affecting the landfall of the prevailing westerlies. Given knowledge about mountain ranges, it is possible to make quite accurate predictions about rainfall patterns in the proposed South Pacific continent.

CONCLUSION

At one level, this discussion of models is obvious. Scientists will agree that they view the world from different perspectives, that they alternate between perspectives depending on which view is appropriate for the problem at hand, that they often check a conclusion derived from one model by testing it against another model, and so forth. So this chapter really is arguing for a position quite close to the commonsense view that scientists already have about their own knowledge.

At another level, however, the proposal that knowledge about complex systems must be represented in multiple models has radical implications both for representing knowledge in intelligent CAI systems and for education generally. We briefly indicate a few of those implications.

The major implication for intelligent CAI systems is that it is not sufficient to build the system based on a single perspective of the domain, nor exclusively to use static representations such as and/or graphs or scripts. Our proposal is that expert systems need multiple models that can be used generatively to test out novel hypotheses and make predictions about new situations. Furthermore, they must have specific strategies that determine when to invoke one model and when another, and how to map back and forth between models. In sum, representation of expert knowledge must be further removed from the surface forms in which people talk than most current systems contemplate. Unfortunately, this makes many aspects of building expert systems more difficult.

The implications for education are equally profound. This view suggests that multiple models should be taught explicitly as alternative points of view about a topic. The emphasis should be on the kinds of situations and problems for which each model is applicable, and on how to apply them to solve different types of novel problems. At the same time, students should learn the limitations of each model and how to test out a solution derived from one model against another. Students might also be taught how various distortions of a model lead to different misconceptions, and how any model can be systematically refined to increase its predictive accuracy.

ACKNOWLEDGMENTS

REFERENCES

Brown, J. S., & Burton, R. Multiple representations of knowledge for tutorial reasoning. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Brown, J. S., Burton, R., & Zdybel, F. A model-driven question answering system for mixed-initiative computer-assisted instruction. *IEEE Transactions on Systems, Man and Cybernetics,* 1973, *3,* 248 257.

Charniak, E. A partial taxonomy of knowledge about actions. In *Advance papers of the Fourth International Conference on Artificial Intelligence.* Tbilisi, USSR: 1975.

Collins, A. Processes in acquiring knowledge. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge.* Hillsadale, N.J.: Lawrence Erlbaum Associates, 1977.

Collins, A., Brown, J. S., & Larkin, K. Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension.* Hillsdale, N.J.: Lawrence Erlbaum Associates, in press.

Hoyt, J. *Man and the earth.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

James, P. *A geography of man.* Waltham, Mass.: Xerox College Publishing, 1966.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill, 1975.

Schank, R., & Abelson, R. *Scripts, plans, goals and understanding.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Stevens, A., & Collins, A. The goal structure of a Socratic tutor. In *Proceedings of the Association for Computing Machinery Annual Conference,* 1977. (Also available as BBN Rep. No. 3518 from Bolt Beranek and Newman Inc., Cambridge, Mass., 02138.)

Stevens, A., Collins, A., & Goldin, S. Misconceptions in students' understanding. *International Journal of Man-Machine Studies,* 1979, *11,* 145-156.

United States Department of Agriculture. *The Yearbook of Agriculture, 1949.* Washington, D.C.: Author, 1950.

21 Complex Learning Processes

John R. Anderson

Paul J. Kline

Charles M. Beasley, Jr.
Yale University

ABSTRACT

This chapter describes the ACT theory of the learning of procedures. ACT is a computer simulation program that uses a propositional network to represent knowledge of general facts and a set of productions (condition⇒action rules) to represent knowledge of procedures. There are currently four different mechanisms by which ACT can make additions and modifications to its set of productions as required for procedural learning: designation, strengthening, generalization, and discrimination. Designation refers to the ability of productions to call for the creation of new productions. Strengthening a production may have important consequences for performance, because a production's strength determines the amount of system resources that will be allocated to its processing. Finally, generalization and discrimination refer to complementary processes that produce better performance by either extending or restricting the range of situations in which a production will apply. Each of these four mechanisms is discussed in detail and related to the available psychological data on procedural learning. The small-scale simulations of learning provided as examples are drawn from the domains of language processing and computer programming. Our ultimate goal is for ACT to learn the complex procedures required in such domains.

INTRODUCTION

We are interested in understanding learning. For many years, learning theory was practically synonymous with experimental psychology; however, its boundaries

have shrunk to such an extent that they barely overlap at all with those of modern cognitive psychology. Cognitive psychologists, by and large, concern themselves with a detailed analysis of the mechanisms that underlie adult human intelligence. This analysis has gone on too long without adequate attention to the question of how these complex mechanisms could be acquired. In an attempt to answer this question, we have adopted one of the methodological approaches of modern cognitive psychology: Results of detailed experimental analyses of cognitive behaviors are elaborated into a computer simulation of those behaviors. The simulation program provides new predictions for a further experimental testing, whose outcome is then used to modify the simulation, and the whole process then repeats itself.

Our computer simulation is called ACT; this chapter describes its learning processes as well as describing some initial contact between empirical data and predictions derived from these learning processes. The ACT system embodies the extremely powerful thesis that a single set of learning processes underlies the whole gamut of human learning—from children learning their first language by hearing examples of adult speech, to adults learning to program a computer by reading textbook instructions. If we can show that ACT's learning processes can acquire some of the cognitive skills required to master these two very different domains, we will have made a beginning toward establishing this bold thesis. The failure of traditional learning theory invites skepticism of the claim that a single set of processes underlies all learning. However, because the consequences of such a thesis, if true, are so important, and because it is now possible to construct more sophisticated theories of learning processes by the use of computer simulation, another attempt to establish this thesis seems appropriate.

Chomsky (1965) and others have advocated the opposing point of view that special mechanisms are required to learn language. In fact, an earlier simulation program, LAS, developed by the first author to model language acquisition (Anderson, 1974, 1975, 1977, 1978), used learning mechanisms that were not applicable to other cognitive skills. However, it now appears that LAS's learning mechanisms can be seen as manifestations of more general learning mechanisms.

There were a number of inadequacies in the LAS program. (These are reviewed in detail in Anderson, 1978.) LAS was unable to make discriminations, to correct errors, to deal with nonhierarchical aspects of language, or to account for the gradualness of human learning. There were also reasons for doubting that LAS was properly modeling the procedural aspects of language or that it was properly modeling human limitations in language learning and performance. In one way or another, each of these problems could have been handled by additions to the LAS theory—but at great cost to the overall parsimony and elegance of that theory. It seemed that a more elegant resolution was possible only by stepping back to a more general learning approach. We expect that ACT will reproduce many of LAS's learning feats; however, it will do so in a way that will naturally extend to the many problems LAS could not handle. Thus, LAS established what

could be done by a set of learning mechanisms, and ACT is an attempt to generalize what we have learned from LAS.

The organization of the rest of this chapter is as follows: First there is a short description of the nonlearning aspects of the ACT production system. Following this, there are sections discussing each of the three ways the system has of forming new productions: designation, generalization, and discrimination. The next topic discussed is production strength, which serves to integrate the new productions into the behavior of the system to produce better performance. The final sections contain speculations on the origin of designating productions and some directions for future work.

THE ACT PRODUCTION SYSTEM[1]

The ACT production system can be seen as a considerable extension and modification of the production systems developed at Carnegie-Mellon (Newell, 1972, 1973; Rychener & Newell, 1978). ACT represents its knowledge of general facts in a propositional network. This propositional network uses nodes to represent ideas (roughly) and labeled links, which connect nodes, to represent various types of associations between ideas. Information is organized into propositional units where each proposition is a tree interassociating a number of nodes. Although the network aspects of this representation are important for such ACT processes as spreading activation, for most purposes ACT's data base may be thought of as consisting simply of a set of propositions. For example, ACT might represent the addition problem 32 + 18 by the set of propositions:[2]

    (ADD 32 18)
    (BEGINS 32 2)
    (AFTER 2 3)
    (ENDS 32 3)
    (BEGINS 18 8)
    (AFTER 8 1)
    (ENDS 18 1)

ACT represents its procedural knowledge as a set of productions—that is, (condition ⟹ action) rules. The condition is an abstract description of a set of propositions. If propositions can be found in the data base that satisfy this

---

[1]The version of ACT described in this chapter is called ACTF. Earlier publications (e.g., Anderson, 1976) described the previous version, ACTE.

[2]To simplify the exposition, a relation-argument syntax for propositions is used in this chapter. This is a departure from the actual ACTF syntax, which relies on infix operators such as * and OF as described in previous publications (see Anderson, 1976; Anderson, Kline, & Lewis, 1977). Also in the interests of simplicity, type-token distinctions required to represent several occurrences (tokens) of the same digit (type) in an addition problem properly are being ignored here and throughout this chapter.

abstract description, the production will perform its action. Actions can both add to the contents of the data base and cause the system to emit observable responses.

Propositions that are added to the data base are treated as sources of *activation*. The total amount of activation given to a source is divided up among all the terms contained in that proposition and then spread from them out over the links in the propositional network to activate other propositions containing these same terms. The activation of these propositions causes them to be treated as sources in turn (but with a reduced amount of activation), and the process continues until the activation spread to a proposition is less than the amount the system requires to consider a node active at all. The amount of activation that will accumulate at any given node will depend on the number, strength, and directness of its connections to the original sources of activation.[3]

ACT productions can only have their conditions satisfied by active propositions—a requirement that insures that the system will be most responsive to changes in the contents of its data base. ACT's basic control structure is an iteration through successive *cycles*, where each cycle consists of a production-selection phase followed by an execution phase. On each cycle an APPLYLIST is computed that is a probabilistically defined subset of all of the productions whose conditions are satisfied by active propositions. The probability that a production will be placed on the APPLYLIST depends on the *strength (s)* of that production relative to the sum (S) of the strengths of all the productions whose conditions mention active nodes; that is, this probability is proportional to s/S. Discussion of the process of assigning a strength to a production is postponed until a later section; all that needs to be said here is that this strength reflects just how successful past applications of this production have been. Thus one component of the production-selection phase consists of choosing out of all the productions that could apply those that are most likely to apply successfully. Further discussion of the details of production selection and execution is best conducted in the context of an example.

Sample Production System

Table 21.1 presents a set of productions for adding two numbers.[4] Since it is difficult to grasp the flow of control among the productions in Table 21.1, this information is presented diagrammatically in Fig. 21.1, which may be useful in

---

[3]No discussion of link strength is provided here. Similarly, the whole question of decay of activation is being ignored. A more complete treatment of spreading activation can be found in Anderson (1976, Chap. 8), although the current ACTF implementation of the spreading activation process differs substantially from the implementation discussed there.

[4]The productions presented in this chapter are translations of the formal syntax of the implemented productions into (hopefully) more readable prose. The reader interested in the details may write to the authors to request listings of the implemented versions and examples of their operation.

TABLE 21.1
A Set of Productions for Adding Two Numbers

P1: IF the goal is to add LVnumber1 and LVnumber2
    and LVnumber1 begins with LVdigit1
    and LVnumber2 begins with LVdigit2
  THEN add LVdigit1 and LVdigit2
    and set GVdigit1 to LVdigit1
    and set GVdigit2 to LVdigit2

P2: IF GVdigit1 and GVdigit2 are being added
    and LVsum is the sum of GVdigit1 and GVdigit2
  THEN set GVsum to LVsum

P3: IF GVdigit1 and GVdigit2 are being added
    and LVsum is the sum of GVdigit1 and GVdigit2
    and there is a carry
  THEN set GVsum to LVsum and note that a
    carry must be added to GVsum

P4: IF GVsum has a value
    and there is no carry
    and GVsum is not > 9
  THEN write GVsum
    and go to the next column

P5: IF GVsum has a value
    and there is a carry
    and LVsum1 is the sum of GVsum plus 1
    and LVsum1 is not > 9
  THEN write LVsum1
    and go to the next column

P6: IF GVsum has a value
    and GVsum > 9
    and GVsum is the sum of LVdigit3 and 10
    and there is no carry
  THEN write LVdigit3
    and go to the next column with a carry

P7: IF GVsum has a value
    and there is a carry
    and GVsum > 8
    and GVsum is the sum of LVdigit3 and 9
  THEN write LVdigit3
    and go to the next column with a carry

P8: IF sent to the next column with no carry
    and there is a digit, LVdigit3, after GVdigit1
    and a digit, LVdigit4, after GVdigit2
  THEN set GVdigit1 to LVdigit3
    and set GVdigit2 to LVdigit4
    and add GVdigit1 and GVdigit2

P9: IF sent to the next column with no carry
    and there is a digit, LVdigit3, after GVdigit1
    and there is no digit after GVdigit2
  THEN set GVdigit1 to LVdigit3

(continued)

TABLE 21.1
(continued)

and write GVdigit1
and go to the next column
P10: IF sent to the next column with no carry
and there is a digit, LVdigit4, after GVdigit2
and there is no digit after GVdigit1
THEN set GVdigit2 to LVdigit4
and write GVdigit2
and go to the next column
P11: IF sent to the next column
and there is no digit after GVdigit1
and there is no digit after GVdigit2
THEN problem completed
P12: IF sent to the next column with a carry
and there is a digit, LVdigit3, after GVdigit1
and a digit, LVdigit4, after GVdigit2
THEN set GVdigit1 to LVdigit3
and set GVdigit2 to LVdigit4
and add GVdigit1 and GVdigit2
and note the carry in the new column
P13: IF sent to the next column with a carry
and there is a digit, LVdigit3, after GVdigit1
and there is no digit after GVdigit2
and LVdigit3 is not = 9
and LVsum is the sum of LVdigit3 and 1
THEN set GVdigit1 to LVdigit3
and write LVsum
and go to the next column
P14: IF sent to the next column with a carry
and there is a digit, LVdigit4, after GVdigit2
and there is no digit after GVdigit1
and LVdigit4 is not = 9
and LVsum is the sum of LVdigit4 and 1
THEN set GVdigit2 to LVdigit4
and write LVsum
and go to the next column
P15: IF sent to the next column with a carry
and there is a digit, LVdigit3, after GVdigit1
and there is no digit after GVdigit2
and LVdigit3 = 9
THEN set GVdigit1 to LVdigit3
and write 0
and go to the next column with a carry
P16: IF sent to the next column with a carry
and there is a digit, LVdigit4, after GVdigit2
and there is no digit after GVdigit1
and LVdigit4 = 9
THEN set GVdigit2 to LVdigit4
and write 0
and go to the next column with a carry

TABLE 21.1
(continued)

P17: IF sent to the next column with a carry
  and there is no digit after GVdigit1
  and there is no digit after GVdigit2
THEN write 1
  and problem completed

---

understanding the discussion of the addition productions that follows. There are a number of notational conventions in this figure: Productions are represented as arrows connecting states represented by circles. Each arrow is labeled by the production it represents. The state circle at the head of an arrow shows the action of the production. The arrows for other productions that need these actions performed in order to apply are shown originating from this state circle. When two or more productions originate from a state circle, additional information from the data base must be examined in order to decide which production should apply. Such additional conditions are represented in diamonds adjacent to the production numbers. The state circle at the tail of a production arrow along with the adjacent diamond totally account for the condition of that production.

Suppose that the addition problem 32 + 18 is in ACT's data base in the format described earlier. Then the condition of production P1 is satisfied by making the following correspondences between elements of the condition and propositions in the data base:

add LVnumber1 and LVnumber2 = (ADD 32 18)
LVnumber1 begins with LVdigit1 = (BEGINS 32 2)
LVnumber2 begins with LVdigit2 = (BEGINS 18 8)

In making these correspondences, the variables LVnumber1, LVnumber2, LVdigit1, and LVdigit2 are bound to the values 32, 18, 2, and 8, respectively. The *LV* prefix indicates that these are local variables and can be bound to anything. They only maintain their binding within the production. Other productions are not constrained to match these variables in the same way. In contrast, there are global variables (*GV* prefix), which, once bound, keep their values in subsequent productions unless explicitly rebound.

The action of P1, *add LVdigit1 and LVdigit2*, becomes, given the values of the variables, an instruction to place the proposition (ADD 2 8) into the data base. The action of P1 also sets global variables to the digits in the first column.

After the execution of P1, the first element of the condition of production P2 is satisfied:

If GVdigit1 and GVdigit2 are being added = (ADD 2 8).

The remaining condition of P2 matches a proposition in the data base about integer addition:

206
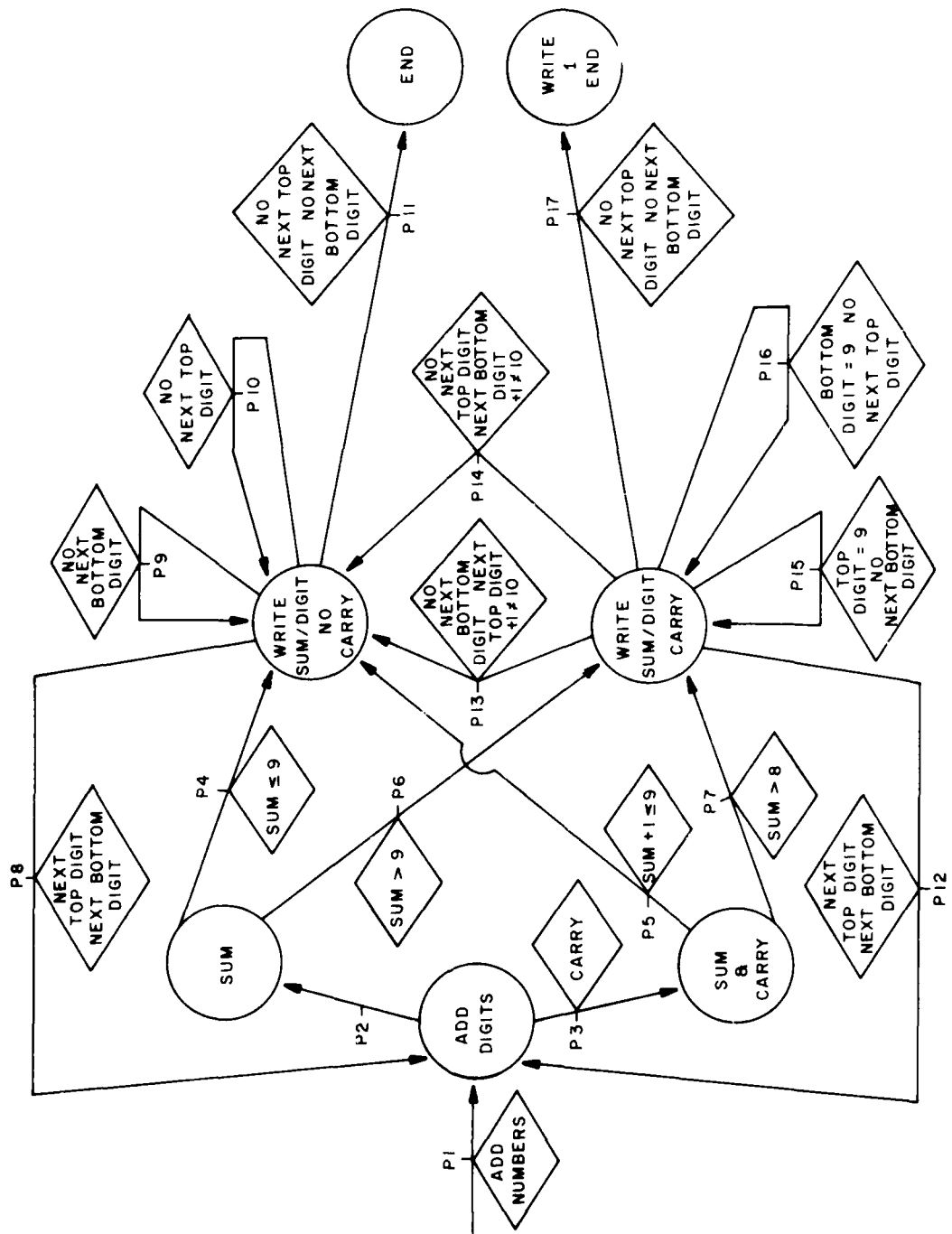
LVsum is the sum of GVdigit1 and GVdigit2 = (10 = 2 + 8)

The action of P2 simply sets the global variable, GVsum, to this sum.

Productions that require that GVsum have a value can now apply. In particular, production P6 is matched as follows:

GVsum > 9 = (10 > 9)

GVsum is the sum of LVdigit3 and 10 = (10 = 0 + 10).

Since this is the first column in the problem, the final requirement of P6,— that there be no proposition in the data base indicating a carry into this column— is obviously satisfied. The action of P6 writes out 0 as the first digit in the answer and places a proposition in the data base, (DO-NEXT 2 8 CARRY), to the effect that this column is finished and a carry should be made into the next column.

It may be worth considering why no other production besides P6 can apply. Production P4 fails because there is a proposition in the data base, (10 > 9), inconsistent with the requirement that *GVsum is not > 9*. Productions P5 and P7 do not apply because there is no carry into the first column. One might wonder why P1 or P2 do not apply again, since their conditions were satisfied once by data base elements that have not been changed. The current version of the ACT production system does not allow production conditions to match twice to exactly the same data-base propositions. This constraint serves to avoid unwanted repetitions of the same productions and the danger of infinite loops.

Production P12 applies next, resetting GVdigit1 to 3 and GVdigit2 to 1 and entering (ADD 3 1) into the data base so that the next column can be added. Production P3 sets GVsum to 4, obtained from the data base proposition (4 = 3 + 1). P3 applies here rather than P2, although the condition of P2 is also satisfied. This is because the condition elements of P2 are a proper subset of those of P3. This principle is referred to as *specificity ordering* in what follows, because it results in more specific productions applying in place of more general ones.

Production P5 adds the carry to GVsum and writes out the second digit of the answer, 5. P11 then applies, noting that the problem is finished.

This example illustrates a number of important features of the ACT production system.

1. Individual productions act on the information in long-term memory. They communicate with one another by entering information into memory and setting global variables.

2. Productions tend to apply in sequences where one production applies after another has entered some element into the data base. Thus the action of one production can help evoke other productions.

FIG. 21.1.   *(Opposite page)*   The flow of control among the productions in Table 21.1.

3. The condition of a production describes an abstract pattern of propositions in the data base. The more propositions a condition requires in its pattern, the more difficult it is to satisfy that condition. Similarly, the more a condition relies on constants instead of variables to describe its pattern, the more difficult it is to satisfy that condition.

PRODUCTION DESIGNATION

ACT needs the ability to augment its set of productions with new productions. For this reason, productions can designate the construction of other productions in their actions in much the same way that they designate the construction of memory structure. Production designation is an important means by which ACT learns procedural skills.

Encoding of Procedural Instructions

As a first example of procedural learning, let us consider how production designation can be used to assimilate the lessons provided by instruction. Consider how ACT might assimilate the following rules defining various types of LISP expressions (adapted from the second chapter of Weissman, 1967):

1. If an expression is a number, it is an atom.
2. If an expression is a literal (a string of characters), it is an atom.
3. If an expression is an atom, it is an S-expression.
4. If an expression is a dotted pair, it is an S-expression.
5. If an expression begins with a left parenthesis, followed by an S-expression, followed by a dot, followed by an S-expression, followed by a right parenthesis, it is a dotted pair.

After receiving this instruction, ACT will have the sentences expressing these rules represented in its data base. However, this representation by itself does not allow it to perform any of the cognitive operations that would normally be thought of as demonstrating an "understanding" of these rules. In order to obtain such an understanding, a means of integrating these rules into ACT's procedural knowledge is required. Because these rules have the form of conditionals (antecedent implies consequent), they can be translated in a fairly straightforward manner into the condition-action format of productions. Table 21.2 illustrates four ACT productions for performing such a translation.[5] Production P18 handles

---

‘These productions and some others in this chapter embody some clearly oversimplified notions about language comprehension; a more adequate treatment would only distract attention from the learning processes that are the matters of present interest, however. For a discussion of language

TABLE 21.2
A Set of Productions for Encoding Rules
About LISP Structures

P18: IF there is a sentence beginning
    "If an expression is an LVword...,"
    where LVconcept is the concept for LVword
    THEN save *IF there is an LVconcept* for a new
        condition by attaching it to GVhold
    and set GVword to LVword

P19: IF there is a sentence ending
    "... GVword it is an LVword,"
    where LVconcept is the concept for LVword

    THEN BUILD:
```
IF GVhold

THEN it is an LVconcept
```

P20: IF there is a sentence beginning
    "If an expression begins with an LVword..."
    where LVconcept is the concept for LVword
    THEN save *IF an expression begins with an LVconcept*
        for a new condition by attaching it to GVhold
    and set GVword to LVword
    and set GVconcept to LVconcept

P21: IF a sentence has a phrase
    "... GVword followed by an LVword..."
    where LVconcept is the concept for LVword
    THEN save *IF there is a GVconcept before an LVconcept*
        for a new condition by attaching it to GVhold
    and set GVconcept to LVconcept
    and set GVword to LVword

---

the antecedents of the first four conditionals. For example, P18 matches the segment *If an expression is a number, ...* of Rule 1 by binding LVword to the word *number* and LVconcept to the concept @NUMBER that ACT considers underlies the word. Its action is to save the proposition *If there is a (a NUMBER* by attaching it to GVhold.

Production P19 is respons⁻ⁱ⁻, for actually building the productions encoding these rules. It obtains the conditions of these new productions from the global variable GVhold, which is given a value by other productions, and it obtains the actions from its own processing of the consequent parts of the rules. For example, in the case of Rule 1, P19 applies after P18, matching the remainder of the sentence *...number, it is an atom.* GVword had been previously fixed to

---

processing within the ACT framework, see Anderson, Kline, and Lewis (1977). (One complication necessary to any complete analysis of language comprehension is, nevertheless, being observed in some of the examples in this chapter—the distinction between words and the concepts underlying them.)

number by P18; the local variables LVword and LVconcept had no prior con-
straints (by the definition of a local variable) and received values of *atom* and
(*α* ATOM, respectively, in the process of matching. The action of P19 builds the
production:

> **P22:** IF there is a *(α* NUMBER
> THEN it is an *(α* ATOM

Production P22 is the mechanism by which ACT can actually make the in-
ferences authorized by Rule 1.

Productions P20 and P21 are responsible for processing complex conditionals
like Rule 5. P20 processes the first *begins* phrase and P21, each subsequent
followed by phrase. GVhold has as its value all of the condition elements col-
lected by P20 and P21. After the antecedent of the conditional has been entirely
processed, production P19 will apply to process the consequent and then desig-
nate a production. In the case of Rule 5, this production would be:

> **P23:** IF an expression begins with a *(α* LEFT-PARENTHESIS
> and this *(α* LEFT-PARENTHESIS is before an *(α* S-EXPRESSION
> and this *(α* S-EXPRESSION is before a *(α* DOT
> and this *(α* DOT is before an *(α* S-EXPRESSION
> and this *(α* S-EXPRESSION is before a *(α* RIGHT-PARENTHESIS
> THEN it is a *(α* DOTTED-PAIR

Designation With Substitution

The power of the designation mechanisms can be greatly increased by simply
allowing substitutions of one item for another throughout a designated produc-
tion. For example, consider the following production that might be useful in
learning by modeling:

> **P24:** IF when LVmodel sees LVevent1
> another event, LVevent2, occurs
> consisting of LVmodel doing LVaction
> THEN BUILD:
> > IF LVevent1
> > THEN LVevent2
>
> substituting ACT for all occurrences of LVmodel

Applied in a situation where Mommy says *Hi* to Alice after seeing her wave, P24
will designate:

> **P25:** IF ALICE waves to ACT
> THEN ACT say "Hi"

The substitution mechanism also allows ACT to handle implicit variables in
definitions. For example, when CONS(A B) = (A . B) is offered as a definition
(rather than an example) of the LISP function CONS, A and B are implicitly

variables. ACT knows this, in the sense that when it designates a production to encode this definition of CONS, it substitutes variables for the constants appearing as arguments.

GENERALIZATION

It is the ability to perform successfully in novel situations that is the hallmark of human cognition. For example, *productivity* has often been identified as the most important feature of natural languages, where this refers to the speaker's ability to generate and comprehend utterances never before encountered. Traditional learning theories are generally considered inadequate to account for this productivity, and ACT's generalization abilities must eventually oe evaluated against this same standard.

Although it is possible for ACT to designate new productions to apply in situations where existing ones do not, this kind of generalization requires having designating productions that correctly anticipate future needs. It is plausible that ACT could have such designating productions to guide its generalizations in areas in which it possesses some expertise. For example, if ACT were learning a second language, its experience with its first language might reasonably lead it to expect that the syntactic rules of this new language would treat whole classes of morphemes as equivalent (e.g., the class of all nouns), rather than including different syntactic rules for each individual morpheme. ACT's ability to substitute variables for constants when designating new productions would allow it to capitalize on this expectation and immediately generalize its competence beyond those sentences in the second language that it had actually observed.

It would be much more controversial to attribute such sophisticated expectations to ACT when it learns a first language; and even if it turned out to be justified in this case, it is highly unlikely that sophisticated expectations are available in all cases in which people can make generalizations. For this reason, ACT has the ability to create new productions automatically that are generalizations of its existing productions. This ability, though less powerful than the ability to designate generalizations, is applicable even in cases where ACT has no reliable expectations about the characteristics of the material it must learn.

Examples used to illustrate ACT's automatic generalization mechanism draw on productions from Table 21.3. Production P26 is a designating production that builds comprehension productions. It takes a sentence spoken by a teacher and makes it the condition of a production whose action is ACT's representation of the event the teacher is thought to be describing. When ACT hears this sentence in the future, this comprehension production will allow it to understand that another instance of the event the teacher described has occurred.

Productions P27 and P28 were built by production P26 based on pairings of the sentences *John gave the ball to Jane* and *Bill gave the dolly to Mary* with the

TABLE 21.3
The Productions Involved in Learning Two Possible
Sentence Structures for the Verb *Gave*

P26: IF the LVteacher says an LVsentence
    while pointing to an LVevent

THEN BUILD: IF LVsentence
              THEN LVevent

P27: IF there is a sentence "John gave the ball to Jane"
    THEN understand from this sentence that
        John caused a change in the possession
        of the ball from John to Jane

P28: IF there is a sentence "Bill gave the dolly to Mary"
    THEN understand from this sentence that
        Bill caused a change in the possession
        of the dolly from Bill to Mary

P29: IF there is a sentence "LVagent gave the LVobject to LVrecipient"
    THEN understand from this sentence that
        LVagent caused a change in the possession
        of the LVobject from LVagent to LVrecipient

P30: IF there is a sentence "Mary gave to John the ball"
    THEN understand from this sentence that
        Mary caused a change in the possession
        of the ball from Mary to John

P31: IF there is a sentence "Bill gave to Jane the dolly"
    THEN understand from this sentence that
        Bill caused a change in the possession
        of the dolly from Bill to Jane

P32: IF there is a sentence "LVagent gave to LVrecipient the LVobject"
    THEN understand from this sentence that
        LVagent caused a change in the possession
        of the LVobject from LVagent to LVrecipient

---

events they describe. ACT's automatic generalization mechanism forms a new production P29, which has variables in place of the constants that differ in these two designated productions. Production P29 will handle any sentence of the form *LVagent gave the LVobject to LVrecipient* and thus extends ACT's competence far beyond the specific examples encountered.

Formal Definitions

Further discussion of the properties of ACT's automatic generalization mechanism requires a formal definition (adapted from Vere, 1977): A production $C_1 \Rightarrow A_1$ is considered a *generalization* of $C_2 \Rightarrow A_2$ if $C_1 \Rightarrow A_1$ can apply in every circumstance that $C_2 \Rightarrow A_2$ can (and possibly others); and in these circumstances $C_2 \Rightarrow A_2$ would cause just the same changes to the data base as $C_1 \Rightarrow A_1$.

We can specify the conditions under which one production will be a generalization of another: Consider any consistent scheme for replacing local variables and constants in C_2 by local variables in C_1. We refer to this as a substitution θ. Let θC_2 denote C_2 after these substitutions have been made. Similarly, let θA_2 denote the action after the same substitutions. Then $C_1 \Rightarrow A_1$ is a generalization of $C_2 \Rightarrow A_2$ if and only if there is some θ such that $C_1 \subseteq \theta C_2$ and $A_1 = \theta A_2$.

Consider how this definition can be applied to show that production P29 (which corresponds to $C_1 \Rightarrow A_1$ in the definition) is a generalization of production P28 (which corresponds to $C_2 \Rightarrow A_2$): The substitution θ will replace *Bill* in P28 by *LVagent* from P29. Similarly, *dolly* will be replaced by *LVobject*, and *Mary* by *LVrecipient*. After the substitution θ, the two productions are identical; in the terms of the definition, $C_1 = \theta C_2$ and $A_1 = \theta A_2$. The fact that P29 is a generalization of P28 will be denoted by P29 < P28.

The result $C_1 = \theta C_2$ is stronger than what is required by the definition of generalization ($C_1 \subseteq \theta C_2$), which means that in forming the generalization P29 from P27 and P28, ACT could have deleted some condition clauses as well as substituting variables for constants. The reason no clauses were deleted is that ACT forms *maximal common generalizations* (this concept is also due to Vere, 1977). P29 is a maximal common generalization of P27 and P28 because P29 < P27 and P29 < P28, and there exists no production P such that P29 < P, P < P27, and P < P28. A maximal common generalization of P27 and P28 is one that deletes the minimum number of their clauses and replaces the minimum number of their constants by variables.

Productions P30 and P31 are the immediate results of a sequence of training trials whose eventual outcome is the generalization P32. P32 will comprehend all statements of the form *LVagent gave to Lvrecipient the LVobject*. These training trials were performed to demonstrate that ACT would properly distinguish the two different sentence structures for the verb *gave* and would not form a generalization of them that would handle all sentences containing this verb. There is no way to substitute corresponding variables from the condition of P29 into P32 to produce the identity of actions required by the definition of generalization.

There are occasions on which the maximal common generalization of two perfectly reasonable productions is a production that we would not want ACT to have. For example, consider the following pair of productions:

P33: IF there is an LVlocation in Asia
        that is wet and hot and flat
      THEN rice can be grown in this LVlocation
P34: IF there is an LVlocation in Vietnam
        that has roads,
        that is near the river,
        but that is not in the mountains
      THEN rice can be grown in this LVlocation

Their maximal common generalization is:

> P35: IF there is an LVlocation in an LVplace
>   THEN rice can be grown in this LVlocation

Here the perceived lack of commonality among these two sets of requirements for rice growing has led to the spurious generalization that rice can be grown anywhere. To avoid such obviously spurious generalizations, a restriction is placed on the number of constants that can be deleted in producing a generalization. If k is the number of constants in the smaller of the two conditions, then no generalization will be formed if more than .5 k constants must be deleted.

The Problem of Efficiency

A number of other researchers (e.g., Hayes-Roth & McDermott, 1976; and Vere, 1975, 1977, 1978) have also worked on generalization routines for production systems. Their routines use different computational techniques to produce generalizations of pairs of productions. ACT's generalization routine uses a rather brute-force technique that tries to put clauses from the two productions into correspondence by substitution of variables. Clauses that have no corresponding member in the other production are not included in the generalization. If there are n clauses in the condition of one production and m clauses in the other ($n > m$), there are potentially $n!/(n - m)!$ ways to assign correspondences. ACT's generalization routine manages to achieve some efficiency by the use of heuristics to guide the search for corresponding clauses. However, there is a sense in which research directed toward discovering efficient algorithms for generalizing two productions is hopeless. Hayes-Roth (1977) has observed that the generalization problem in its most general form is an NP-complete problem. Because it is widely believed that the time required to solve NP-complete problems must be an exponential function of the complexity of the problem, there is probably no entirely satisfactory algorithm for generalization.

Several features of ACT's generalization routine were motivated by this inevitable computational inefficiency. The first of these is that a limit is placed on the amount of computing time that will be spent trying to generalize any pair of productions. The second is that an attempt is made to generalize as few pairs as possible. A realistic simulation of an adult human's entire procedural knowledge would require hundreds of thousands of ACT productions. Under these circumstances, it would be disastrous to attempt to generalize all possible pairs of productions. Not only would this be astronomically costly but it would produce many spurious generalizations as well. ACT only attempts to form generalizations when a new production has been designated. Although no potential generalizations would be missed if a generalization were attempted for each possible pairing of this newly designated production with an existing production, an enormous computational cost would be required even under this scheme. For this reason, generalizations are attempted only for pairings of newly designated

productions with the productions on the APPLYLIST. Because a production is on the APPLYLIST only if the constants it references are active and it has met a strength criterion (see p. 202), this implies that attempts to generalize will be restricted to productions that are relevant to the current context and that have had a fair history of success.

Overgeneralization

Because ACT's automatic generalization mechanism extrapolates beyond observed situations, it is bound to make errors. However, given the goal of a realistic psychological simulation, such overgeneralizations on ACT's part would actually be desirable if it could be shown that people also overgeneralize in similar ways. For example, children learning language (and, it appears, adults learning a second language; see Bailey, Madden, & Krashen, 1974) overgeneralize morphemic rules. Thus a child will generate *mans, gived,* and so forth. ACT will do the same.

The following example illustrates some of the ways in which ACT will overgeneralize. Suppose that ACT has the set of productions shown in Table 21.4 for learning the syntactic structure of simple agent-action-object sentences. ACT brings to this effort the knowledge that certain morphemes refer to certain semantic categories. For instance, it knows that *dog* refers to the category *(a* DOG. When it encounters a known morpheme, it will assume that the semantic category is being referred to and will build this information into the production. However, when it encounters an unknown morpheme, it skips over it. In this example, ACT starts out not knowing how morphemes signal tense and number.

To learn the syntactic structure of a simple sentence, the productions in Table 21.4 require that the sentence can be paired with the event it describes. Productions P36 through P39 step through the sentence, collecting all the semantic relations provided by the morphemes whose meanings are known to ACT. Once they are finished, production P40 designates a production that will say this sentence in response to any other event that has occurred during the same time and that is given the same semantic categorization by these known morphemes. For example, when an adult model says *The dog chases the cat* as a description of some event occurring at TIME1, these productions will cause the designation of:

P41: IF *(a* DOGS are *(a* CHASING *(a* CATS at TIME1
        and the morpheme "dog" refers to the category *(a* DOGS
        and the morpheme "chase" refers to the category *(a* CHASING
        and the morpheme "cat" refers to the category *(a* CATS
     THEN say "The dog chase + s the cat"

Once ACT has P41, it will say *The dog chases the cat* in response to events that should actually be described by *The dogs chase the cat.* It will also use this sentence to describe events that should be described, *The dogs chased the cat*

TABLE 21.4
A Set of Productions for Learning the Syntactic
Structure of Simple Agent-Verb-Object Sentences

P36: IF GVmodel begins an LVsentence
     with an LVmorpheme that is used
       to refer to objects in LVcategory
    THEN save the proposition
    LVmorpheme refers to LVcategory
    by attaching it to GVrelations
    and set GVmorpheme to LVmorpheme
    and set GVsentence to LVsentence

P37: IF GVmodel begins an LVsentence
     with an LVmorpheme that is not known to
       refer to any LVcategory
    THEN set GVmorpheme to LVmorpheme
    and set GVsentence to LVsentence

P38: IF in GVsentence GVmorpheme is followed
     by LVmorpheme that is used to
       refer to objects in LVcategory
    THEN save the proposition
    LVmorpheme refers to LVcategory
    by attaching it to GVrelations
    and set GVmorpheme to LVmorpheme

P39: IF in GVsentence GVmorpheme is followed by
     LVmorpheme which is not known to
       refer to any category
    THEN set GVmorpheme to LVmorpheme

P40: IF the GVsentence ends with GVmorpheme
     and the GVmodel uses this sentence
       to describe the event of some number
       of LVagents LVacting on some number of
       LVobjects at LVtime

  THEN BUILD:
> IF some number of LVagents are
>   LVacting on some number of
>   LVobjects at LVtime and there
>   are the GVrelations
>   between these semantic
>   categories and some morphemes
> THEN say the GVsentence

(when TIME1 is no longer present). This shows that whereas the productions in Table 21.4 will designate only correct sentences if all the relevant morpheme-to-semantic-category correspondences are known, they will designate overgeneral productions in the absence of complete knowledge (i.e., of what ` +s`` and `` + ed`` signal). Thus, directly designated productions can be overly general even before automatic generalization comes into the picture.

In addition, the automatic generalization mechanism can be shown to act in such a way as to compound this overgeneralization. The distinction between

morphemes and the semantic categories they refer to is ignored in what follows to simplify the exposition. Thus, for example, the production designated when an adult says *The cat kills the rat* to describe an event is abbreviated by:

P42: IF cat happens to kill rat at TIME2
      THEN say "The cat kill+s the rat"

In response to the pair of productions P41 and P42, the automatic generalization mechanism will produce production P43 in Table 21.5. P43 generates what is really a present-tense, singular-subject sentence regardless of the actual tense or plurality requirements of the event that sentence is supposed to describe. The other productions in Table 21.5 are similar overgeneralizations produced in response to other examples of grammatical speech. For example, production P46 is a generalization over the productions designated in response to the adult sentences *The dogs chased the cat* and *The cats killed the rat*. Because all four productions in Table 21.5 have identical conditions, as far as these productions are concerned, the choice of inflection for subject and verb is entirely arbitrary. Another overgeneral feature of the productions in Table 21.5 is that they would apply to irregular words, generating items like *mans* and *gived*.

Thus, with the acquisition of the productions in Tables 21.4 and 21.5, ACT has passed from a state of never using the morphemes that express tense and number to a state in which they are used more or less haphazardly. Although there is evidence for similar transitions in the empirical literature on language acquisition, it is also the case that people eventually learn to correct their over-generalizations. The correction of overgeneralizations is primarily the responsibility of ACT's automatic discrimination mechanism.

DISCRIMINATION

One response to the problem of overgeneral productions is to designate new productions that apply in a more limited range of circumstances. However, just

TABLE 21.5
An Overgeneral Set of Productions for Generating
Agent-Verb-Object Sentences Referring to Singular
or Plural Subjects in Either Present or Past Tense

P43: IF LVagent happens to LVact on LVobject at LVtime
      THEN say "The LVagent LVact+s the LVobject"
P44: IF LVagent happens to LVact on LVobject at LVtime
      THEN say "The LVagent+s LVact the LVobject"
P45: IF LVagent happens to LVact on LVobject at LVtime
      THEN say "The LVagent LVact+ed the LVobject"
P46: IF LVagent happens to LVact on LVobject at LVtime
      THEN say "The LVagent+s LVact+ed the LVobject"

TABLE 21.6
Correctly Discriminated Versions of the Productions
in Table 21.5

P47: IF LVagent happens to LVact on LVobject at LVtime
         and LVagent is singular
         and LVtime is present
      THEN say "The LVagent LVact+s the LVobject"
P48: IF LVagent happens to LVact on LVobject at LVtime
         and LVagent is plural
         and LVtime is present
      THEN say "The LVagent+s LVact the LVobject"
P49: IF LVagent happens to LVact on LVobject at LVtime
         and LVagent is singular
         and LVtime is past
      THEN say "The LVagent LVact+ed the LVobject"
P50: IF LVagent happens to LVact on LVobject at LVtime
         and LVagent is plural
         and LVtime is past
      THEN say "The LVagent+s LVact+ed the LVobject"

---

as in the case of designated generalization, the existence of the required designating productions is plausible only for domains in which ACT already possesses some expertise. In such domains, ACT could possess the knowledge required to debug its own errors intelligently, but in the majority of cases, it will rely on its automatic discrimination mechanism. For example, ACT's automatic discrimination mechanism can form the new, correctly restricted productions of Table 21.6 from their overgeneral counterparts in Table 21.5 without recourse to any specific hypotheses about the nature of the material that must be learned.

An Earlier Discrimination Algorithm

A comparison of any pair of corresponding productions from these tables shows that the correctly discriminated member of the pair contains additional propositions in its condition involving the variables that occur in the condition of the overgeneral member of the pair. These additional propositions function to restrict the variable bindings that will satisfy the condition of the discriminate production to some subset of those variable bindings that will satisfy the condition of the overgeneral production. If the automatic discrimination mechanism can find additional propositions that restrict the set of variable bindings in just the right way, then the overgeneralization will be corrected.

Every time a production applies, there is an opportunity to obtain a new set of bindings for its variables. A proposition can then be chosen out of all those in the data base that mention any of these new bindings. This proposition (appropriately variabilized) can then be added to those in the condition of the production that has just applied to form a new discriminate production with the same action. (It

should be emphasized that the discriminate production does not replace the one it was formed from; productions used as the basis for discrimination or generalization continue to exist in the system alongside their "offspring.")

For example, the overgeneral production P43 might apply to generate the sentence *The girl hits the boy* to describe an event that occurred at TIME3. If there is a proposition in the data base stating that TIME3 is the present time, this proposition could be chosen to produce the discriminate production:

> P51: IF LVagent happens to LVact on LVobject at LVtime
>   and LVtime is present
>   THEN say "The LVagent LVact+s the LVobject"

A subsequent discrimination of P51 that chooses a proposition stating that the agent is singular would be required to produce the correct production P47 in Table 21.6.

As long as appropriate propositions are somewhere in the data base, a random choice out of all the propositions that mention new variable bindings is all that is required to guarantee that correct discriminations will eventually be found without any recourse to specific hypotheses about the nature of the material that must be learned. However, the power of random choice is always bought at some cost in efficiency. An earlier version of the automatic discrimination mechanism did randomly choose a proposition to form a new discrimination after every production application. However, very large numbers of discriminations were generated before the correct one was formed.

The Current Discrimination Algorithm

A new discrimination algorithm was developed that greatly increases efficiency. This algorithm makes a distinction between correct and incorrect actions. Productions place new propositions into the data base and emit observable responses; either of these actions can be declared incorrect by a human observer or by ACT itself. In the absence of such a declaration, an action is considered correct. That is, the only distinction made by the discrimination mechanism is between negative feedback and its absence (a later section takes up a possible role for positive feedback). Since the way in which ACT declares that the action of a production is incorrect is to apply another production that makes such a declaration as part of its own action, arbitrarily complex ACT computations can be performed to decide the correctness of any particular action.

The current automatic discrimination mechanism will only attempt to discriminate a production when it has both a correct and an incorrect application of that production to compare. Consider two applications of P43, one of which correctly generates *The boy hits the girl* to describe a present-tense situation and the other, which incorrectly generates this same sentence to describe a past-tense situation. Suppose the only difference between the variable bindings in these two

applications was that LVtime was bound to TIME4 when the present-tense sentence *The boy hits the girl* was correctly generated, and bound to TIME5 when it was incorrectly generated—that is, when the action took place in the past. Thus assuming that ACT has received the appropriate feedback, it can correct its behavior if it can discover the relevant difference between TIME4 and TIME5.

A search is made for propositions mentioning the binding that occurred in the later of the two applications. In the case where the correct binding, TIME4, occurred in the later application, this search might find the proposition *TIME4 is present*. However, before using this proposition to form the discrimination P51, a check is made that the analogous proposition *TIME5 is present* was not also in the data base at the time of the first, unsuccessful application. Finding such a proposition would show that the contemplated discrimination P51 would not have avoided the error made by the overgeneral P43. An attempt would then be made to find another proposition mentioning TIME4 that might better discriminate between successful and unsuccessful applications. If all propositions examined in this way fail, ACT forms no new production—it is possible that the feedback it received was unreliable.

In the case where the later of the two actions was the unsuccessful one, the proposition *TIME5 is past* might be found, which mentions the binding of interest. Because the analogous proposition *TIME4 is past* was not in the data base at the time of the earlier, successful application, a discriminate production with an *absence* condition is formed:

```
P52:  IF LVagent happens to LVact on LVobject at LVtime
         and LVtime is not past
      THEN say "The LVagent LVact+s the LVobject"
```

The current automatic discrimination mechanism also attempts to speed up the process of finding useful discriminations by its method of selecting propositions from the data base. Though still using a random process so as to maintain the guarantee that if the appropriate propositions are in the data base, they will eventually be found, this random choice is biased in certain ways that reflect general hypotheses about what sorts of propositions are likely to be incorporated by correct discriminations. Since the greater the amount of activation that has spread to a proposition, the more relevant this proposition is likely to be to the current situation, the discrimination mechanism chooses propositions with probabilities that vary with their activation levels. Because the strength of a proposition's interconnections to associated propositions is an overall indicator of its past usefulness, the discrimination mechanism also chooses propositions with probabilities that vary with their average strengths of association.

Discrimination by Specificity Ordering

The use of all these efficiency-promoting devices allows the automatic discrimination mechanism to correct rather quickly the overgeneral productions in Table

21.5 when provided with feedback about the sentences these productions generate. However, our experience with the simulations performed to date is that although correct behavior on ACT's part is obtained rather quickly, it is produced by a somewhat different set of productions than the completely discriminated ones shown in Table 21.6. Although discriminations that add one additional proposition (e.g., P51) are obtained in all four cases, once completely discriminated productions are formed in two of the cases, they block the erroneous applications required to complete discrimination in the remaining two cases.

To be more specific, suppose we have formed the discriminations shown in Table 21.7. Two of these productions, P48 and P49, are from Table 21.6. Each of these is included in just one cell in Table 21.7, showing that they are applicable to only one combination of tense and number; that is, they are completely discriminated.

On the other hand, Table 21.7 also contains the incomplete discriminations P53 and P54:

> P53: IF LVagent happens to LVact on LVobject at LVtime
>   and LVagent is singular
>   THEN say "The LVagent LVact+s the LVobject"

> P54: IF LVagent happens to LVact on LVobject at LVtime
>   and LVagent is plural
>   THEN say "The LVagent+s LVact+ed the LVobject"

Each of these is included in two cells, reflecting their overgeneral status. Cells in which they are the sole occupants indicate the combinations of tense and number for which they generate correct sentences, whereas membership in other cells indicates circumstances in which they will apply and produce errors. However, the left-to-right ordering of productions in these latter cells corresponds to their specificity ordering (p. 207); so, for example, if P49 is selected, it will apply instead of P53, thereby preventing an error. In effect, the specificity ordering provides the needed additional discriminations. The control structure we have in Table 21.7 can be indicated:

> If singular Then   If past Then apply P49
>   Else apply P53
> Else If plural Then   If present Then apply P48
>   Else apply P54

TABLE 21.7

A Categorization by Number and Tense of
the Situations in Which Four
Discriminate Productions Can Apply

	Singular	Plural
Present	P53	P48, P54
Past	P49, P53	P54

Whereas an *if-then* control structure is easily implemented in a production system, this *if-then-else* structure is possible in ACT only because of the specificity-ordering principle. The participation of P53 in this *if-then-else* control structure restricts its application in exactly the same way that the addition of the new condition clause *and LVtime is not past* would in an *if-then* control structure. This is the sense in which the specificity-ordering principle can produce discrimination.

Because P48 will prevent P54 from making errors in a similar fashion, these four productions in Table 21.7 produce errorless performance as long as the completely discriminated ones get selected (a probabilistic process) whenever they can apply. However, erroneous applications of P53 and P54 are just what are required to produce the completely discriminated productions that would occupy the major diagonal cells of Table 21.7 (i.e., P47 and P50).

Irregular Verbs

There are cases in which productions produced on the way to obtaining those in Table 21.6 are more than mere stepping stones. Notice that the productions in Table 21.6 generate grammatical sentences only for regular verbs—they would generate *The girl hitted the boy* to describe an event occurring in the past. Based on experiences where the sentences *The girl hit the boy* and *The boy hit the girl* were paired with the events they described, the following generalization would have been formed:

> P55: IF LVagent happens to hit LVobject at LVtime
>   THEN say "The LVagent hit the LVobject"

This production would sometimes apply incorrectly, perhaps to describe an event that would be correctly described by *The girl hits the boy*. Punishment of such errors would eventually result in a discrimination that correctly handles the irregular verb *hit:*

> P56: IF LVagent happens to hit LVobject at LVtime
>   and LVtime is past
>   THEN say "The LVagent hit the LVobject

PRODUCTION STRENGTH

The usual situation is for a number of ACT's productions all to have their conditions satisfied at the same time. On one hand, this gives ACT a capability for parallel processing that, we have argued elsewhere (Anderson, Kline, & Lewis, 1977), is crucial for an accurate simulation of complex cognitive skills like language processing. On the other hand, the assumption of the ACT model of procedural learning is that the acquisition of most complex cognitive skills requires trying out competing sets of productions for performing the same task.

These competing productions would all tend to have their conditions satisfied at the same time and to differ only in the appropriateness of their actions. The strength of an ACT production is a number that is interpreted as a predictor of this appropriateness. Decisions about which productions will actually apply, out of all those satisfied in any given situation, are made largely on the basis of their strengths. Consequently, ACT's ability to adjust the strengths of productions is an important component of its learning.

Adjustments to Strength

Because a production will not apply if it is not strong enough to be placed on the APPLYLIST (see p. 202), the impact of a production on ACT's performance depends crucially on that production's strength. ACT has a number of ways of adjusting the strength of a production in order to improve performance. Productions have a strength of .1 when first created. Each time it applies, a production has its strength increased by .025. However, when a production applies and receives negative feedback, its strength is reduced by a factor of .25. Because a multiplicative adjustment produces a greater change in strength than an additive adjustment, this "punishment" is much more effective than a reinforcement.

Although these two mechanisms are sufficient to adjust the behavior of any fixed set of productions, additional strengthening mechanisms are required to integrate new productions into the behavior of the system. Because these new productions are introduced with low strength, they would seem to be victims of a vicious circle: They cannot apply unless they are strong, and they are not strong unless they have applied. What is required to break out of this circle is a means of strengthening productions that does not rely on their actually applying. This is achieved by taking all of the strength adjustments that are made to a production that applies and making these adjustments to all of its generalizations that are in the system as well. Since a general production will be strengthened every time any one of its (possibly) numerous specializations applies, new generalizations can quickly amass enough strength to extend the range of situations in which ACT performs successfully.

For purposes of strengthening, recreation of a production that is already in the system, whether by designation, generalization, or discrimination, is treated as equivalent to a successful application in the sense that the recreated production receives a .025 strength increment, and so do all of its generalizations. One implication of this principle is that repetition of instructions has cumulative benefits for performance.

Interaction Between Strength and Specificity

Although selection rules based on strength can make some of the required choices among competing productions, it is clear that strength cannot be the sole criterion. For example, people reliably generate irregular plurals (e.g., *oxen*)

under circumstances in which the "add *s*" rules for regular plurals are presumably also applicable. This reliable performance is obtained despite the fact that the productions responsible for generating regular plurals are applied much more frequently than those for irregulars and therefore should be much stronger. ACT's solution to the problem of exceptions to strong general rules relies on the specificity-ordering principle to decide which productions on the APPLYLIST should actually execute. This principle accounts for the execution of a production generating an irregular plural, since its condition presumably contains all of the requirements for generating the regular plural and must, in addition, make reference to the specific noun to be pluralized.

The precedence of exceptions over much stronger general rules does not imply that exceptions are impervious to feedback, however. In order to benefit from the specificity-ordering principle, exceptions must first have achieved the amount of strength necessary to be placed on the APPLYLIST. Furthermore, because this amount depends on the strengths of the other productions that could apply, the stronger a general rule is, the more strength its exceptions need in order to apply reliably. But exceptions are designated with such low strength that one of the two mechanisms that can strengthen productions that have not actually applied must rescue them if they are ever to come to apply reliably. As it is unlikely that a newly designated exception is a generalization of any existing productions, inheriting the strengthenings given to specializations is not a solution in this case. Instead, repeated designations of the exception can provide the initial strength required for occasional placement on the APPLYLIST. Once this is achieved, a series of successful applications will be enough to produce consistent execution of the exception instead of the general rule.

The following example, which shows ACT learning to refer to objects with definite and indefinite articles, illustrates this interaction between strength and specificity. The example begins with ACT in the situation of a young child who knows how to refer to objects with nouns, but who does not yet know how to modify them with articles. ACT's knowledge here takes the form of the production:

```
P57:  IF the goal is to refer to LVobj
         and LVc is the concept for LVobj
         and LVword is the word for LVc
      THEN say LVword
```

By some unspecified process, ACT forms the general hypothesis that the speaker's choice of article is determined by the listener's relation to the object being referred to. This hypothesis also takes the form of a production:

```
P58:  IF GVmodel is referring to LVobj
         and LVc is the concept for LVobj
         and LVword is the word for LVc
         and the listener has LVrelation to LVobj
         and the model says "LVword1 LVword"
```

THEN BUILD:

```
IF the goal is to refer to 'LVobj'
   and 'LVc' is the concept for 'Lvobj'
   and 'LVword' is the word for 'LVc'
   and the listener has LVrelation
       to 'LVobj'
THEN say "LVword1 'LVword' "
```

(In the process of designating new productions, P58 will substitute variables—see p. 210—for the items in single quotes.) Whenever there are new data relevant to this general hypothesis about the dependence between the speaker's choice of article and the state of the listener, P58 designates a production to embody the specific hypothesis supported by these new data. In particular, one of the productions P59 or P60 will be designated by P58 on almost every occurrence of articles in adult speech:[6]

```
P59:  IF the goal is to refer to LVobj
          and LVc is the concept for LVobj
          and LVword is the word for LVc
          and the listener is aware of LVobj
       THEN say "THE LVWORD"
P60:  If the goal is to refer to LVobj
          and LVc is the concept for LVobj
          and LVword is the word for LVc
          and the listener is unaware of LVobj
       Then say "A LVword"
```

The conditions of P59 and P60 are both supersets of the condition of P57. Therefore, if either one of these productions that use articles in referring is on the APPLYLIST, it will apply instead of P57, which only uses nouns, by the specificity-ordering principle.

Once ACT has the designating production P58, the course of learning may be observed. A training trial consists of providing an example of reference using articles. The amount of learning that has occurred can be assessed with test trials, produced by entering propositions into the data base that satisfy the productions that have been designated. There must be a proposition to the effect that ACT has the goal of referring to an object. There must also be a statement about the listener's awareness/unawareness of the object in question. For example, if the listener is said to be aware of the dog that ACT wants to refer to, either production P57 will apply generating *dog*, or production P59 will apply generating *the dog* (errors like *a dog* were not possible in this simulation). The proportion of test trials on which an article is used is a measure of ACT's learning.

---

[6]Choice of article is more complicated than implied here; see Brown (1973, pp. 340-350) for a discussion.

The details of the simulation were as follows: Production P57, which refers without articles, was given an initial strength of 20. The designating production P58 was given a strength of only .1, reflecting the fact that it is a new hypothesis about articles. Training trials alternated with test trials, and definite articles alternated with indefinite articles. Thus a series of four trials had the form: train with definite article, test use of definite article, train with indefinite article, test use of indefinite article. A complete simulation of learning to use articles required 10 such blocks of four trials. (ACT undoubtedly learns too rapidly to be an accurate model of humans; however, the computational expense of a more accurate simulation would be prohibitive.) Ten replications were performed of the complete simulation in order to obtain proportions of article use in each block. The course of learning was different in each replication because of the probabilistic nature of production selection.

In qualitative terms, the results of the simulations were as follows: On the first few training trials, the designating production P58 applied unreliably due to its low strength; even when it did apply, the productions it designated (P59 and P60) were too weak themselves to apply reliably on test trials. However, when P58 did manage to apply, it was strengthened, resulting in more reliable designation on subsequent training trials; this led, in turn, to the strengthening of P59 and P60. The combined strengthening influences of frequent redesignation and successful application were enough to produce reliable generation of articles by the end of the simulation.

The results are shown in quantitative terms in Fig. 21.2. There is a relatively rapid, but not all-or-none, change in the level of performance. The best and the
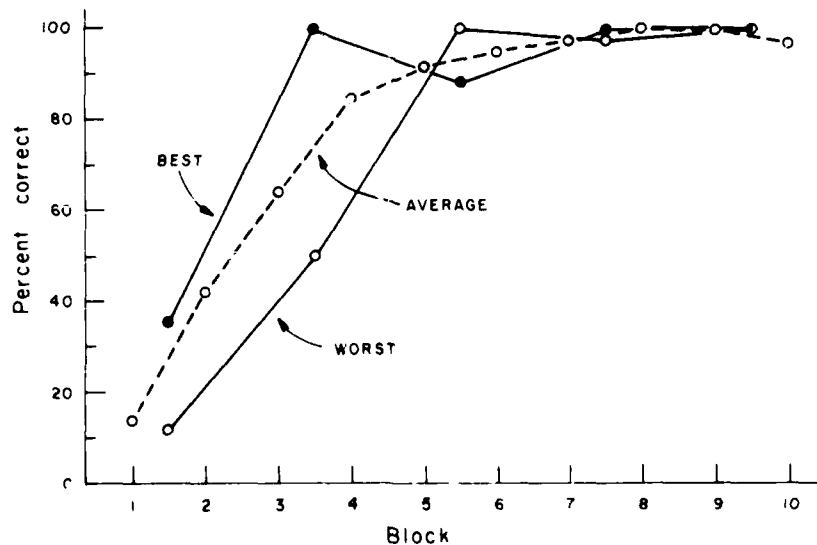
FIG. 21.2.   The increase with time of ACT's use of the articles *a* and *the*. Successive blocks are averaged in reporting the best and worst individual runs.

worst simulations show much the same pattern as the average of all 10. These rapid changes can be explained by the tendency for success to feed on itself in ACT. A successful execution of a production results in an increase in its strength and consequently greater opportunity for further execution and strengthening. Roger Brown (1973) reports that young children show just these sharp, but not all-or-none, changes in their percentage of correct use of grammatical morphemes.

Designation Takes Precedence Over Strength

An argument can be made for adding yet another principle of production selection to those already operative in the previous example: For several cycles following the designation of a production, an attempt should be made to apply this new production before applying any of the productions on the APPLYLIST.[7] This principle allows us to explain the fact that with some effort, it is possible for adults to override highly overpracticed rules deliberately. For instance, it is possible to replace the ''add *s*'' rule for pluralizing nouns with an ''add *er*'' rule (e.g., *three booker*). The explanation runs as follows: The production that implements the ''add *er*'' rule is repeatedly designated as long as a deliberate effort is being made to perform the new pluralization. By virtue of having been just designated, it is applied in preference to the ''add *s*'' rule. When the deliberate effort is no longer maintained, designation ceases, the ''add *er*'' production fails to be placed on the APPLYLIST because of its low strength, and the strong ''add *s*'' rule reasserts itself.

The results of some experiments by LaBerge (1973) have a similar explanation involving the precedence of designation over strength. LaBerge had subjects make same–different judgments for familiar alphabetic symbols and for unfamiliar letterlike symbols. Reaction time in this task can be thought of as determined by the number of cycles required to select the relevant productions. This quantity will be inversely related to the strengths of the productions unless designation causes automatic selection. Because alphabetic symbols presumably have very strong productions responsible for their recognition, the reaction-time advantage usually found for these symbols can be explained as due to strength differences. However, when subjects knew ahead of time what symbol would be involved in the judgment, there was no advantage for the familiar symbols. This can be explained as due to the automatic selection of productions designated to recognize the expected symbol.

---

[7]In the simulations discussed in the previous section, the ability to apply various productions was used to assess the amount of procedural learning that ACT had accomplished. However, the principle being proposed now means that a production that applies easily after designation might be very difficult to apply later on. These earlier simulations were run without giving preference to designated productions. This is equivalent to having many intervening events between each simulated event.

Discrimination by Restriction Versus
Discrimination by Exception

There is an important distinction to be made in ACT between two types of discrimination, only one of which can be formed by automatic discrimination. ACT's automatic discrimination mechanism cannot form an exception to a general rule because the exception would need a different action. Productions with new actions can only be formed by designation. The automatic discrimination mechanisms merely modify the range of situations in which an existing action will be performed; that is, they correct overgeneralizations of that action. This we call discrimination by *restriction* to distinguish it from the discrimination by *exception* required in the pluralization example of the previous section.

It is interesting to compare the ways in which exceptions and restrictions are integrated into the behavior of the system. First, consider the similarities: We have seen previously that after being designated with low strength initially, repeated redesignation allows exceptions to accumulate the strength required for occasional placement on the APPLYLIST. Nothing prevents the automatic discrimination mechanism from choosing, on different occasions, the same proposition from the data base to use in forming new productions. Thus, in all likelihood, the same restriction of an overgeneral production will be formed multiple times; therefore, just as is the case with exceptions, it is possible for multiple formations to provide the strength necessary for placing restrictions on the APPLYLIST. Once occasional placement on the APPLYLIST is achieved, a history of successful applications will increase the strength of both exceptions and restrictions to the point where they will apply reliably in the future.

However, interesting differences between exceptions and restrictions emerge when we consider circumstances in which these discriminations do not apply. When an exception is not applicable, its general rule will take over and presumably be strengthened for correct performance. The intention is that both the exception and the general rule should coexist in the system, and, in fact, as long as occasions to apply the exception are frequent enough, neither will grow in strength at the expense of the other.

On the other hand, assuming that our restriction is the right one (i.e., its action is called for in just those situations described in its condition), whenever this restriction is not applicable, any application of its overgeneral source results in errors. These errors will presumably be punished, costing the overgeneralization .25 of its strength each time. Here the intention is that the correct restriction should come to replace its overgeneral source in the operation of the system, and, in fact, the restriction grows rapidly in strength relative to its source. It can be strengthened in all situations in which its source is strengthened; but it avoids all the punishment the source receives for misapplication.

It is *relative* loss of strength of the source that is important here. Because production selection evaluates the strength of a production relative to the

strengths of all productions with active constants, a production will be selected for the APPLYLIST with a probability of 1.0, regardless of its strength on occasions in which it is the only active production. This implies that negative feedback would not be effective in the ACT system if it only reduced strength and did not also result in the creation of competing productions through automatic discrimination. On this issue, ACT is supported by the learning literature (Estes, 1970; Hilgard & Bower, 1966), which indicates that negative feedback works not so much by "stamping out" behaviors as by producing alternative behaviors.

Another prediction that follows from the ACT model is that negative feedback should play an important role in the learning of any complex procedure, since without it, the automatic discrimination mechanism cannot operate. This prediction is in direct conflict with the widespread belief that negative feedback is completely ineffectual in first-language acquisition. For example, Cazden (1965) has reported that providing children with corrected versions of their ungrammatical utterances does not result in more rapid acquisition of the correct forms. If this claim is accurate (and there is some evidence that it is not; see McNeil, 1970), then it can only be explained in ACT terms by assuming that the children were for some reason incapable of determining just which productions should have been punished from the negative feedback that was provided.

THE ORIGIN OF DESIGNATING PRODUCTIONS

Although procedural learning involves the acquisition of new behaviors, as noted earlier, ACT's automatic generalization and discrimination mechanisms cannot add new actions to productions. The designation process is thus indispensable to the ACT theory of procedural learning because it alone has the ability to introduce productions with new actions into the system. Once this is appreciated, it becomes necessary to account for the acquisition of the designating productions themselves. In our work to date, the only requirement we placed on ourselves in proposing designating productions for ACT in learning some skill was that a human learner of that same skill might plausibly possess the knowledge incorporated in those productions. Given our interest in the learning of complex procedures, this seemed like a good strategy since it would be very difficult to give any detailed account of the origins of the sophistication that is d anded from the learner of any complex procedure. Of course, this is only defensible as a short-term strategy—the ACT learning theory is distressingly incomplete as long as the origin of designating productions is unexplained. The function of the present section is to present some speculations on the origin of ACT's designating productions.

Experience can always be expected to function, in at least a crude way, to recommend certain new behaviors; it would be reasonable for ACT to start out

already having designating productions that capitalize on this expectation. For example, we saw the following modeling production earlier:

P24 IF when LVmodel sees LVevent1
        another event, LVevent2, occurs
        consisting of LVmodel doing LVaction
     THEN BUILD:

        ┌─────────────────┐
        │ IF LVevent1   │
        │ THEN LVevent2   │
        └─────────────────┘

     substituting ACT for all occurrences of LVmodel.

Actions performed by models in various situations have a high likelihood of being appropriate for ACT in those situations as well, and this makes P24 a good candidate for membership in the set of original designating productions. Other candidates for this set are inspired by the principles of traditional learning theory. For example, there is production P61 of Table 21.8, which incorporates a reinforcement principle.

Now it might appear that production P61 is useless for producing new behaviors because it requires that ACT has already performed the behavior in question. However, in conjunction with a mechanism that randomly generated all the behaviors of which ACT is capable, P61 would enable a reinforced behavior to be incorporated into a production where it could be performed under stimulus control for the first time. A (rather anthropomorphic) example would have ACT reinforced for accidentally saying *mama* when its mother is near. The following production would be designated, which represents a modest, but necessary, step toward the lexicalization of natural language; that is, it introduces a connection between the word *mama* and the concept (*a* Mommy:

P66: IF ACT sees *a* Mommy
      THEN ACT say "mama"

Alternatively, the environment can act in a highly directive way to produce a passive action on ACT's part—as, for example, when an adult takes a child's hand and makes it go through the motions required to tie a shoe. Production P61 would allow ACT to produce such behaviors on its own subsequently.

It is just possible that original designating productions of these sorts, in combination with the automatic generalization and discrimination mechanisms, is all the "innate endowment" that ACT requires to account for human procedural learning. The remainder of this section attempts to provide support for this possibility by demonstrating that one of the designating productions required to comprehend verbal instructions can be formed from generalizations and discriminations of some original designating productions. The original designating productions that are used are the reinforcement production P61 and production P63 from Table 21.8. Production P63 designates new productions that predict the consequences of ACT's behavior. These new productions will apply whenever that behavior is performed in the future and will predict the same consequences

TABLE 21.8
Two Innate Designating Productions (P61, P63),
Two Discriminations (P62, P64) and
One Generalization (P65)

P61: IF LVevent occurs just before ACT performs
　　　　LVaction which is followed by reinforcement

　　THEN BUILD:　　IF LVevent
　　　　　　　　　　THEN LVaction

P62: IF LVevent occurs just before ACT performs
　　　　LVaction which is followed by reinforcement
　　and a teacher has said
　　　　"If LVclause1 then LVclause2"
　　and LVevent is the meaning of LVclause1
　　and LVaction is the meaning of LVclause2

　　THEN BUILD:　　IF LVevent
　　　　　　　　　　THEN LVaction

P63: IF ACT performs LVaction which is
　　　　followed by LVeffect

　　THEN BUILD:　　IF LVaction
　　　　　　　　　　THEN LVeffect

P64: IF ACT performs LVaction which is followed by LVeffect
　　and a teacher has said
　　　　"If LVclause1 then LVclause2"
　　and LVaction is the meaning of LVclause1
　　and LVeffect is the meaning of LVclause2

　　THEN BUILD:　　IF LVaction
　　　　　　　　　　THEN LVeffect

P65: IF a teacher has said
　　　　"If LVclause1 then LVclause2"
　　and LVcondition is the meaning of LVclause1
　　and LVaction is the meaning of LVclause2

　　THEN BUILD:　　IF LVcondition
　　　　　　　　　　THEN LVaction

that were obtained previously. The automatic discrimination mechanism can form two new productions, P62 and P64 in Table 21.8, from the original designating productions P61 and P63. Both of these discriminations result from ACT's observation that occasions on which useful designations are formed are often those on which teachers use a particular kind of sentence (*if-then*) that refers to the events involved in the designation.[8]

---

[8]Actually, at present there is no way to punish the designating productions, as is required to produce these discriminations. First of all, they have as their actions the creation of new

Once these two discriminations have been formed, a generalization over them produces the designating production P65 in Table 21.8, which is responsible for comprehending verbal instructions. Thus by processes of discrimination and generalization, two designating productions that record events surrounding ACT's own actions ultimately give rise to a designating production of a very different character. Our hope is that all of the designating productions ACT requires for procedural learning can be produced in this same manner.

FUTURE DIRECTIONS:
INSPECTION OF PRODUCTIONS

Currently, one ACT production cannot inspect the contents of another ACT production, because the productions themselves are not represented in the data base. As a consequence, it is impossible to use productions to analyze the procedures that ACT has available for performing some task in order to isolate and correct "bugs" in those procedures. The idea that procedural learning consists of a debugging process has motivated a great deal of recent work in cognitive science (Brown, Burton, Hausmann, Goldstein, Huggins, & Miller, 1977; Goldstein, 1974; Sussman, 1975). Although we think that debugging processes require too much domain-specific knowledge to account for much of human procedural learning, it is undeniable that experts can analyze the procedures they are using to find and correct bugs. An example comes from our experiences in learning to program in the language C, where all indexing initiates at 0 rather than the more customary 1. Introspection suggests that this requires systematically reworking familiar procedures for searching arrays, looping, and so forth to compensate for this convention, which was unfamiliar to us. To make it possible to model such debugging processes, we intend to modify the ACT system to allow productions to treat other productions as data—that is, to allow productions to test for the existence of various other kinds of productions and, upon finding them, to add to them or make other modifications.

Although the primary motivation for this change is to expand ACT's learning capabilities, it appears that making productions inspectable will provide benefits for the nonlearning (performance) aspects of the system as well. One expected benefit is that it should become easier for ACT to direct its behavior in service of

productions wholly internal events that could never be evaluated by an outside observer. Secondly, neither the internal event of designation nor the new production that results from that event can currently be evaluated by the production system itself, because they are not represented in the data base that the system has available for inspection. However, as is discussed in the next section, we anticipate expanding ACT's data base to take care of this difficulty.

its goals. For example, in the LISP-learning simulation discussed earlier (p. 210), we had a production for categorizing a sequence of symbols as a dotted-pair:

```
P23:  IF an expression begins with a (a LEFT-PARENTHESIS
          and this (a LEFT-PARENTHESIS is before an (a S-EXPRESSION
          and this (a S-EXPRESSION is before a (a DOT
          and this (a DOT is before an (a S-EXPRESSION
          and this (a S-EXPRESSION is before a (a RIGHT-PARENTHESIS
      THEN it is a (a DOTTED-PAIR
```

Notice that this production depends on the subsequences having already been categorized as S-expressions; that is, it assumes a bottom-up sequence of processing where all decisions about high-level constituents must wait on decisions about all low-level constituents. The difficulty with this scheme is that the failure of a single production to apply—due to low strength or to a failure to spread activation to all of the required memory structure—holds up the entire sequence of processing. In addition, there is a great deal of wasted effort, because low-level categorizations are made without regard to their usefulness for deciding between the various high-level categorizations that are viable at the moment.

Giving productions the ability to inspect other productions makes it possible to implement a top-down scheme that avoids some of these difficulties. Productions will respond to the top-level goal of showing that a particular expression is a dotted-pair by searching for other productions that make this categorization as part of their action. This search will find production P23, and then productions will notice that the condition of P23 can be satisfied if there are S-expressions on both sides of the dot. This leads, in turn, to a search for productions that categorize symbol sequences as S-expressions, and the entire process repeats itself until a production is found whose condition is satisfied but that has not yet applied. If it is low strength that has prevented this production from applying previously, then redesignating it will enable it to apply now. Alternatively, because the process of finding this production involved focusing the system's attention on successively smaller constituents of the dotted-pair, this refocusing can be expected to activate any memory structure whose inactivity blocked the application of this production previously. In any case, the ability to implement this top-down process should result in more reliable achievement of the system's goals.

It is generally acknowledged that the design of a performance system will have strong influences on the learning system. That is, our learning principles will be strongly influenced by our conception of what the end product of the learning process is like. On the other hand, it is also the case, as just illustrated, that work with a learning theory will affect the performance theory. There is a complex and intimate relationship between the two. It is preferable—and fortunately, it is possible for us—to pursue both endeavors in parallel.

ACKNOWLEDGMENTS

REFERENCES

Anderson, J. R. *Language acquisition by computer and child* (Tech. Rep. No. 55). Ann Arbor: University of Michigan, Human Performance Center, 1974.

Anderson, J. R. Computer simulation of a language acquisition system: A first report. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Anderson, J. R. *Language, memory, and thought*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Anderson, J. R. Induction of augmented transition networks. *Cognitive Science*, 1977, *1*, 125-158.

Anderson, J. R. Computer simulation of a language acquisition system: A second report. In D. LaBerge & S. J. Samuels (Eds.), *Perception and comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Anderson, J. R., Kline, P., & Lewis, C. A production system model for language processing. In P. Carpenter & M. Just (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Bailey, N., Madden, C., & Krashen, S. D. *Is there a "natural sequence" in adult second language learning?* Unpublished manuscript, English Language Institute and Linguistics Department, Queens College and the Graduate Center, City University of New York, 1974.

Brown, J. S., Burton, B. R., Hausmann, C., Goldstein, I., Huggins, B., & Miller, M. *Aspects of a theory for automated student modelling* (BBN Rep. No. 3549). Bolt, Beranek, & Newman, Inc., Cambridge, Mass.: 1977.

Brown, R. *A first language*. Cambridge, Mass.: Harvard University Press, 1973.

Cazden, C. G. *Environmental assistance to the child's acquisition of grammar*. Unpublished doctoral dissertation, Harvard University, 1965.

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

Estes, W. K. *Learning theory and mental development*. New York: Academic Press, 1970.

Goldstein, I. P. *Understanding simple picture programs* (Rep. No. AI-TR-294). Cambridge, Mass.: M.I.T. A.I. Laboratory, 1974.

Hayes-Roth, F. *The role of partial and best matches in knowledge systems* (Rep. No. P-5802). Rand Corporation, Santa Monica, Calif.: 1977.

Hayes-Roth, F., & McDermott, J. *Learning structured patterns from examples*. Proceedings of the Third International Joint Conference on Pattern Recognition, November 1976.

Hilgard, E. R., & Bower, G. H. *Theories of learning*. New York: Appleton-Century-Crofts, 1966.

LaBerge, D. Attention and the measurement of perceptual learning. *Memory & Cognition*, 1973, *1*, 268-276.

McNeil, D. *The acquisition of language*. New York: Harper & Row, 1970.

Newell, A. A theoretical exploration of mechanisms for coding the stimulus. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston, 1972.

Newell, A. Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Rychener, M. D., & Newell, A. An instructible production system: Basic design issues. In D. A. Waterman & F. Hayes-Roth (Eds.), *Pattern-directed inference systems*. New York: Academic Press, 1978.

Sussman, G. J. *A computer model of skill acquisition*. New York: American Elsevier, 1975.

Vere, S. A. *Induction of concepts in the predicate calculus*. Proceedings of the Fourth International Joint Conference on Artificial Intelligence, Tbilisi, USSR, 1975.

Vere, S. A. *Induction of relational productions in the presence of background information*. Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Boston, August 1977.

Vere, S. A. Inductive learning of relational productions. In D. A. Waterman & F. Hayes-Roth (Eds.), *Pattern-directed inference systems*. New York: Academic Press, 1978.

Weissman, C. *LISP 1.5 primer*. Belmont, Calif.: Dickenson, 1967.

22 Discussion: Teaching, Learning, and the Representation of Knowledge

Donald A. Norman
University of California, San Diego

The problems of teaching and of learning center around the problem of representation. The problem can be viewed as that of getting the knowledge of the topic matter into the mind of the student. This requires some understanding of how knowledge is represented within a person's memory structures. Several different forms of knowledge are relevant: There is knowledge of the topic that is to be learned; there is the knowledge that the student already has; there are strategies used by both teacher and student in the attempt to acquire topic matter knowledge, making use of what is already known. In our quest for understanding the teaching and learning process, we must come to understand how all these different forms of knowledge are represented. But representation is not enough. We must also come to understand those processes that operate upon the representation   understand how information is used.

The fo   ·pters under review represent quite different approaches to different facets o. lese problems. Four different chapters, four quite different topics, yet with one common theme underlying representation. Let me go over those papers, covering each one quickly, discussing some of the major issues. This review is short. The chapters themselves are detailed and require careful study. Here, I am concerned primarily with the critical aspects of the chapters to the theme of this conference: teaching and learning. But these four chapters go beyond this topic. They all contribute to our general theoretical understanding of learning, of teaching, and of the problems of representation. Thus, they are important to cognitive psychologists in general, not just those interested in learning and teaching.

LEARNING AND UNDERSTANDING THROUGH
THE USE OF CONCEPTUAL MODELS

How should one teach? What is the nature of the basic process by which a student acquires a new topic matter? These are the questions emphasized by two of the chapters—the one by VanLehn and Brown on procedural morphisms and the chapter by Stevens and Collins on conceptual models. Stevens and Collins show how a student's conceptualization of a situation can lead to difficulties when it proves to be too simple for the phenomena. Natural phenomena, such as the control of weather patterns, are too complex to be understood by a complete, formal model. As a result, we use simplified descriptions of the situation, simple models that help us understand the essential variables. These simplified models must be chosen and used with some care, however, lest they lead to conceptual difficulties. The chapter by Stevens and Collins illustrates the kinds of models used by students and the forms of problems that result.

VanLehn & Brown examine what happens when a topic matter is taught by analogy to another topic. Thus, in teaching arithmetic, one technique is to introduce instructional blocks that can be used to help understand the numerical operations. Various forms of blocks exist—some set up to make clear the radix operations of arithmetic, others designed to clarify the notions of addition and subtraction of number, others with different emphasis (including some with ill-defined characteristics because the block or toy designer did not appear to have a clear notion of the conceptual structure of the arithmetic operations that were to be clarified). The point in introducing blocks (or some other model structure) is that the teaching and learning be enhanced by forming an analogy between the model and the topic to be learned. A teacher attempts to teach a topic domain by introducing a model that has several properties: The model must itself be easy to learn, it must provide an appropriate analogy to the target domain, and the mapping of attributes of the model onto attributes of the topic must be clear. The philosophy underlying the use of a model is that it provides a simplification of the overall path toward understanding of the topic. Ideally, the model is related to the target topic in a direct and straightforward way. Each aspect of the model should represent an aspect of the target. A unique, one-to-one mapping from one domain onto the other is called an isomorphism. In their chapter, VanLehn and Brown examine the possible mappings between model and task—morphisms. The chapter attempts an important first step toward a formalization of learning through analogy.

Analogy and Metaphor: Key Concepts in Understanding

Although the chapters by Stevens and Collins and by VanLehn and Brown talk about the use of conceptual models in different ways and for different purposes, the common, underlying thrust is the same. When students learn, they build upon

existing knowledge, and the mechanism by which the new knowledge is formed is that of construction of an analogy or metaphor.

The Chapter by Stevens and Collins.   What did Stevens and Collins do? They showed how students used conceptual models in their attempt to understand any given topics. These models are used not only to give a cohesive picture but also to expand the general knowledge, to answer questions about aspects of the topic not thought about before. I believe this chapter to be an important one, for the study of how people use their conceptual models must surely be at the heart of how people understand.

A critical aspect of these models is the selection of attributes. All the models have in common the fact that they simplify the structure that is to be understood. But simplification poses dangers, for by its very nature it must ignore or smooth over some of the complexities of the actual situation. It is simply impossible to understand a real-world topic that is as complex as weather patterns by un er- standing the detailed interaction of air masses, temperatures, and moisture   ͻ the complexities of terrain and real geography. Some simplification is requi   ᶦ. The problem is to choose the appropriate simplification. It is mos᷄ likely tɻ ᴝ different simplifications are required for different purposes, and it i� ᷄mportant thaᵗ a student realize the nature of the conceptual model that is useᵈ. All thᶦ conceptual models are erroneous in that they do not capture all that goes on. But all are correct in their description of some essential aspect of behavior. Thᶦ point is that the student must understand the nature of the models, of the simplifica- tions, and the appropriate ways in which they can be used.

The Chapter by VanLehn and Brown.   VanLehn and Brown wish to for- malize the building of one knowledge structure based upon an explicit analogy with another, an analogy presented by the teacher specifically for the purpose of building from a topic understood by the student to one that is to be acquired. I think this chapter is of potential great importance, for it could lay a formal formation for the understanding of learning by analogy.

Unfortunately, I find the chapter to be flawed. I have two major objections. The first is that, to me, large sections of the chapter are quite unintelligible. The second problem is that I believe the work not to be fully developed. I am much more sympathetic toward the second problem than toward the first. VanLehn and Brown are attempting an extremely important and difficult task, the formaliza- tion of the problem of learning and teaching by analogy. That this first attempt should be incomplete is quite understandable. I urge them to continue the de- velopment of these ideas.

The problem of intelligibility is less defensible. Indeed, given the chapter's emphasis on developing clear understanding of a topic matter, building up slowly from what is understood in nice, simple, direct steps, I find this flaw somewhat amusing. We are told (twice!) that "if AI has contributed anything to cognitive

psychology, it is an appreciation that ignoring trivial detail often leads to overlooking nontrivial problems." Ouch. That certainly is not the lesson I thought I had learned from the studies of artificial intelligence. If that sentence means anything, it means that when one is just beginning the study of a topic matter, what appears to be a trivial detail often in practice is not. When one lacks knowledge, the only way to distinguish trivial details from important concepts is to build a complete system. This is a common problem in all sciences. The system that one builds can be of many different forms: a physical model (as in the case of the solution of the structure of DNA), a mathematical model, or perhaps a computer simulation. Along the way, one discovers the critical features of the system and just which pieces of knowledge are important and which are irrelevant. It is from this aspect of the problem that the quotation has arisen: In the building of a complete system, often quite innocuous details turn out to be critical. But then, when the model is complete, one knows what is important and what is not. You tell the reader the important stuff and leave the trivia for the appendix, or for technical papers in the specialized journals. VanLehn and Brown have begun the modeling process. They are moving in what I believe to be the correct direction, but because they have not yet completed their task, they themselves do not know which aspects are important and which are the "trivial details." Alas, that is no reason to subject the poor reader to page after page of horrendous detail. I was overwhelmed with more information than I could assimilate, underwhelmed by the importance of it all.

I believe firmly that what VanLehn and Brown are attempting to do is important. Moreover. I think the philosophy of the approach is probably correct. This chapter is simply premature; more development is required. I urge you to read the chapter, but for the intent, not for the details.

I must add a positive note to the complaint. VanLehn and Brown have continually revised their paper, even as I write this review. As a result, my comments have always been one draft behind in the multiple versions that they have produced. Therefore, the comments in this chapter are based upon the last version of their paper available to me. The version that is published in this book has probably gone through one more revision. (I suspect that most of my comments still apply.) Each revision, by the way, marked a substantial improvement in their thinking and in their presentation. I trust the last revision has done the same.

COMPLEX LEARNING:
THE CHAPTER BY
ANDERSON, KLINE, AND BEASLEY

The study of complex learning differs from the more conventional study of learning in that its emphasis shifts to the development of appropriate organization and representation of the information being acquired, rather than the simple

formation of associative bonds that is the major interest of most existing psychological learning theories. Anderson, Kline, and Beasley examine the mental processes used by a learner in constructing an internal representation of a new topic matter. The theoretical tool is the combination of propositional representation and production systems. The result is a computer simulation that is capable of doing some learning, providing a step on the way toward a theory of human learning.

The attempt here is important. It is the development of a theory of the learning of complex topics based upon the notions of modern processing concepts and upon the development of the representational structures of the learner. Anderson et al. present four different mechanisms for learning: designation, strengthening, generalization, and discrimination. I fear that in their current form, these will not suffice.

Before I begin a critique, let me explain. I think the work that is represented in this chapter is important and must go on. Many of the problems of the current approach are known by the authors and, indeed, are spelled out within the chapter. There is nothing the matter with making early attempts that are known to be insufficient but, nonetheless, will advance our general understanding of the principle under study. I believe that the mechanisms described here for generalization and for discrimination are quite insufficient, but these insufficiencies are pointed at within the chapter itself, so I need not elaborate upon them here. Generalization and discrimination must be two important aspects of a learning theory, and for the topic being studied here, the form adopted seems adequate.

The major shortcoming of the approach, in my opinion, limits it to the study of "incremental" learning: situations in which the basic structure of the topics to be acquired already exists, and what is now being done is the steady accumulation of knowledge about that topic. In an earlier paper, Rumelhart and I (Rumelhart & Norman, 1978) suggested that there were at least three different kinds of learning: accretion, learning, and structuring. The chapter by Anderson et al. essentially deals with accretion, the accumulation of knowledge (and perhaps a little with tuning, the making more efficient the use of knowledge). A major aspect of the learning of a complex topic is the development of a new conceptual framework within which to interpret the new information. The studies conducted by me and by my students indicate that learners actively interpret and reinterpret the information before them. And so I see that some of the major activities that learners perform are not covered by the approach of Anderson et al.

Consider the lessons of the chapters by Stevens and Collins and By VanLehn and Brown. These chapters told us about the use of conceptual models, how these models were used both to guide the acquisition of new knowledge and also for the understanding of a current situation. It is this aspect of learning that was completely absent from the work reported by Anderson et al.

I pick this problem for a simple reason: My own approach to the study of learning was once very similar to that of Anderson et al., but it came to grief. It

simply was not adequate to explain what students did. Students built models of a situation. Students went far beyond the information presented to them in their attempt to construct a sensible interpretation of what they had experienced. The students' underlying conceptual structures were not formed by simple additions of information as they read through the textbook, but rather grew as they formed new conceptual models. Information that seemed discrepant to the model was ignored. Sentences, even paragraphs, of the text seemed to be skipped. The student was an active learner, constructing models to describe what had happened, forming new structures, and not at all behaving in the simple, sensible way that I had postulated. (A detailed analysis of the active model building in which the student engages is provided by the thesis of Bott, 1978. More informal descriptions of these results and also of our overall studies are provided by Norman, 1980a, 1980b.)

Overall, I think that Anderson, Kline, and Beasley are pursuing an important objective, one that will lead toward a theory of complex learning. I think the work they have done has moved us in a correct direction, but it is insufficient. I believe the major deficiency resides in the treatment of the human learner as a systematic collector of information, adding a new knowledge structure here, generalizing there, always interpreting what is happening in a nice, systematic manner. Real subjects simply do not learn that way, at least not for any long length of time, not when there is complex material.

MENTAL IMAGES:
THE CHAPTER BY KOSSLYN AND JOLICOEUR

Kosslyn and Jolicoeur clearly and explicitly spoke about representation. I thought the chapter an interesting one. A few years ago I was quite unsympathetic to this line of approach. I thought the arguments too simple. I thought the arguments about propositional and imaginal representation missed the point. I am pleased to say that I have changed my mind, become converted, if you will, by the overwhelming weight of the evidence. Kosslyn and his colleagues have demonstrated important results and put them together in a nice, cohesive package. This doesn't mean that I am necessarily happy with everything in that package, but the experiments force us into considering some sort of imaginal representation. I believe that the chapter by Kosslyn and Jolicoeur makes an important contribution, far more important than any single experiment in the overall picture: a melding of different representations for different purposes.

Kosslyn and Jolicoeur examine a major unsolved aspect of representation: the representation of mental images. Here, the goal is to determine the role of imagery in mental operations and the kind of representational systems that are necessary to account for people's performance in a variety of tasks. The authors

suggest that a dual representation is required—one propositional, one imaginal—with a "race" sometimes occurring between the two systems.

An interesting offshoot of the study of images is the development of a test for individual differences. If individual differences in imaging reflect major differences in processing strategies, then development of sensitive tests could prove to be an important tool for many different aspects of psychology. There already exist numerous tests for the ability of a person to form and use images. These tests, however, are not very convincing, primarily because they seem to have no obvious contact with our understanding of the underlying mechanism and process. The nice thing about the work of Kosslyn and Jolicoeur is that they have attempted to develop a test based upon an understanding of the theoretical structures and representations of imaging: They attempt to test the underlying processes.

This is not the place to go over in detail the particular assumptions of the model presented by Kosslyn and Jolicoeur. Basically, they suggest that images are stored, essentially as a matrix of light-intensity values (a "dot" matrix) represented in polar coordinates. Each image is available in a "file." The files themselves are then represented in some sort of propositional representation, perhaps within a network representation. When a question is to be answered, one must find the appropriate image information, then generate the image from the stored representation, and finally search the image for the information appropriate to the question.

Personally, I do not take all this too seriously; the models being proposed must be but a first approximation. The present formulation of models strikes me as much too simple. Few would quarrel with the notion that we use multiple means for representing the information about the world, that some of this information leads to images, and that the images themselves can then be manipulated and inspected. We do not quite know what an "image" is, but it is unlikely to be a simple matrix of points that are mentally illuminated within the recesses of the mind. So what? At this early stage of research, we need not take the preliminary models all that seriously. The framework of the model is a useful direction in which to move.

Studies of imagery have been hampered because there have not been reasonable models on which to base one's interpretations. We had good models of propositional representation but none of other formats. The recent work by Kosslyn and his colleagues provides us with a sensible starting place. People who disagree with the format now have a chance to attempt to improve it. Moreover, as the title of the chapter indicates, with a decent model of mental imagery, one can then examine the various parameters and aspects of the model, asking how they might be differently reflected in different populations of peoples. In this way, one can make a start toward the true analysis of individual differences in mental imagery.

OVERVIEW

I have reviewed four chapters on four different topics. I have voiced many complaints. But I could not have asked for a better set of chapters to review. They represent four of the most important new directions of research now being taken. All represent extremely important directions in our study of mental processes, leading not only to better teaching and learning but also to better understanding of human information-processing systems. Do not let my complaints cause you to miss this point.

I believe that we are entering a new phase of understanding of cognitive systems in general, about the mechanisms underlying learning, and about appropriate strategies for instruction. The work of Anderson, Kline, and Beasley starts us toward the development of complex learning, examining the changes in mental structures that occur during the course of the stages of accretion in learning. The work of VanLehn and Brown tells us about the role of a prior model in the learning process, formalizing the process of learning by analogy. The work of Stevens and Collins tells us something about the way that people use conceptual models in order to understand a given topic matter and to extend their knowledge to new aspects of that topic. And the work of Kosslyn and Jolicoeur tells us something about the representation for images, the long-neglected aspect of representation.

ACKNOWLEDGMENTS

REFERENCES

Bott, R. *A study of complex learning: Theory and methodologies.* Unpublished doctoral dissertation, University of California, San Diego, 1978.

Norman, D. A. What goes on in the mind of the learner? In W. J. McKeachie (Ed.), *Cognition, college teaching, and student learning.* San Francisco: Jossey-Bass, 1980. (a)

Norman, D. A. *Human learning and memory.* New York: Scientific American, 1980. (b)

Rumelhart, D. E., & Norman, D. A. Accretion, tuning and restructuring: Three modes of learning. In J. W. Cotton & R. Klatzky (Eds.), *Semantic factors in cognition.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

23 Models of Concept Formation

Richard B. Millward
Brown University

The position taken in this review of experiments on concept formation is a general one, intended more to comment on the field as a whole than on any one position in particular. Although researchers have advanced the study of the meaning of *concept* significantly during the past 5 years, many of their conclusions are incompatible. The purpose of my analysis is to show that although individual researchers argue that their own experiments have been definitive and have made untenable some other positions, no one theoretical position is broadly enough based to handle all existing results adequately. Further, there is generally a sense of correctness about each position, because no single position rests solely on a single experimental paradigm, and because each position is based on some intuitively reasonable psychological principle. A more comprehensive theory must explain all the experimental results and the general principles. The answer to the question "How are concepts formed?" will be, not one of the current theories, but rather a theoretical structure that accounts for aspects of existing theories that are correct. It would be ad hoc and unparsimonious to construct a general theory that was simply an amalgamation of existing positions.

Another purpose of this review is to relate concept formation studies and semantic memory studies. At present, the relationship is rarely expressed; when it is, it is rather vague. For example, in the semantic memory literature, a distinction is made between type and token nodes. A type node represents an abstract definition of a concept, and a token node points to a type node to represent a particular use of a concept. Yet the type node is not defined except in its relationship to other nodes in the net. In the literature on concept formation, models of how a class of items becomes represented as a concept are discussed,

but relationships between concepts are not mentioned. One aim of this paper is to suggest how a single, unifying theory might be constructed.

There are currently four major theoretical positions on concept formation. The models of these positions can be labeled *prototype, exemplar, frequency,* and *rule* models. (There are also differences within each approach; these differences are also treated, although not in extensive detail.) It should be noted that these four classifications are those of the author and not necessarily those of other researchers in the area. Further, the work of some researchers is interpreted according to more than one of these models.

PROTOTYPE MODELS

During the last decade there has been great interest in and support for the idea of a prototype. Historically, the idea goes back at least to Bartlett (1932/1967). Attneave (1957) and later Posner and Keele's (1968) experiment on the genesis of ideas are responsible for its current revival.

In the typical prototype experiment, subjects memorize or learn to classify a small number of examples of one or more concepts. After the learning task, the subjects are presented with a large set of test instances, some of which have been presented previously and some of which have not. Sometimes instances are presented that are not examples of any of the concepts. The subjects classify these test items, and the correctness of their decisions is recorded. Often a second measure is taken in order to estimate the ease with which the subjects made the classification response. The latency of the response is one such measure; another is the subject's confidence in his or her own decision. In general, old and new examples receive high confidence ratings, sometimes with and sometimes without distinction between old and new examples. Instances that are not examples of a concept receive much lower ratings.

Average-Distance Model

Posner and Keele (1968) defined a prototype as the stimulus with the average value on each dimension of a set of stimuli varying along a number of dimensions. Such a stimulus is the stimulus least distant from all the other stimuli. Their experiments involved three concepts and four examples of each concept. Each concept consisted of a random dot pattern as a "prototype" from which the examples were generated by random displacements of the dots according to different levels of distortion. Posner and Keele's subjects made fewer errors to the prototype from which the examples were derived than to other equally unfamiliar stimuli. They also remembered the specific instances presented better than new ones and showed an awareness of the amount of variability around each

prototype. From these results, Posner and Keele concluded that subjects computed an abstract stimulus similar to the prototype from which the examples were generated.

The underlying principle of the average-distance model is that the values presented in different exemplars are averaged, forming an average case to which other instances are compared in order to categorize them. Reed's (1972) various experiments used Brunswik faces, meaningful stimuli that were clearly dimensionalized. He refined the whole concept formation analysis by using distance measures derived from both physical and psychological spaces. The physical measures on the various dimensions may not be linear with the psychological values, and therefore an average computed with the physical measures alone may be incorrect. Thus a psychological scale is important in testing a prototype model.

Another of Reed's innovations was to assume that the weights of the different dimensions could vary. For example, subjects may find the nose more salient in faces than the eyebrows and consequently may judge an instance with a nose length similar to that of the prototype as closer to the prototype than is an instance with eyebrows similar to the eyebrows of the prototype. With these refinements, precise tests of model predictions could be made. Although Reed's innovative approach was successful, it proved very difficult to discriminate among the various models because they all fit the data about equally well. Reed found that an average-value model and an average-distance model were about equally good at predicting the confidence ratings made by subjects. (See the later section on "The Exemplar Model.")

The Integration Model

An alternative conception of a prototype was introduced by Bransford and Franks (1971). Here the prototype was a proposition expressing a complex concept. For example, the sentence "The ants in the kitchen ate the sweet jelly which was on the table" can be broken down into simple propositions ("The ants ate the jelly"; "The jelly is sweet"; and so on, into complex ideas consisting of two propositions ("The ants ate the sweet jelly"), or into complex ideas with three propositions ("The ants in the kitchen ate the jelly which was on the table"). Four different complex concepts (propositions) were presented in the form of two sentences from each of the one-, two-, and three-proposition concepts. Later, for each concept, the six sentences presented earlier (old sentences) and the remaining six possible sentences (new sentences) were presented in a recognition task that asked the subjects to indicate whether they had seen the sentence before and how confident they were in their recognition. The subjects were unable to distinguish old from new sentences that were conceptually correct, but they rejected nearly perfectly propositions consisting of a combination of ideas from different

concepts. Further, the more propositions a complex concept contained, the more familiar a subject found it to be, even though the complete concept (all four idea units) was never presented.

In an earlier study, Franks and Bransford (1971) had used visual abstract forms as stimuli in a concept formation study. Their model for that experiment also assumed the formation of a prototype. Judgments of whether a stimulus had been presented before depended on the number of elementary transformations required to generate the stimulus from the prototype; that is, the fewer the transformations required, the more familiar the stimulus. Thus, a major assumption in Bransford and Franks' work is that the distance of an exemplar from the prototype is measured in discrete units—the number of ideas or the number of operations required to transform the prototype into the exemplar.

Implicit Learning

Another experimental situation that illustrates the breadth of the idea that a prototype is abstracted from memorized information is Reber's (1967) implicit-learning paradigm. (See also Evans, 1967, for a definition of schema theory, which emphasizes this same aspect of concept formation.) Explicit rule learning is a hypothesis-testing situation. (See the later "Rule Model" Section.) In Miller's (1967/1969) project Grammarama, one of the inspirations for Reber's study, Miller instructed subjects to learn the rules of a grammar by generating exemplars. He found that subjects were poor at discovering the underlying set of rules. Reber felt that it was important that Miller's subjects did not learn the rules of the grammar even though they were explicitly instructed to do so. Reber's thesis was that subjects learned implicitly while learning the items generated by the grammar, not by trying to discover the rules directly.

Reber's experimental task required subjects to reproduce strings of letters generated by a finite-state grammar (Chomsky & Miller, 1958). Fifteen grammatical strings varying in length from three letters to eight letters were presented three strings at a time. The subjects were required to reproduce all three strings correctly twice before going on to the next set of three strings. These strings were learned much faster than strings of random letters, but the facilitation did not show up until the third set of strings. After all the strings were learned, the subjects were given a recognition memory test for old grammatical, new grammatical, and nongrammatical strings. Subjects performed better than chance, but, more importantly for Reber's hypothesis, they could not make sensible statements about the rules of the grammar or the constraints on the letter strings. Reber concluded that the subjects were not learning by hypothesis testing but rather that they learned the grammar implicitly.

Like Bransford and Franks' experiment, Reber's study used only one category of items, so the items themselves were learned and not just associated to some response class. Of course, Bransford and Franks used existing schemata to or-

ganize incoming information, whereas Reber emphasized the abstraction of a schema (grammar rules) from the input. This appears to be a difference between the paradigms, but the theory proposed later accounts for this difference, at least in principle.

In Summary

We have seen the formation of integrated structures with four different types of materials: random dot patterns, dimensionalized faces, propositions, and letter strings. A prototype model must be applicable to all these different materials, or else we must postulate different mechanisms for what appears to be the same phenomenon. Posner and Keele (1968) have made a strong case for the extraction of an entity that is an average of the input instances. But they also noticed that subjects remembered individual exemplars and encoded variability.

Reed's (1972) results are interesting in that he can place his stimuli into a four-dimensional space either on the basis of the physical dimensions of the attributes themselves or on the basis of multidimensional scaling. Presumably, we can multidimensionally scale everything and therefore give all stimuli a psychological space. If that is correct, then the prototype model may be correct. However, the idea of scaling letter strings and/or propositions does not seem reasonable, and therefore the use of scaling is not a general solution.

Bransford and Franks (1971) also emphasized that various exemplars were integrated into a single memory structure, although, since they were dealing with propositions, this could hardly be an average. They also assume that subjects abstracted transformations from their experience with a set of stimuli. These transformations provided a mechanism for judging the similarity of stimuli to the integrated memory structure, thus accounting for the subjects' differences in confidence ratings and for their success in recognizing new instances that were legitimate transformations of the prototype.

The Reber study points up another characteristic of all these experiments—namely, that subjects are unaware of the basis of their responses. His experiment also provides another example of concept formation, albeit not one that lends itself easily to prototype theory; we include it here because of the similarity of his experimental results. Reber claims that the implicit learning of grammatical rules is like learning legitimate transformations. It certainly appears unparsimonious to use different models for all these results.

THE EXEMPLAR MODEL

The exemplar model presents a strong contrast to the prototype model in that rather than assuming that information is integrated in some fashion, it assumes that each presented instance is stored as a unique memory item. When new

instances are presented for classification, they are compared to the individually stored memory items. Two procedures have been suggested for this comparison process: One, the *average-distance* procedure, measures the distance of the new stimulus from the stored exemplar of a given category and computes an average distance from the given category. The item is then classified as belonging to the category with the smallest average distance. The second, the *nearest-neighbor* procedure, compares the new exemplar with all the stored exemplars and assigns it to the same category as its closest match. In either case, if the exemplar model is to work well, each exemplar that is learned must be stored in memory. The model does not, however, require that every feature of each exemplar be stored.

One argument in favor of the exemplar model is Posner and Keele's finding that subjects remember old dot patterns better than new ones, even though the old and the new patterns are equally distant from the prototype. Recent studies (e.g., Griggs & Keen, 1977) using the Bransford and Franks paradigm have provided evidence that subjects can distinguish between the old and the new sentences. Further, in the Reber paradigm, the old finite strings are judged grammatical with higher confidence ratings than are the new ones. On the other hand, Hyman and Frost (1975) found little evidence to support the nearest-neighbor version of the exemplar model.

Two aspects of the exemplar model are counterintuitive. One is the fact that it requires very good memory for all the learned exemplars. Many experiments have demonstrated that memory for specific items is very poor. A second problem is the amount of computation required in order to make a decision. Of course, if this computation is done in parallel and by some analog process, it is not impossible.

Learning Individual Items in Concept Formation

Reed (1978) has recently introduced an experimental procedure to determine whether concept formation requires learning individual exemplars, or whether features common to a category can be extracted from exemplars before the exemplars themselves are learned. This is important because, as mentioned above, Reed (1972) could not clearly decide between an average-value prototype model and an exemplar model by using statistical procedures. His idea was to make an experimental distinction rather than rely on the goodness-of-fit technique. He introduced a paradigm that mixed both paired-associate learning and concept formation. Outside the laboratory, we usually learn examples as well as categorize them, and thus this mixed paradigm provides a more realistic test of natural concept formation.

In terms of traditional approaches to concept formation, Reed is distinguishing between discrimination processes and generalization processes. The former processes assume that when first encountered, many items will look the same, and cues must be discovered to differentiate one item from another (Gibson &

Gibson, 1955). The latter processes assume that objects are differentiated from the beginning and are later combined into categories on the basis of common characteristics (Bruner, 1957). Reed's new paradigm requires subjects to learn to identify each item of a list of items as well as to classify them into categories. Reed argues that since an exemplar model requires that individual items be stored in memory so that an item to be classified can be compared to them, identification must occur before classification. If classification learning occurred before or independently of identification, the exemplar model could be rejected because this would imply that a common prototype was being formed upon which decisions were made before there was differentiation among the individual examples. We look at Reed's experiment in some detail because it is an important issue.

Reed (1978) used Brunswik faces, as before, which vary on four dimensions. They were divided into two categories so that they were linearly separable. This means that they formed distinct prototypes but does not mean that there was a simple rule for classification. Reed ran four different experiments, which varied primarily with respect to the task given to the subjects. Two kinds of learning were possible: (1) identification learning in which each of the 10 faces was assigned a unique response (one of the numbers 1 to 10); and (2) classification or categorization learning, which required that each face be assigned to one of two categories. The experimental groups were:

Exp. I: Each trial required only categorization responses.

Exp. II: Each trial required two types of responses, identification and categorization. The faces in one group were given responses 1 to 5, and the faces in the other group, responses 6 to 10, thus allowing an identification response to indicate what class the item belonged to and reducing the number of response choices if the category was known.

Exp. III: Odd trials required category responses and even trials required identification responses. The faces of one group were given responses 1,3,7,8, and 10, and the faces of the other group, responses 2,4,5,6, and 9.

Exp. IV: The response assignment and the procedure were the same as Exp. III, but the faces were mixed up so that they were not linearly separated. Thus, categorization based on prototype was impossible because the two prototypes were nearly identical.

To analyze the two learning processes more precisely, Reed determined the exact trials on which identification and categorization learning occurred for each item. He did this by assuming an all-or-none learning process and using maximum-likelihood estimates on the sequence of error and correct responses. Hence, for each item, he could say whether it was identified first or classified first.

The results of the four experiments provide answers to a number of questions. First, identification learning was faster in Exp. II than in Exp. III or Exp. IV.

This can most easily be explained by assuming that subjects can guess better about the identification response in Exp. II than in the other experiments, because knowing the category to which an item belongs reduces the number of possible alternatives from 10 to 5. Of course, if the subjects learned to identify items first, it would also help them classify the items more easily. But there was no difference in categorization behavior between Exp. II and Exp. III, indicating that no advantage was gained for categorization by the response redundancy.

One other problem remains. Although Reed found categorization learning faster than identification learning, this difference could be accounted for by the fact that the latter required selecting 1 response from a possible 10 (or 5, if the category were known), rather than from only 2. To demonstrate that classification learning was in fact faster because of the abstraction of a prototype, the categorization results of Exp. III and Exp. IV were compared. In Exp. IV, no prototype formation could occur because the faces in the two categories were not separable, yet the categorization process still required only two responses. The results are clear: Exp. IV showed just as good identification learning as Exp. III but very much poorer classification learning. On the basis of these results, Reed rejected the exemplar model in favor of a prototype model.

Using Learned Items in Classification

A series of experiments by Brooks (1978) clearly illustrates how an exemplar model can be used to classify new examples. Brooks' logic is much the same as Reed's except Brooks employs it to demonstrate that learned individual items can be used to classify new items, whereas Reed employs the same logic to disconfirm the need to learn individual items. Brooks was interested in the contrast between the usual concept-learning paradigm and Reber's grammar-inducing memory paradigm. (Note that in most prototype experiments, the initial training is in a classification task, not a memorization task. The Reber paradigm requires the subject to memorize the stimulus items, not simply to classify them as in the usual concept-learning paradigm.)

As the stimuli in a paired-associate learning procedure, Brooks used letter strings generated by two different finite-state grammars. As responses, he used either city or animal names unrelated to the string dichotomy. After the paired-associate learning, a recognition task was given to see if the subjects had any knowledge about the two string types. They did not. Subjects did notice that some responses were cities and others animals but did not notice another feature of this experiment—namely, that the responses (cities or animals) could be further divided into New-World (*Chicago-moose*) or Old-World (*Cairo-baboon*) items and that this difference corresponded perfectly to the two grammars from which the strings were selected. In theory, the subjects could have learned this relationship if they had noticed either of two differences—the difficult discrimination between the stimuli or the easier Old-World/New-World dichotomy. (This latter difference was of course masked by the city-animal contrast.) Brooks then asked

his subjects to sort new letter strings drawn from the two grammars into their correct categories. The subjects were told about the Old-World/New-World difference and then were given the new strings. They performed very well, about as well as Reber's subjects in the implicit-grammar study. Brooks concludes that the good performance resulted from a comparison of test items to individually learned stimuli. He assumes that new stimuli were matched to those learned earlier as paired associates, and when a match was found, the response associated with the matching stimulus was used to determine the correct category. Brooks argues that one way to form a concept is by comparing new instances to individual items in memory, and inferring the associated response by analogy. It is an interesting experiment and a good demonstration of how the individual item can mediate categorization processes.

Brooks' general argument is that the analytic structure of most concept-learning experiments has caused researchers to miss the more natural way we learn concepts. In most concept-learning experiments, subjects analyze a series of individual examples, whereas in "real life," we tend to learn a great deal about a few examples. In other words, the role of memory for individual items has been ignored in concept-learning experiments. Brooks also makes the point that individual objects can be classified into a number of different categories simultaneously, a process not typical of concept-learning procedures but a natural process for individual objects in real life. Because one does not know how an object will eventually be classified, all categorization is done immediately. Thus, the ability to store information about an individual item is important if this information is to be used later.

Reber (1976) recently has reported a somewhat surprising result that deserves further discussion in light of Brooks' thesis. Reber ran two groups of subjects who differed in th. type of instructions they received—explicit or implicit. Subjects in the explicit group received instructions emphasizing that the strings were patterned and that finding the rule that governed the generation of the strings would help them learn the strings. In the instructions for the implicit group, no mention was made of the fact that the strings were structured. Both groups learned five sets of three strings each and then were given a recognition test on new strings. Recognition performance was well about chance for both groups. The interesting point is that the explicit group did not do as well as the implicit group. Reber explains the results by assuming that the explicit instructions interfered with the implicit-learning process of abstraction. Apparently, one can disrupt the formation of prototypes by introducing a hypothesis-testing set.

We replicated Reber's experiment in our laboratory and added a new condition based on the "observation" technique (Reber & Millward, 1968). We did not get exactly the same results as Reber but perhaps provided an even more interesting variation to consider. We used 27 sentences, presented 3 at a time in 9 learning sets. Again, implicit and explicit instructions were given. The proportion of correct responses for the explicit group was .79, and the proportion correct for the implicit group was .74. A third group was also run, an observation

group for which the 27 strings were presented for only 3 seconds each. On one-third of the trials, each presentation was followed by an instruction to recall the string just presented. This was essentially an immediate memory span paradigm. The 27 strings were repeated 3 times to form a block of 81 trials, with 9 different strings given a recall request on each pass through the list. The block of 81 trials was then repeated once more, so that subjects recalled each string from immediate memory only twice. By most standards this would not be considered a very deep level of processing. Nonetheless, this group of subjects had a proportion of .70 correct responses, nearly as good as the other two groups.

These results certainly support Reber's notion that recognition memory is good even without explicit hypothesis testing, but they raise a question about the kind and amount of information required to perform above chance in a recognition experiment of this type. Does the fact that subjects can learn to categorize from minimal stimulus processing contradict Brooks' thesis? In one sense it seems to, but Brooks' main point, like Reber's, is that concept-learning experiments produce an analytical set that inhibits learning individual items naturally. It is possible that the observation technique does not inhibit such learning and, even though the depth of processing is limited, that it is sufficient to allow subjects to pick up enough cues to perform well in later recognition tests. Of course, we are not sure exactly what kind of cues are picked up, so we do not have a good understanding of how subjects make such classification decisions.

One final comment: Subjects in even the implicit and observation groups perform some analysis on these strings. Our college students left the experiment with all kinds of hypotheses and "facts" about the strings. These facts did not always correspond to the truth, and the subjects performed in ways that indicated that their decisions were made independently of any of their hypotheses. Nonetheless, it was clear that hypotheses existed and some explicit information had been encoded.

FREQUENCY MODELS

Frequency models emphasize a memory structure that reflects the frequencies of features in the exemplars studied. Unfortunately, the definition of a feature is not very sharp; to test the model, ad hoc assumptions must be made about what aspects of the stimulus are stored and counted. Further, in frequency models, the rule for classifying new instances is not always determined by the assumptions about storage. Nonetheless (or perhaps because of its flexibility), this class of models has received substantial empirical support.

The Cue-Validity Model

The cue-validity model computes the conditional probability of a category given a cue (or feature or value on a dimension). Conditional probabilities are summed

over all features of an instance to determine the probability that that particular stimulus belongs to a given category. In a classification task, a new inst ance is assigned to the category with the highest conditional probability. Notice that this model does not provide a distance measure for the items classified within a given category. (One could get a distance measure by considering the likelihood ratio, but I do not think this has been done.) Reed introduces a priori probabilities to take into account the different frequencies of presentation of different cues in different categories. The conditional probability can be 1.0 if a cue is presented 10 times and is always followed by the category, or if the cue is presented only once and is followed by the category. The use of a priori probabilities makes less frequent cues less important in the final decision. Reed found the cue-validity model a poor fit and has discarded it. However, frequency models are very special, and their failure in a given instance could be due to inappropriate selection of cues.

The N-Gram Frequency Model

Reitman and Bower (1973) developed a frequency model to account for Bransford and Franks' results (see the section on "The Integration Model"). Reitman and Bower were critical of a number of procedural details in Bransford and Franks' experiments. First, by presenting many similar ideas more than once, Bransford and Franks increased the interference effects of their items. When new instances consisting of recombinations of old concepts were presented, their formal similarity to the old ideas and their high familiarity made the new instances appear old. Second, the more ideas that were presented together, the more likely it was that the subject would recognize old concepts and thus increase his or her confidence ratings. In other words, one does not need to talk about prototypes and distance measures to account for the Bransford and Franks results. Using simple letter strings, Reitman and Bower were able to partially reproduce the Bransford and Franks results. Reitman and Bower's model assumes that each letter (a 1-gram), each pair of letters (a 2-gram), each triplet (a 3-gram), etc., was encoded in memory and received an increment in strength each time it occurred. If a new stimulus contained any new single letters (1-grams), subjects rated the stimulus new with a high confidence. Otherwise, they rated it old with a confidence that depended on the strength of the n-grams encoded in memory. Despite its problems and its ad hoc status, Reitman and Bower's model demonstrates that prototype results can be obtained by assuming only that frequency information about features is stored and that the integration process is possibly not necessary.

The Attribute-Frequency Model

Perhaps one of the most important experiments supporting the frequency models is that reported recently by Neumann (1977). Neumann believes that subjects

form prototypes, but prototypes based on the frequency of cues presented and not on the average of values along a dimension. For example, if the values of the experienced instances form a circle in a two-dimensional space, the prototype would be the center of such a circle, whereas the prototype based on the modal frequencies of the presented cues would be at some point on the circumference of the circle.

In a series of experiments that manipulated the discriminability of dimensions along with the frequencies with which cues were presented, Neumann showed that the mode and not the mean was the determining value for later confidence ratings. He also introduced a new way to consider the effects of continuous variables. When a specific value on some continuous dimension is presented to a subject, that particular value is obviously not the only one encoded because subjects cannot discriminate that precisely. Neumann assumes that each cue presented is actually encoded as an interval on the continuous dimension. The importance of this idea is that when features on a continuum are very close together, there is a tendency for adjacent cues also to receive a frequency increment, leading to prototypes formed at the center of a continuum rather than at the extremes. This analysis could account for the success of the prototype models. Reed's results, for example, could be drastically reinterpreted if principles expanded by Newmann's study prove applicable to Reed's study.

Family Resemblances

All the experiments discussed so far, as well as almost all learning experiments, have implicitly assumed a definition of "concept" that depends on common elements. Yet Wittgenstein (1953) pointed out that concepts cannot be satisfactorily defined in terms of common elements. Almost any attempt to do so for natural concepts is easily contradicted with some special example. Alternatively, one might consider a concept as being the union of a number of sets of intersecting elements. Thus, each instance has one element, and usually more, in common with some other instance, but no element is common to all instances and absent from all contrasting categories. This idea is the basis of Rosch and Mervis' (1975) notion of family resemblances (see also Rosch, Mervis, Gray, Johnson, & Boyes-Broem, 1976).

A description of a specific procedure to measure family resemblances should make the ideas clear. Consider a set of representative examples of some category. The examples have features $a, b, c, \ldots$ Count the number of examples that have each feature. Call these counts $Fa, Fb, Fc, \ldots$ for features $a, b, c, \ldots$ Then give each example a measure of "goodness" by adding up the frequency counts of its features. If, for instance, an example has features $a, c, d,$ and g, its measure would be $Fa + Fc + Fd + Fg$. The example with the highest measure (of "family resemblance") can be called the prototype since; in one sense, it is at the "center" of the category. Rosch and Mervis have shown that

such a measure correlates highly with ratings of "goodness of example," which are made simply by having subjects rate each example on a 5-point scale.

A second principle of family resemblances relates each item of a category to other categories. It states that items with high family-resemblance scores in one category will tend to have low scores in other categories. In other words, the best examples of one category will not be good representatives of other categories. In terms of the features measured earlier, the best example of a category will have the most features overlapping with other examples of the category and the fewest features overlapping with the best example of other categories.

The relevance to concept formation of this theory about the structure of natural categories is based on two arguments. First, Rosch and Mervis are dealing with natural categories: Facts about natural categories must be pertinent to how concepts are formed. Second, they have applied this theory to artificially constructed categories with some success, indicating that the theory can be experimentally demonstrated.

To illustrate that frequency models and prototype models have much in common, Reed (1978) notes the high correlation between prototype ratings by subjects and the measure of family resemblance. However, Rosch and Mervis (1975) do not view these high correlations as supportive of a prototype model:

> However, in one sense, the purpose of the present research was to show that it is not necessary to invoke attribute intersections or higher order gestalt properties of stimuli . . . in order to analyze the prototype structure of categories. That is, even at the level of analysis of the type of discrete attributes normally used in definitions of categories by means of criterial features, we believe there is a principle of the structure of stimulus sets, family resemblances, which can be shown to underlie category prototype structure.

In Summary

A frequency model can be used for all types of features and for categories without common elements. It seems harder to apply the prototype models to the same wide range of stimuli unless one assumes that every type of stimulus is represented in some multidimensional, probably continuous, psychological space. However, the poor fits using the cue-validity model and the very interesting results in semantic memory using multidimensional scaling are not observations to be dismissed easily (Smith, Shoben, & Rips, 1974).

THE RULE MODEL

The concept formation experiments already described emphasized concepts that were so complex as to preclude simple rules. Also, in some cases at least, the

emphasis was not on categorizing the stimuli but on learning the items themselves. Even where categorization was involved, the similarity of the stimuli required the subject to learn a great deal about the features of the stimulus. However, subjects often approach a concept formation task by forming hypotheses about the rule governing the classification of the stimuli involved. This is a reasonable strategy because if such a rule exists or could be formulated, a subject can easily solve the classification problem once and for all. Also, a rule model of concept formation does not require that subjects formulate and test hypotheses explicitly; they might do so unconsciously. When stimuli are classified by simple rules and subjects are explicitly instructed to find the rule involved, the experimental paradigm is called *concept identification* (to distinguish it from the procedure in which a new concept must be formed). The main assumption of the rule model makes it particularly well suited to concept identification and problem-solving situations. First, subjects are able to generate hypotheses appropriate to the situation. Then, as exemplars are presented, subjects apply one or more hypotheses to them in order to decide how to classify them. When they receive feedback about the correct classification, the disconfirmed hypotheses are rejected. Bruner, Goodnow, and Austin (1956) and Hunt (1962) have summarized the early research on concept-learning and hypothesis-testing models. More recently, Millward (1971) has reviewed and Millward and Wickens (1974) have developed a general mathematical theory for this class of models.

Although hypothesis testing no doubt occurs in all these paradigms—concept formation as well as concept identification—it does not seem to characterize correctly the "learning" that occurs in most concept formation studies. First, a rule model requires the subject to formulate classification rules, whereas many of the tasks already discussed require only that the subjects learn to reproduce stimuli and not that they classify them. Even when the instructions explicitly ask subjects to classify stimuli in a concept formation study, the rule structure governing the grouping of the stimuli is so complex that subjects will have great difficulty in determining it. Second, hypothesis-testing models imply all-or-none learning; the gradual improvement generally observed in concept formation studies would be hard to account for by such models. However, if subjects built up a series of conditional rules—no one of which was totally correct, but each of which correctly handled some individual conditions—learning could progress gradually, even according to a hypothesis-testing model.

Third, the hypothesis-testing strategy is particularly explicit and yields declarative propositions upon which subjects can base their responses. In contrast, as Reber (1967) emphatically noted, much of the concept formation process seems to be implicit and to yield improved performance, the basis of which is unclear to the subjects themselves.

Hyman and Frost (1975) investigated the possibility that subjects were covertly testing hypotheses by adding to a set of Posner and Keele dot patterns a featural distinction that could be used as a rule to categorize the stimuli. Hyman

and Frost expected to measure a "natural progression" from exemplar models to rule models as a function of practice but found no evidence for such a progression. Instead, they found evidence, at all stages of learning, that the two categories based on two different prototypes were always analyzed better by two different models. One category was best described by an exemplar model at first and then a rule model, whereas the second category was best described by a prototype model throughout learning. In other words, the success of the model depended on the particular pattern used. Even when subjects were given instructions about the critical feature governing the rule, the category encoded as a prototype produced some responses appropriate for a prototype model and not consistent with a rule model. The Hyman and Frost results are far from definitive and, if anything, suggest that there are serious problems with all models because concept formation is not supposed to depend on the particular pattern used to generate the stimuli.

Hypothesis-testing experiments are inappropriate for studying concept formation for other reasons besides an inappropriate paradigm. Rosch and Mervis' demonstration that natural categories conform to family resemblances precludes describing the members of a category by a simple rule. A general rule defining a category would have to consist of the disjunction of a large number of conjunctive attribute sets. Carried to extremes, such a rule would then begin to specify individual exemplars and therefore approach an exemplar model.

Another, more fundamental failure of rule models is that the rule that defines a concept in hypothesis-testing experiments specifies the attributes of the objects conforming to the concept. But this is not the correct rule. The concept itself is the rule, and the concept is not a specification of the physical characteristics of the objects denoted by the concept. The concept defines the set of objects; the set of objects do not define the concept. We can, of course, define objects on the basis of their features, and the family-resemblance rule provides a mechanism for such classification. However, before fami'v resemblances are possible, the concept must be known and used to form   class of objects from which feature frequencies can be computed. Th   ʳⁱⁱ⅄   'hat the typical rule model is overly tied to features and provides a deₐ₁ₐᵢⁱon of a concept that is too stimulus bound. The clarification of the definition of a concept will take up most of the remainder of this chapter. It is critical to the theoretical position developed here.

A COMPUTATIONAL THEORY

The Functional-Core Concept

Nelson (1974) has presented a definition of a concept that does not depend solely on physical characteristics. She considers the concept of a *ball* and argues that

when a child sees a ball for the first time, that single experience with the ball is sufficient for the child to form a "ball concept." Associated with that experience will be actions, expectations, and other memories of an episodic type. These experiences add up to much more than a cluster of features concerning the object. The features are not completely separated from the actions involved in the use of the ball or from observations of the ball's behavior. All these relationships become an integral part of the concept of *ball*, its "functional core."

Nelson assumes that individual objects are recognized as objects before concepts are formed. She believes that a concept does not depend on there being more than one instance of the concept nor that multiple instances be distinguished one from the other. For example, a child rarely sees exactly the same bee but still has a concept for *bee* whereas he or she may see only one example of a dog, his or her own pet, and has a concept for *dog* also. The organizing principle for a concept is its functional core, and it is on that basis that objects and events are conceptualized. If such be the case, there are serious questions about what we are investigating when we study, for example, random dot patterns. Each instance of random dots is recognized as a set of unorganized dots. Distinguishing between any pair of them is difficult because of their highly similar components. What kind of functional core can subjects generate from such stimuli? Whatever it may be, if indeed there is one, it will probably not be the same for all subjects. If we accept Nelson's ideas, we have to question the appropriateness of the concept formation paradigm as an analogue to natural concept acquisition.

Nelson (1977) has more recently refined her ideas, relating them to scripts. She assumes three parts to the conceptualization process: (1) finding identifying attributes; (2) generating functional cores; and (3) establishing scripts. The identifying attributes allow a concept to be used out of its context. Finding the attributes is of course the focus of the classical definition of concept based on features and depends on repeated experience with the concept. Functional cores are basic to the meaning of events and things because they represent how concepts are used. Functional cores can be established on the basis of a single instance. Scripts organize the flow of concepts, providing an interrelationship among them. (More is said about scripts later.) One might consider the experimental literature reviewed thus far as being concerned with the attribute selection process. If that process can be isolated, then perhaps the other aspects of conceptual behavior can be ignored. The isolability of the attribute selection process is a stong assumption; even if it is true, we still have only one aspect of the process (according to Nelson's tripartite division).

In other words, as Nelson emphasizes, a concept is not a catalogue of the features of exemplars that meet the concept rule. In one sense, the concept is the rule itself. But a concept is more complex than a rule, as Nelson's three-part conceptualization process implies. Miller and Johnson-Laird (1976) have stated this emphatically:

The meaning of "book" is not the particular book that was designated, or a perception of that book, or the class of objects that "book" can refer to, or a disposition to assent or dissent that some particular object is a book, or the speaker's intention (whatever it may have been), or the set of environmental conditions (whatever they may have been) that caused him to use this utterance, or a mental image (if any) of some book or other, or the set of other words associated with books, or a dictionary definition of "book," or the program of operations (whatever they are) that people have learned to perform in order to verify that some object is conventionally labeled a book. We will argue that the meaning of "book" depends on a general concept of books; to know the meaning is to be able to construct routines that involve the concept in an appropriate way, that is, routines that take advantage of the place "book" occupies in an organized system of concepts [pp. 127-128].

Frames, Scripts, and Schemata

When one considers abstract concepts, the problems with prototype, exemplar, frequency, and rule models become more obvious. Consider such nouns as *room, restaurant* and *underdog* and verbs such as *break, throw,* and *give.* These words represent concepts in everyday life, but it would be very difficult to set up a traditional concept formation study to present them to subjects. Nelson has provided us with some hints about their representation. Each of these words describes one or more experiences. If a person stores the experiences associated with each situation, then after a number of experiences, that person might be able to put together a fairly complex structure that "defines" the concept. Current work in psychology, linguistics, and artificial intelligence converge in suggesting that experiences are coded by rather elaborate mental entities: frames, scripts, and schemata. Minsky (1975) defines a structure called a *frame* in order to describe the perceptual characteristics of a *room;* Schank (1975a, 1975b) uses a *script* to represent complex episodic events such as those associated with a *restaurant;* and Bobrow and Norman (1975) and Rumelhart and Ortony (1977) use the term *schemata* to define abstract nouns like *underdog* and verbs like *break, throw,* and *give.*

If entities like schemata can be found for concepts as abstract as *give,* it seems reasonable to believe that they can be found for letter strings, stylized faces, and dot patterns. But the latter are pseudoconcepts because there is no functional core, and the data that serve as input are rather poorly organized. In other words, random dots have no significance for the subject: There is no way for him or her to interact with them, and they do not form discrete objects with any meaning for the subject. Nonetheless, we shall assume here that the same kind of processes that are at work when we learn about rooms and restaurants are at work when we learn about stylized faces, random dots, and letter strings.

One problem with studying concepts as the abstraction of physical features is illustrated by Minsky's (1975) discussion of a birthday party. He took the follow-

ing example from Charniak's (1974) thesis. Consider the following story: "Jane was invited to Jack's birthday party. She wondered if he would like a kite. She went to her room and shook her piggy bank. It made no sound [p. 241]."

A number of inferences must be made in order to understand this story. We make them easily, in fact so easily that it is hard to imagine not making them. We know that birthday parties require giving presents; therefore, Jane is considering whether Jack would like a kite in order to decide what present to buy for him. Minsky argues that for such a concept as *birthday party*, we have special frames, sequences of expectations, actions, facts, rules of behavior, and so forth, and when the term *birthday party* is used, these frames are called out to be matched to the current state of the world. The match does not have to be perfect, and there are certain default options and backtracking procedures to be used when an initial assumption fails. The understanding system thus looks for certain kinds of contexts and uses frames to inject reasons, motives, and explanations for them.

The "frames" definition of a concept has no simple rule based on features, no set of defining features, and no prototype. A good portion of the concept could be built up from the memory of a single birthday party. The idea that we use individual cases as the foundation for a concept, much as Brooks and Nelson argue, makes a good deal of sense.

Although there are differences among the three ideas of *frames*, *scripts*, and *schemata*, certain general features are common to all three. [We frequently use the term *schemata* for all three because it is the most general term and has historical precedent (Bartlett, 1932/1967).]

1. All three concepts are data structures representing stereotyped situations. No very strict conditions are specified for either the structure or the type of situation for which they are intended. Thus, examples range from descriptions of objects, through situations and events, to sequences of events.

2. The data structure itself is nonatomic. It is hierarchical in a loose sense, or *heterarchical*, to use Minsky's term. Each organizational unit calls upon others as a team of experts might interact. Associated with each schema are cues indicating how to use it, what is to be expected from its use, and what to do if it fails. Each schema has variables that are matched to context at the time it is involved. If a match is impossible, a default value is determined, which depends on the values of other variables. Hence default values are not necessarily constants.

3. Schemata are organized by other schemata into structures that represent organization of knowledge. These organized structures assist in the use of concepts since the same assignment of values to variables can be made for a number of different schemata at the same time. Such organized structures are essentially schemata themselves, so the concept of a schema is recursive and dependent on other schemata to which it is related. This implies that one cannot talk about the definition of a concept without also specifying related concepts.

4. The selection and use of schemata is not a simple one of finding a match to some environmental context but is more like a problem-solving situation.

This discussion of schemata is related to three important psychological contrasts. One contrast concerns two types of memory—episodic and generic. A second contrast involves ways to process information—goal directed versus data driven. A third contrast concerns ways of storing knowledge—either as facts (declaratively) or as processes (procedurally).

Episodic and Generic Memory

Tulving (1972) introduced a distinction between episodic and semantic memory that has become fairly well accepted in studies of memory. Episodic memory is the record of experiences fixed with respect to time, space, and context. Semantic memory is knowledge of the meaning of words, although the context in which these meanings were acquired has been lost. Here we use the term *generic* instead of *semantic* to emphasize the abstract and universal characteristics of meaning as opposed to the specific meaning implied by episodic experiences.

The formation of a concept can be thought of as the transformation of information from an episodic to a generic representation. Schank (1975b) argues that the distinction between episodic and generic is fallacious and that all we have in memory are episodes. Schank's position is similar to Brooks' and emphasizes the use of examples in understanding. However, Schank also talks about scripts, which appear to be a kind of abstraction from many episodes. As such, they represent generic information and suggest some summarization, integration, and abstraction process at work. Schemata in general must be built up from a number of experiences and therefore require some "concept formation" process. Schank's denial of its existence seems completely wrong. Of course, positing an abstraction process does not imply that all episodic memory disappears. Insofar as we can remember details about specific experiences that are not part of the general concept, it is necessary to assume that specific experiences are given unique memory representations. Likewise, insofar as the individual experiences influence the general concept, it is necessary to postulate a process that builds schemata from episodes. Here we are actually restating the question posed by the studies reviewed above, but we are suggesting a more complex model for the abstracted structure. This model, the schemata, can subsume the theoretical models discussed earlier.

Goal-Directed and Data-Driven Processing

The second psychologically important contrast emphasizes how information is processed rather than how it is represented. Consider what is involved in understanding a sentence. One can begin by assuming that a sentence (S) consists of a

noun phrase (NP) followed by a verb phrase (VP); that is, S → NP + VP. Given such a representation, then, in understanding a sentence, the first step is to look for a noun phrase. Here the system is acting in a goal-directed manner.

If, on the other hand, the first words of the sentence are "The big ship . . . ," then the parsing might begin with the word *The* and continue to the word *ship*, coding these as *determiner* + *adjective* + *noun*, and then establish that this sequence is a noun phrase. The data entered into the system drive the parser until it organizes a recognizable structure. Of course, there must be a representation for a noun phrase if a noun phrase is to be recognized, so higher-level structures are required for either type of parsing.

These different parsing techniques become important when one tries to model systems that understand. When there is a great deal of structure in the information to be processed, then goal-directed approaches are most efficient because they direct the search. But when the structure is weak, mistakes of interpretation occur that require backtracking. The advantage of data-driven processing mechanisms is that the system is directed by the data themselves. On the other hand, because the data are not always presented in an organized way, there can be a great deal of unnecessary processing due to misinterpretation of some piece of data. Obviously, some combination of goal-directed processing and data-driven processing is required for situations beyond the narrow and highly structured concepts utilized in laboratory experiments.

Human knowledge seems to be organized in such a way that when people have to respond to stereotyped situations, they do so efficiently, when they have to respond to unfamiliar situations, however, they do so adaptively, despite the fact that they have no preconceptions about them. This flexibility in human information processing suggests a system that utilizes goal-directed processing when the situation is appropriately structured and data-driven processing when the situation is unstructured. Schemata are data structures suited for both kinds of processing. First, a schema can be very simple and appropriate for a very limited situation. For example, a student's knowledge of the *corpus callosum* might consist of a single fact: "It is a part of the brain." But schemata can also represent highly structured systems of knowledge—for example, the student's knowledge of *corpus callosum* after a course on the brain.

Schemata both direct the processing of data and act upon new input conditions. Once a schema is evoked, it begins to process any data that enter the system and also begins to look for appropriate information to complete its meaning. If no appropriate data enter or are found, then the schema is replaced by some other schema. In working on data, an active schema might evoke a number of schemata both at lower levels and at higher levels. When a situation is highly predictable and has been experienced repeatedly, a standard battery of schemata are elicited. There will be some organizational schema that, once evoked, will guide the further processing of the experience. Some of the schemata evoked may be specialized and applicable to a number of different situations. Unfamiliar

situations will tend to produce trial-and-error application of a number of specialized schemata, none of which are designed specifically for the experience. The combination of schemata is sufficient to produce adaptive behavior. One definition of concept formation, then, is that it is the process of organizing the evoked schemata into some higher-level schema.

Procedural and Declarative Knowledge

A declarative representation of knowledge is one that emphasizes the storage of facts. Representing knowledge declaratively means storing it independently of the use made of it. The advantage of such representation is that knowledge can then be used in a large number of ways without being changed. A procedural representation stores knowledge as programs, as specific routines that carry the function of the knowledge with them. (See Winograd's 1975 article for a good discussion of these ideas.)

In psychology there have been systems that take either the declarative or the procedural position as fundamental. In a way, a stimulus-response or simple associative theory is all procedural. We know only what we can do. Cognitive maps, on the other hand, are highly declarative, and Tolman was rightly criticized for "leaving his rat frozen in thought." Today, we have highly declarative systems (e.g., Quillian's semantic networks, 1968) and highly procedural systems (e.g., Newell's PS, 1972), with Anderson's (1976) ACT and Norman and Rumelhart's (1975) MEMOD providing mixes of the two conceptual schemes.

Although it is difficult to make exact comparisons among the different ideas recently proposed as theories of knowledge and understanding, one might think of Nelson's core as primarily procedural, and of the kind of information represented by prototypes, exemplars, features, and even rules as declarative. Frames, scripts, and schemata appear to be a combination of both the procedural and the declarative (Winograd, 1975).

Anderson (1976) has three criteria for distinguishing between procedural and declarative knowledge:

1. Declarative knowledge is all-or-none while procedural knowledge is partial.
2. Declarative knowledge is acquired suddenly while procedural knowledge is acquired slowly over time.
3. One can communicate declarative knowledge but not procedural knowledge [p. 117].

If we accept these three distinctions, we have to assume that concept formation as it has been reviewed here is primarily procedural: The subjects in these experiments form concepts in a way that seems to be partial, gradually acquired, and

difficult to communicate. Yet the models emphasize the declarative component of the resulting experience. It is important to note that there are procedural components associated with the prototype, exemplar, feature, and rule models, but these procedural components are not made very explicit in the models.

Anderson's ACT model takes particular note of the declarative-procedural distinction. He postulates a semantic memory consisting of concept nodes linked by relationships. These linked nodes represent facts known by the system. Anderson then introduces *productions* as procedural entities that act on the semantic network. A *production* is a pair of symbolic entities—a *condition* and an *action*. The *condition* part of a production represents some context or state of the system that, if realized, will cause the production to *fire*. The *condition* is realized (or matches the state of the system) when a property (or its absence) is present in active memory. A property is a particular value of a variable, an active node, or some active link between two concepts. When a production fires, its *action* is implemented. *Actions* are procedures or a sequence of procedures that modify semantic memory by activating a node, by building a new memory structure, by binding or unbinding variables, or by allowing transfer of information to or from memory.

Two further aspects of Anderson's model should be mentioned. One is that at any given point in time, only a small portion of long-term memory is active. From the active portion of long-term memory, a selected set of nodes are put on an active list (the ALIST), where they will not weaken in activity. The ALIST acts as a short-term memory buffer. Second, all matching productions are selected and placed on an APPLYLIST. More than one production can fire at one time. Thus, parallel processing is a part of the system.

Sketching Out a Theory of Concept Formation

A naive organism can be thought of as one without concepts appropriate for the environment it finds itself in. Concept formation is the process of learning appropriate concepts for a situation. A naive individual must be guided primarily by schemata that are not particularly appropriate to the situation. Since he or she has no higher-level schemata, his or her behavior is not integrated into efficient and purposeful organized sequences. The naive organism is generally data driven rather than goal directed. The events of experience are nonetheless encoded—all experience requires some kind of encoding—but the manner in which the information is encoded may be inappropriate for later retrieval and use. The "facts" stored declaratively are simple structures with no relevant connections to other knowledge structures. When an organism is in this naive state, learning is gradual, memory partial, and reconstruction of the experience nearly impossible.

After repeated interaction with any environment, an organism almost invariably becomes less naive. This change is brought about by a number of processes, one of which is concept formation. The definition of a concept being

proposed here is complex, and the process involved in forming a concept is also complex. Indeed, there may be a number of different processes involved in what we term *concept formation*. The delineation of these processes and the specification of how they interact are both goals of a theory of concept formation. What processes produce the increased efficiency of experienced subjects? One possibility is that existing schemata are modified so that they are more appropriate for the environment. A second possibility is that schemata are sequenced or combined into a higher-level schema appropriate for the new situation. Still a third possibility is that new schemata are created for the specific context.

Existing schemata are modified to make their application easier and to make them handle the situation more efficiently. When we enter a new culture, we often find the same general concepts, but they are applied in slightly different ways. Generally, the old concepts or schemata are adequate but need to be corrected. This means changing the parameters of the schemata. For example, our schemata for automobiles may undergo changes as a result of living abroad for an extended period of time. On our return, American cars seem very large. Almost everything else about cars is the same, so the adjustment is a minor change in the range and expectation of the size dimension. Such a change reflects recent experience and is an example of how concepts can be modified to fit new situations.

The second possibility, organizing a sequence of existing schemata into a hierarchical structure, also *implies* the third possibility, creation of new schemata. In the second process, existing schemata are organized to form a higher-level schema for use in some specific (stereotyped) situation. Schank's restaurant script contains schemata for "dinner" (or "lunch" or "breakfast") and one for paying for service received, schemata that presumably are already understood. The restaurant script also includes some unique other elements (waiting to be seated, ordering from a menu, tipping) peculiar to the restaurant schema. The creation of a schema for a restaurant requires building up a new schema, but it is one composed of old schemata as well as new ones. In the formation of the schema for a restaurant, the tipping schema would be constructed. Tipping is an interesting concept because it depends on a number of context-sensitive factors. There is an appropriate time for tipping, conditions when it is appropriate, rules for how much to tip, and the social convention explaining why it is done. One's first schema for tipping might be rather limited in definition and applied too broadly. Children might expect their parent to tip in a McDonald's restaurant, where no service was given at a table. At the same time, children's experience of tipping in a restaurant would not be generalized to tipping a taxi driver. In other words, simply adding more features to the existing concept is not adequate, since in some cases, one has to generalize and, in other cases, restrict the application of the schema.

The computational theory discussed here can be presented in terms of Anderson's ACT model. Rather than schemata, ACT has subsets of productions that

fire in sequence, depending on conditions in semantic memory. ACT explicitly separates the procedural and the declarative, whereas the computational model creates a data structure containing both. The computational model's modification of schemata would seem to correspond to simple modifications in ACT's productions. Combining schemata is parallel to a recombination and integration of productions. Creating new schemata is equivalent to the creation of new productions. In the following sections of the paper, an interpretation of the four models already discussed (the prototype, exemplar, frequency, and rule models) is given in terms of the computational model. To do this, ideas from both ACT and schema theory are used for illustrative purposes. Both the experiments reviewed earlier and the theoretical explanations they suggested to account for concept formation are diverse: Stimuli may consist of random dot patterns, faces, letter strings, propositions, and natural categories. Theoretical support exists for prototype, exemplar, frequency, and rule models. Is our choice limited to these theories? Is it the case that different theories apply to different experimental situations or stimuli? Are the theories simply equivalent versions of each other, or do they make alternative predictions? The computational theory suggested here, albeit not very precisely stated, is potentially able to handle these different results and give an explanation for the different theoretical positions.

One of the difficulties in dealing experimentally with such diverse stimulus material is that the subject already has schemata suitable for many of the stimuli. The Bransford and Franks propositions are easily integrated because of existing high-level schemata designed for integrating meaningful material. The existence of schemata for faces was also evident in Neumann's experiments, and letter strings surely evoke some schemata, although the violation of the rules of English orthography probably leads to less than coherent structures. There are probably no preexisting schemata for dot patterns, but the fact that old dot patterns are recognized better than new ones attests to some kind of retrievable storage.

The initial process of analysis varies in all these experiments, so the degree of elaboration of the material in memory is different in each case. All the studies have in common a lack of meaningful encoding schemata of a high integrative level. (The Bransford and Franks propositions are an exception, but the Reitman and Bower study can be substituted for their study since it is similar in design but different in stimulus material.) Thus, only partial, unorganized, declarative meaning structures are likely to result. If more exact and complete codes were available, then memory for individual items or some integration of items would be better. When new items are presented for recognition, enough information exists about past experiences to allow better than chance, but not perfect, classification and to allow some, but not all, old items to be remembered.

An important feature of such experiments is that they usually provide a fairly long learning process. During this learning phase, there is an opportunity for the three processes of schemata change to take place. Where schemata exist, they

must be given new parameters and be modified to some extent. For example, the face stimuli present "new faces" that are coded by existing schemata with the wrong set of parameters. These parameters are shifted with practice, providing a modified set of schemata or copies of existing schemata modified for this particular experimental situation. (Adaptation studies in speech perception may be doing a very similar thing.)

String stimuli are less well coded, yet low-level schemata may exist for them. In learning string stimuli, the major modification of the system might be integration of existing low-level schemata into higher-level schemata. Reber (1967) emphasized the process of finding the "phrases" of the grammar—that is, the repeated subcomponents. Each subcomponent might represent a schema built up on the basis of letter patterns.

Visual perception is so overdeveloped that no stimulus is new enough to be responded to by primitive procedural processes. Schemata of some type already exist. Yet there are no highly organized schemata for random dot patterns, so their processing is rather uncoordinated. The major focus of learning may be on developing a new set of schemata just for these stimuli. One handicap in learning to process dot stimuli is their lack of meaning with respect to higher-level structures. Even if a schema is developed for a set of dots, this schema is not integrated into any higher-level structure. We do not know how important it is to embed schemata into other schemata, but such embedding could be an important factor in developing truly natural concepts. This point about the web-like nature of concepts relates to Nelson's notion about integrating concepts into scripts.

What about the role of examples in concept formation? Because the individual stimuli in experimental situations are poorly encoded and are not expressed in any kind of a meaningful episode, they are generally poorly remembered. Brooks' argument that in real life we overlearn a few examples of concepts is important here. In these experiments, even if a few examples are overlearned, they are still not embedded in any kind of meaningful context. Reber's strings are not at all like going to a birthday party. Nevertheless, as Brooks has demonstrated, if exemplars are learned well enough, they can be used to classify new stimuli that match them.

Reed, however, argues that some kind of abstraction occurs *before* individual items are learned. If so, this suggests that higher-level schemata can be used to organize whatever lower-level schemata are evoked by the stimuli prior to the development of new or modified lower-level schemata designed especially for the new stimuli. If the stimuli used were real faces, individual faces might be recognized more easily, and identification learning might occur before concept formation. When individual items are stored in memory, the concept formation process can abstract information from them, rather than only from items as they are presented. Further, if a set of well-coded stimuli exist in memory or can be generated by a simple set of productions (such as the simple *size-color-shape* stimuli used in many concept identification experiments), then hypothesis testing

may be an important form of concept formation. Here, the subject makes an explicit and conscious search of possible rules, and when a rule is successful, it is stored in declarative memory and integrated into the schema for processing such stimuli. Although we usually process speech sounds and graphemic patterns automatically (that is, implicitly), we can sometimes explicitly analyze them in order to form rules about their structures. These rules may produce behavior that violates the normal schema, as when we use the spelling of a word to modify our pronunciation—the hyperurbanism response (e.g., when the letter *t* is pronounced in the word *often*).

Interestingly enough, the frequency models seem, all in all, to be the most successful. Neumann's experiment shows that prototypes are not necessarily formed, and Reitman and Bower's analysis of Bransford and Franks' experiment suggests an alternative to integration. Rosch's analysis of family resemblances provides a reasonable explanation for both goodness-of-example and formation of category membership. According to the computational model, a frequency analysis is not enough (although no one has explicitly said it was), and an explanation of how frequency works is as important as the fact that frequency is an effective variable in concept formation. In a way, frequency has to be important if the mind is to be attentive to the most important events in the environment. But even more central to developing concepts is the way experiences are organized. That is, the system proposed here is heterarchical, and it is not the simple frequency of some feature that is crucial, but rather the role of the frequency of some feature in some context. Having a theory of how context is compartmentalized is as important as having a rule for the effect of frequency.

The computational theory of concept formation begins by assuming that episodes are stored in memory, having been encoded by whatever schemata exist for processing the episodes. Data presented to the system cause schemata to become active (data-driven). Each activated schema has built into it an anticipatory function (goal-oriented) that elicits other schemata. It seems reasonable to assume that data activate schemata by matching features and, in particular, by weighting these features according to their frequencies or joint frequencies. Hence, the frequency of features is especially important for perceptual schemata. Higher-level schemata activate and order the sequencing of other schemata; frequency may play a role here also. For example, where alternative schemata exist, the most frequent one might be activated first or with more strength. However, the sequence of schemata is not influenced by the frequencies of features.

Newmann's experiment clearly shows the trade-off between using existing schemata (realistic faces) and new ones (geometric designs). Frequency was more important in the latter case. But frequency will affect even existing schemata. Therefore, schemata should not be thought of as fixed entities but rather as dynamic structures constantly undergoing change. We are continually debugging our schemata and modifying their sequence of actions to make them function more efficiently.

By using a highly discriminable rule, Reitman and Bower replicated the Bransford and Franks results; that is, examples that fit the "full" concept more exactly (i.e., more complex sentences) received higher confidence ratings. Why should this be? Reitman and Bower's frequency model requires that all n-tuples be looked at, and that the past frequencies of experience of each n-tuple be summed to provide an overall indication of familiarity. Why all n-tuples? Why n-tuples at all? I would agree with Reitman and Bower that n-tuples are a natural way to encode such stimuli; but contrary to the way their model works, I would argue that the size of the n-tuple processed depends on a higher-level schema used to organize such complex stimuli. The higher confidence rating is due not simply to frequency but also to the presence of the higher-level schema. Here, we must assume that the lack of success of the schema triggers a "debugging" routine that examines the reasons for the lack of success. If the reason is that the initial match on letters fails (noncases), the schema is simply inappropriate, and the subject gives a low confidence rating. If it only partially succeeds because it does not meet all the conditions of the full concept, then it is rated according to how well it meets the conditions. Bransford and Franks used material that depended on existing schemata and so, got this result. With a low discriminability rule, Reitman and Bower found that confidence ratings decreased with an increased number of elements. Because the subjects did not realize there was a rule, they did not indicate how well the example matched the rule. The argument here is that the ratings reflect the subject's awareness of how well the example matches the rule, not how confident he or she is about whether the test item is old or new. The notion of integration does not depend on an old–new difference since the increased confidence ratings indicate the existence of a schema suitable for integrating the stimuli.

Rosch's ideas about family resemblances can easily be incorporated into our computational theory. A central point of her theory is that not all exemplars of a category are equally "good"; that is, people judge some exemplars as more representative than others. Rosch introduces an algorithm for measuring the goodness-of-example of an exemplar of a category. This algorithm requires counting the frequencies of each feature of all exemplars of a category and assigning the count for each feature as a weight for that feature. The goodness-of-example of an exemplar is, then, the sum of the weights of its features. Now assume that each exemplar has associated with it a schema designed to match it on the basis of features. If this matching process is to be efficient, it should take the frequencies of features into account.

It is assumed that as each object is recognized, its schema is activated. Further, activated schemata will tend to activate the schema or schemata that represent superset categories. As there are many possible supersets, the particular superset chosen will depend mainly on the kind of goal that is guiding the processing. However, the set of features common to the exemplar and the superset schemata will determine the ease and likelihood of selecting the superset. The set of defining features for a superset schema should maximize the weights of

features unique to a particular category and minimize the weights of features that occur in many different categories. Here, again, the Rosch algorithm provides a reasonable rule for selecting a superset category.

The relationship of this theory to studies in semantic retrieval is fairly obvious. Reaction-time studies show that the better an item is as an exemplar, the faster it is judged a member of its category. Smith, Shoben, and Rips (1974) present one version of how such a comparison is made. The computational theory presented here looks at the mechanism behind confidence ratings, goodness-of-example judgments, and reaction times in a slightly different way. We assume that the information about relative frequencies is stored in the schema and that there are procedures that can observe the frequency information stored as part of the schema. The information abstracted from two different schemata is then compared. Presumably, feature frequencies are stored as declarative information and so are potentially observable. The procedure of evoking a superset schema during normal thinking and understanding is not declarative and therefore not observable. We are not aware of inferring the category to which an item belongs, but under instructions, we can make a comparison of some of the information stored in the relevant schemata. Thus, there is a difference between asking the cognitive system if a dog is an animal and having the system use this relationship while processing information. The goodness-of-example ratings used by Rosch come about by this process. The confidence ratings in the integration experiments by Bransford and Franks (1971) and by Reitman and Bower (1973) are also due to computations on the declarative information stored in schemata.

The rule or hypothesis-testing model can eas ′ be made a part of the computational theory. A rule, regardless of how it was acquired, is a declarative statement about the world. Rules are formed by hypothesis testing—that is, by self-instruction—and are based on a discovery of regularity in the environment, including observation of one's own behavior. Rules are also taught directly. Once a rule is known, it can be applied. This is done by a sequence of schemata that match the condition of the rule to the environment and then determine what the consequences of the rule are. The sequence of schemata utilized to apply a rule is complicated and must be distinguished from the sequence that do the computing in "automatic" rule-governed behavior, such as language understanding. The latter sequence defines the rule implicitly; the former applies a rule explicitly. The distinction is not unlike the distinction between running compiled code and interpreting symbolically stored code in a computer.

In Summary

The purpose of this chapter was to present a summary of the major ideas about concept formation. For that purpose, four major models were discussed, using a few of the most important papers representing each model as a focus for the discussion. The major problem with the current state of theorizing about concept

formation is a lack of any integration of a number of reasonable and supported ideas. Researchers taking different points of view do not mention the work of others, even though both are studying concept formation. Because they cannot all be correct as long as they state their positions as being complete, some integration of these various research interests is definitely needed.

The second purpose of the chapter was to present a computational theory of concept formation that, we believe, can serve to integrate the facts from the different research areas. This theory is not a micromodel introduced to account for a limited set of experimental data. Rather, it is a theory broadly based in current research on human information processing, semantic memory, and artificial intelligence. The theory is based on work in computer simulation of human understanding systems, such as Anderson's ACT and Norman and Rumelhart's MEMOD theories. Some general comments were made to illustrate how the computational model would account for some of the results discussed in the first part of the chapter. Obviously, a great deal more work is needed to make this theory quantitatively rigorous and then to test it exactly in the concept formation area.

ACKNOWLEDGMENT

REFERENCES

Anderson, J. R. *Language, memory, and thought*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Attneave, F. Transfer of experience with a class schema to identification learning of patterns and shapes. *Journal of Experimental Psychology*, 1957, *54*, 81-88.

Bartlett, F. C. *Remembering*. Cambridge: Cambridge University Press, 1967. (Originally published in 1932.)

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Bransford, J. D., & Franks, J. J. Abstraction of linguistic ideas. *Cognitive Psychology*, 1971, *2*, 331-350.

Brooks, L. Non-analytical concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), *The nature and principles of categories*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Bruner, J. On perceptual readiness. *Psychological Review*, 1957, *64*, 123-152.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking*. New York: Wiley, 1956.

Charniak, E. *Toward a model of children's story comprehension*. Doctoral thesis (AI-TR-266), Artificial Intelligence Laboratory, MIT, Cambridge, Massachusetts, 1974.

Chomsky, N., & Miller, G. A. Finite state languages. *Information and Control*, 1958, *1*, 91-112.

Evans, S. H. A brief statement of schema theory. *Psychonomic Science*, 1967, *8*, 87-88.

Franks, J. J., & Bransford, J. D. Abstraction of visual patterns. *Journal of Experimental Psychology*, 1971, *90*, 65-74.

Gibson, J. J., & Gibson, E. Perceptual learning: Differentiation or enrichment? *Psychological Review*, 1955, *62*, 32-41.

Griggs, R. A., & Keen, D. M. The role of test procedure in linguistic integration studies. *Memory & Cognition*, 1977, *5*, 685-689.

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

Hyman, R., & Frost, N. H. Gradients and schema in pattern recognition. In P. A. Rabbit & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press, 1975.

Miller, G. A. *The psychology of communication: Seven essays* (Chapter 7, Project Grammarama). Baltimore, Md.: Penguin Books, 1969. (First published by Basic Books, 1967.)

Miller, G. A., & Johnson-Laird, P. N. *Language and perception*. Cambridge, Mass: Belknap Press of Harvard University Press, 1976.

Millward, R. B. Theoretical and experimental approaches to human learning. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth & Schlosberg's Experimental psychology* (3rd ed.). New York: Holt, Rinehart & Winston, 1971.

Millward, R. B., & Wickens, T. D. Concept-identification models. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Learning, memory, and thinking* (Vol. 1). San Francisco: Freeman, 1974.

Minsky, A. A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Nelson, K. Concept, word, and sentences: Interrelations in acquisition and development. *Psychological Review*, 1974, *81*, 267-285.

Nelson, K. Cognitive development and the acquisition of concepts. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Neumann, P. G. Visual prototype formation with discontinuous representation of dimensions of variability. *Memory & Cognition*, 1977, *5*, 187-197.

Newell, A. A theoretical exploration of mechanisms for coding the stimulus. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston, 1972.

Norman, D. A., Rumelhart, D. E., & the LNR Research Group. *Explorations in cognition*. San Francisco: Freeman, 1975.

Posner, M. I., & Keele, S. W. On the genesis of abstract ideas. *Journal of Experimental Psychology*, 1968, *77*, 353-363.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, Mass.: MIT Press, 1968.

Reber, A. S. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 855-863.

Reber, A. S. Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 88-94.

Reber, A. S., & Millward, R. B. Event observation in probability learning. *Journal of Experimental Psychology*, 1968, *77*, 317-327.

Reed, S. K. Pattern recognition and categorization. *Cognitive Psychology*, 1972, *3*, 382-407.

Reed, S. K. Category vs. item learning: Implications for categorization models. *Memory & Cognition*, 1978, *6*, 612-621.

Reitman, J. S., & Bower, G. H. Storage and later recognition of exemplars in concepts. *Cognitive Psychology*, 1973, *4*, 194-206.

Rosch, E., & Mervis, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, *7*, 573-605.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Broem, P. Basic objects in natural categories. *Cognitive Psychology*, 1976, *8*, 382-440.

Rumelhart, D. E., & Ortony, A. The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Schank, R. C. *SAM — a story understander* (Research Rep. No. 43). New Haven, Conn.: Yale A. I. Project, 1975. (a)

Schank, R. C. The structure of episodes in memory. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding* New York: Academic Press, 1975. (b)

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214-241.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization and memory*. New York: Academic Press, 1972.

Winograd, T. Frame representations and the declarative-procedural controversy. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding*. New York: Academic Press, 1975.

Wittgenstein, L. [*Philosophical investigations*] (G. E. M. Anscombe, trans.). New York: Macmillan, 1953.

24

Learning Theory, Instructional Theory, and Adaptation

Thomas J. Shuell
State University of New York at Buffalo

In recent years we have heard a great deal about the two disciplines of scientific psychology (Cronbach, 1957, 1975). The effort to combine the correlational and experimental approaches to the study of human behavior reflects a new and growing interest in the role of individual differences in psychological processes and instructional methods. A new approach for investigating issues in this general area has been created, and several different types of research have been influenced by the resultant merger of thinking about individual differences, learning, cognition, and instructional treatments. The continued success of this combined approach, however, now requires an additional merger. It is now time to call for the unification of the two disciplines of educational psychology.

Within educational psychology, research on learning and teaching is usually approached from two very different perspectives, each with its own paradigm and its own methodology for investigating the problem. On one hand are those investigators concerned with the psychology of learning and cognitive processes, frequently on a rather microscopic level in a more-or-less laboratory tradition. On the other hand are those researchers concerned with teaching as it occurs in the classroom, and their methodology usually is based on observation and description. The assumptions and concerns of the two approaches are usually very different. Although there have been some honest attempts to bridge the gap between these two different areas of concern, most of these attempts have so far overlooked most of the crucial issues that must be addressed if these two disciplines are to be integrated in the most productive manner.

The successful integration of these latter two disciplines, however, cannot ignore the current concern for individual differences that has grown out of the merger of the correlational and experimental traditions of psychology. But the

277

influence must not be one-way. The concern for individual differences within cognitive psychology and the attempt to relate those differences to instructional issues—that is, the aptitude-treatment interaction (ATI) paradigm—will benefit greatly from the consideration of some of the issues raised within the research on teaching and psychology of learning traditions of educational psychology.

The purpose of this chapter is to consider the relationships among some of the issues raised in these different areas and to suggest a direction in which we might move in order to develop a workable integration of these various concerns. A theme that is evident throughout the chapter is a concern for how these issues can be conceptualized in a fashion that will permit us to ask our research questions in the most productive manner. The first part of the chapter is concerned with the relationship between learning (including cognitive processes and information processing) on one hand and instruction on the other. Next the role of individual differences in learning and instruction is discussed. Finally, these two concerns are combined into the more general concern for adapting instruction to meet the need reflected by individual differences among students.

THE RELATIONSHIP BETWEEN
LEARNING AND INSTRUCTION

Almost everyone would agree that the purpose of teaching is to influence learning in one way or another. As a consequence of teaching, the student will either (1) learn something that he or she would not have learned without the instruction, or (2) learn it in a more efficient manner. Although both cognitive and affective outcomes are usually acknowledged as being important, the emphasis is usually on the cognitive. Beyond this point, however, there is little agreement on how the relationship between learning and instruction should be conceptualized. Likewise, there is little agreement on how these combined concerns should be utilized in developing effective instructional materials or in designing effective learning environments.

One characteristic common to all definitions of learning is that learning involves, in one way or another, a change in behavior (Shuell & Lee, 1976). The change that is involved may well be a change in a schema or some other type of system for representing knowledge. The distinction between learning and performance has a long tradition in psychology. Thus, although the change may not always be reflected in performance, most would agree that learning involves a change in a person's knowledge or ability to perform some task and that this change can only be determined by observing some sort of change in the individual's behavior.

Learning psychology has traditionally investigated the relationship among variables thought to be responsible for those changes in behavior. For present purposes, the specific variables that have been investigated are not as important

as the research concern for factors influencing changes in what an individual is capable of doing. "Implications" of learning theory and cognitive theory for education are sometimes discussed, but the specifics of how these implications are to be translated into instructional procedures for use in a particular situation are usually stated in very vague terms. The translation, for all practical purposes, is left entirely in the hands of the teacher or instructional designer.

Research from the learning tradition, however, has a number of limitations when it comes to applying that research to educational problems associated with teaching. Most of the learning research has ignored variables typically occurring in a normal classroom environment—for example, variables characteristic of the dynamics of teacher-student and student-student interactions. Although laboratory studies of learning and instruction can provide rich sources of potential variables that may be useful in further research on teaching (Rosenshine & Furst, 1973), traditional approaches to the psychology of learning are insufficient in several critical ways for purposes of developing a theory of instruction (Gage, 1963; Gagné, 1962; McKeachie, 1974). The qualitative differences between a psychology of learning and a psychology of instruction are discussed more fully in a later section of this chapter.

Research on teaching, on the other hand (for the time being, research on instructional design rising out of the learning tradition is not being included in the rubric "research on teaching," although this approach is considered shortly), has been concerned with observing and describing interpersonal interactions that occur in a typical classroom or with correlations among various teacher characteristics and various criteria of effectiveness (Dunkin & Biddle, 1974). Only a small minority of these studies, unfortunately, have investigated the relationship between classroom interactions (including teacher behaviors) and student achievement. Although reviews of these studies (Dunkin & Biddle, 1974; Rosenshine, 1971a, 1971b; Rosenshine & Furst, 1971, 1973) have isolated several teacher variables that appear to be related to student achievement, the manner in which these variables are related to the learning processes of students is not well understood.

In trying to develop instructionally relevant research on individual difference and ATIs, several things need to be kept in mind. Regardless of the extent to which computer-assisted instruction and instructional systems where the student works independently develop, a large amount of our instructional effort will continue to be spent in group instruction under the supervision of a live teacher. The reasons for this include feasibility—at least in the foreseeable future—cost, and the simple fact that certain objectives best lend themselves to, and may even require, group instruction. Thus, it is important for us to pay attention to variables that reflect the dynamics that exist in group instruction with a live teacher. This concern, however, should not be viewed as being limited only to group instruction. Many of the instructional variables that are reflected in the dynamics of live, group instruction also operate in other types of instructional settings as

well, such as the writing of text materials, developing films and videotapes, computer-assisted instruction, and so forth. All instruction, including instruction carried on by the use of previously prepared instructional materials in the absence of a live teacher, involves the interaction of at least two people. In one case all parties to the instructional act are physically present, whereas in the other case the instructor's influence is more remote. The form and nature of the interaction between teacher and student may differ widely, but the reality of the relationship is still there.

Learning Theory and Cognitive Psychology

The shift in emphasis during the last decade from research on learning, primarily from a behavioristic S-R perspective, to a concern for cognitive processes and information processing has had a number of important consequences. This new emphasis on cognitive psychology, however, has been concerned primarily with describing the various stages involved in the information-processing sequence and in determining the characteristics of these stages rather than with learning per se (i.e., concern for variables responsible for changes in behavior whether internal or external).

Although cognitive researchers and theorists have sometimes talked about learning, most have implicitly viewed learning as being synonymous with the storage and retrieval of new information or strategies. Several researchers (Bransford & Franks, 1976; Greeno, 1974) have been somewhat more explicit by stating that learning is, for all practical purposes, the same thing as comprehension; and Norman, Gentner, and Stevens (1976) have defined learning in terms of schemata modification. So far, however, little work has been done on developing systematic, lawful, and empirically based hypotheses about variables that influence changes in comprehension or the modification of schemata.

The most systematic attempt to deal with issues of learning within the framework of modern-day cognitive psychology is the ACT theory of learning developed by John Anderson (Anderson, Kline, & Beasley, 1978, and Chapter 21, this volume). ACT is a computer simulation program that predicts learning data in a variety of different situations involving cognitive processes. Another, although rather different, attempt to investigate problems of learning within the cognitive framework is several recent studies by some of Piaget's colleagues at Geneva (Inhelder, Sinclair, & Bovet, 1974).

Cognitive psychology has been trying to break away from the strong influence of S-R learning psychology, so it is probably natural to expect that the focus would have been on issues other than learning. But cognitive psychology has come of age, and the situation is changing. There have been several recent attempts (Anderson, Spiro, & Montague, 1977; Klahr, 1976) to relate current cognitive theory and research to instructional issues. Although these two sources

plus the present one are extremely important steps toward the necessary integration of research and theory in cognitive psychology and educational practices, considerable work is still required before the information contained in these sources will be either directly useful or capable of being translated into statements that are useful to a classroom teacher or an instructional designer with less than a high level of sophistication and expertise.

Learning Theory and Instructional Theory

Part of the difficulty of trying to specify the applications of learning theory or cognitive theory to instructional situations is that the very nature of these theories and supporting data precludes their direct application to practical utilization. Knowledge about learning and cognition is essential to the development of effective instructional procedures and materials, but this knowledge is qualitatively different from the type of knowledge required for instructional design (Bruner, 1966; Gage, 1963, 1964; Gagné, 1962). A theory of learning is concerned with the relationship among variables responsible for a change in a person's behavior; in other words, it is concerned with how people learn. A theory of instruction, on the other hand, is concerned with how one person influences the learning of another person. In other words, a theory of instruction is concerned with how the variables specified in a theory of learning can be controlled in a way that will facilitate the student's learning of the desired outcome. For example, a theory of learning might specify that there is a direct relationship between how much time a person spends studying the material being learned and the amount of material actually learned. Such a statement is perfectly appropriate for describing the factors that influence human learning, and this particular relationship has been verified many, many times. Yet there is evidence (Gagné, 1962) that in some instructional settings, simple practice on the task to be learned does not necessarily result in better performance. In order to improve performance on the overall task, the learner may have to identify and become proficient in performing certain subcomponents of the task that are prerequisite to performance on the overall task. A theory of instruction would be concerned with specifying— probably by means of a task analysis—what those important subcomponents are and the sequence in which the student should practice on the various components of the task. A theory of instruction would also be concerned with various ways in which the student might practice on the task—for example, reading text materials, listening to a lecture, performing the task in the laboratory, doing homework, and so forth.

A body of knowledge specifically concerned with the applications or translation of the basic knowledge of learning and cognition to instructional situations is needed but not presently available. The attempts to develop a science of design (e.g., Glaser, 1976a, 1976b) are noteworthy attempts to fill this gap. But the

present efforts still require a great deal of sophistication and knowledge of basic research findings for their implementation, and for the most part, they are pragmatic, atheoretical attempts. As Glaser (1976b) has pointed out, what is needed is a body of knowledge that transcends the skill and talents of an individual investigator. Although the present attempts are theory and data based to some extent, their efforts could be substantially improved if a viable theory of instruction were available.

One stumbling block to research on teaching has been a tendency for researchers to ask inappropriate questions as far as the relationship between teaching and learning is concerned. Research on any topic can be either limited or facilitated by the way in which one conceptualizes the problem being investigated. The development of a useful theory of instruction requires the relationship between learning and instruction to be conceptualized in a manner that permits us to ask questions that get at the crux of the relationship in a specific and direct fashion.

In asking research questions, there is often a tendency to focus on variables that are highly visible and have a fair amount of face validity but that may not be directly related to those things in which we are actually most interested—in this case, those factors and processes that influence student achievement. One example of this conceptualization problem is the large amount of research that has been done over the years on the relationship between class size and student learning.

Class size is an obvious variable that many teachers, administrators, and researchers alike tend to feel is somehow related to teacher effectiveness. Although class size may be a very legitimate variable for certain types of educational research—for example, research on educational administration or organization and classroom management—there is no way that class size can be directly related to student learning unless we want to hypothesize that the amount of human flesh in the immediate environment of the student influences his or her learning. When no concern is given to how the cognitive processes of students responsible for learning are or can be engaged in classes of different sizes, it is little wonder that so much of the research on class size has been inconclusive (Jamison, Suppes, & Wells, 1974).

If the concern is with student learning, then our research questions should be phrased in terms of the way in which activities that *may* occur in classes of different sizes can engage those psychological processes that will result in the desired learning. For example, feedback is known to be an important variable, at least for certain types of learning. Thus, if an investigator were interested in studying the relationship between class size and student learning, he or she might look for ways in which feedback or other learning variables might operate in classes of different size.

Although feedback—especially to an individual student's responses—might be more likely to occur in a small class than in a large class, a small class does

not guarantee that it will occur in a way appropriate for effective learning. A class discussion involving teacher–student and student–student interactions is well suited for feedback to occur in a manner directly relevant to the learning process, although appropriate feedback may not occur if the teacher is inept at providing feedback or permits the discussion to get off on a tangent that is unrelated to the objectives the students are supposed to be learning. Likewise, appropriate feedback is not likely to occur if the teacher lectures to the class (an activity that can be performed in a large class just as easily as a small class) rather than letting the students become involved in a true discussion.

Class size can be a legitimate concern. The purpose of the present example is to illustrate my concern for the importance of conceptualizing educational variables in a manner directly relevant to the teaching/learning process if that is the concern of the investigator. The variables considered, however, should not be limited to traditional learning and cognitive process variables, such as feedback. Social psychological and group-dynamic variables need to be considered as well, especially as they relate to human learning and the instructional process.

Instructional research needs to be guided by a conceptualization of the relationship between learning and teaching that captures the dynamics of both concepts. We have already seen that learning is concerned with the psychological processes responsible for a change in the way a person represents some knowledge or is able to perform some task. *Instruction,* as the term is used here, is concerned with how those psychological processes can be influenced by another person. In a general sense, it refers to any situation in which one *individual* intentionally tries to influence the learning of another individual by structuring the environment of the learner in such a way that the latter will achieve the desired outcome (Shuell & Lee, 1976). At this level, no distinction is made between instruction that is carried on in the presence of a live teacher and instruction that is carried on indirectly through predesigned instructional materials such as textbooks, films, specific curriculum materials, and so forth. Likewise, no distinction is made between instruction that is carried on within the framework of formal education and instruction that is carried on in other types of situations such as counseling, advertising, parent–child interactions, journalism, and so forth.

The instructional process is viewed as being concerned with the question of how the teacher (author, therapist, parent, or the like) can engage or elicit in an appropriate fashion the psychological processes and strategies of the students that are necessary for them to learn the desired outcome. These psychological processes that need to be engaged, however, are not limited to those processes involved in the handling of cognitive information. Motivational and attitudinal processes are also involved. At times it may also be desirable to elicit emotional reactions.

In establishing a learning environment for a student that will help him or her achieve the desired outcome, a teacher or instructional designer must determine

the combination of teaching behaviors to be used in a particular instructional situation. This choice of teaching behaviors should be based on many factors relevant to the teaching/learning process, most notably the type of learning that must be engaged in to achieve the desired outcome and what is known about the way people learn and process information. To consider all of these factors may be an impossible task for a classroom teacher and for many instructional designers, but a useful theory of how the factors are interrelated would help to make the task manageable and increase the likelihood that effective teaching will occur.

Toward Developing a Theory of Instruction

A useful theory of the instructional process should be able to specify teaching behaviors that will maximize desired instructional outcomes (both cognitive and affective) while minimizing undesirable outcomes. Such specification must take into account those factors that define the context and constraints of instruction, but these factors must be conceptualized and defined in terms that have a direct relationship to those information-processing and psychological processes responsible for learning. The heart of such a theory is an explicit description of the relationship between specific teaching behaviors (including both those behaviors exhibited by real teachers and those reflected in prepared instructional materials such as textbooks, films, and the like) and those cognitive strategies and psychological processes responsible for both cognitive and affective learning in the student. In developing such a theory, knowledge and ideas from both learning/cognitive psychology laboratories and from what is known about research on teaching should be incorporated together to form this body of knowledge about the instructional process.

There are at least three different ways that this theoretical endeavor can be approached. For the time being, let me distinguish among them by referring to them as a theory of instruction, a theory of teaching, and a theory of design. These three concerns are clearly related. They do, however, seem to me to represent three different concerns. Making a distinction among them, hopefully, will help to clarify our task and reduce its overall complexity by identifying different aspects of the instructional process.

A *theory of instruction* is perhaps the simplest of the three and may form something of a necessary but not sufficient base for the other two types of concerns. A *theory of instruction*, as it is being used here, refers to the specification of the relationship among (1) various instructional variables (as distinguished from learning variables and cognitive variables); (2) learning/cognitive process variables; and (3) the nature of the material or outcome that the student is going to acquire. The distinction between instructional variables and learning variables parallels the distinction between instructional theory and learning theory discussed in the preceding section. More specific examples of both types

of variables are given shortly. A theory of instruction is not concerned with how effectively or efficiently a teacher or instructional designer may be able to make use of the relevant instructional variables in designing a particular learning environment. It merely states the relationship that exists between instructional variables and learning variables.

The flavor, at least, of the type of conceptualization I have in mind is represented by some of the work of Hilda Taba (1967). In developing an elementary social studies curriculum, various tables were developed that related the types of questions a teacher should ask in order to elicit appropriate cognitive processes for performing specified overt activities. One such table for concept formation is presented in Table 24.1. Although these tables were to some extent theoretically based, there was no systematic attempt to relate the various factors to a psychological theory of learning or cognitive processing. Likewise, no attempt was made to validate empirically the relationships depicted in the tables. Data were collected on the extent to which the suggested teacher behaviors produced corresponding behavior in students, but unfortunately, no attempt was made to relate them to student achievement or learning. Nevertheless, the approach serves as a useful example of an attempt to specify a general correspondence between instructional variables and psychological processes for a particular type of content (concepts) with recommendations for appropriate teaching behaviors.

A *theory of teaching* brings in a new level of complexity by recognizing that teachers and instructional designers may not, for a variety of reasons, be able to capitalize on the relationships specified by a theory of instruction in an optimal fashion; that is, there is a psychology of the teacher or instructional designer that must be considered, and a theory of instruction does not take these considerations

TABLE 24.1
Relationships Among Teaching Strategies, Covert Mental Operations
of the Learner, and Appropriate Eliciting Questions for Teacher to Use
in Concept Formation Lesson[a]

Overt Activity	Covert Mental Operations	Eliciting Questions
1. Enumeration and listing.	Differentiation.	What did you see? Hear? Note?
2. Grouping.	Identifying common properties, abstracting.	What belongs together? On what criterion?
3. Labeling, categorizing.	Determining the hierarchical order of items. Super- and subordination.	How would you call these groups? What belongs under what?

[a]*Source:* Hilda Taba, *Teacher's Handbook for Elementary Social Studies,* © 1967, Addison-Wesley Publishing Company, Inc., Chart #7, page 92. Reprinted with permission.

into account. A teacher or instructional designer may be unable to follow exactly, perhaps because of habit or past experience, those things recommended by a relevant theory of instruction. For example, students affect the teacher's behavior as well as the other way around (Fiedler, 1975; Klein, 1971; Noble & Nolan, 1976). A theory of teaching would consider the interdynamics that are involved in such a mutually interactive exchange between two psychological beings. Concerns for such things as management skills, monitoring, and the like would also be reflected in a theory of teaching. Naturalistic studies of classroom teaching reflect this approach, although most of the studies in this vein have not been concerned with learning processes.

Finally, a *theory or science of design* would be concerned with the mechanics of developing an instructional program or unit. Concerns for such topics as task analysis, determining a student's present state of knowledge, and so forth (including procedures for carrying out these activities) would be an integral and important part of a theory of design but are topics that would not be specified in either a theory of instruction or a theory of teaching. A theory of design would also be concerned with utility or cost-benefit analysis, and it would provide information on how the teacher or instructional designer can develop appropriate matches among desired outcomes, specific learning environments, and the relevant profile of individual differences of the learners. A theory of design needs a theory of instruction from which to work, but it would be concerned with different types of issues.

This chapter is concerned primarily with a theory of instruction. As already noted, a theory of instruction must accurately reflect the psychological processes that are to be engaged or elicited by the instructional variables being considered. But what are the basic psychological processes involved in human learning and information processing that such a theory should incorporate? Most models of human information processing (e.g., Atkinson, Herrmann, & Wescourt, 1974; Bower, 1975) have emphasized the sequence of stages through which information passes, rather than the processes that are encountered along the way. Newell and Simon (1972) have developed a list of the elementary information processes sufficient to produce the full range of information processing encountered. Their list, however, is based primarily on principles of computer science and seems more appropriate for computer simulation than for capturing the psychological reality and richness involved in cognitive psychology, human learning, and instruction.

A tentative list of psychology processes relevant for instruction is presented in Table 24.2. This list is undoubtedly not exhaustive and may be redundant. It merely represents a beginning attempt to specify the basic processes responsible for learning and the cognitive processing of information that need to be reflected in a theory of instruction. The appropriate level of analysis is not completely clear at present. Some of the processes presently on the list may need to be combined into more relevant clusters in either a linear or hierarchical manner.

The level of analysis that will ultimately prove most appropriate for a theory of instruction will need to be microscopic enough to reflect accurately and capture the most basic psychological processes involved in learning while being molar enough to have theoretical and practical utility for instruction.

There are several ways one might approach this problem. The work of Rose (Chap. 3, Vol. 1) represents one approach. Another approach that represents the concern for an appropriate level of analysis, for example, might be an attempt to see how various cognitive-style variables could be made to map onto the more basic processes. In fact, such an attempt might help to define new and more appropriate cognitive-style variables. In any event, the role of individual differences in these various processes will have to be considered. Perhaps the most important thing at this point is to be aware of the need to analyze both the learning process and the instructional process in a way that permits one to establish a direct relationship between the two domains.

A theory of instruction must specify how various instructional variables can be utilized to control each of the cognitive processes depicted. For example, we know that individuals attend to the environment and things in it in a selective manner. A theory of instruction would specify, not *what* variables influence attention, but *how* selective attention can be controlled (perhaps by isolation of relevant dimensions in the learning material by the use of color, highlighting them either verbally by means of emphasis or mechanically by means of a pointer, and so forth). Consideration must be given to the possibility that gaining auditory attention may be different from gaining visual attention, which in turn

TABLE 24.2
Psychological Processes Involved in Human
Learning and Information Processing

Reception
Attention
Feature extraction
Encoding
Search
Comparison of information
Holding
Scanning
Goal setting
Motivation
Grouping
Hypothesis generating
Decision making
Transforming information
Recoding or translation
Response generating
Synthesis of information

may be different from gaining tactical or enactive attention. Also, depending on the objective or criteria involved, certain types of intervention may actually be detrimental to learning (e.g., Samuels, 1967), and these negative influences would also be specified in the instructional theory. It should also specify the important dimensions in a particular type of learning that require attention by the learner. Again, individual differences of the learners must be taken into account. There are several places we can begin to search for relevant instructional variables.

One place to begin is by listing instructional variables that correspond to the psychological processes suggested in Table 24.2. A tentative list of such instructional variables is presented in Table 24.3. The variables listed represent little more than a verbal qualification of the psychological processes suggested earlier, but listing them in this manner may help us to identify the types of variables that must be reflected in a theory of instruction. Other variables may well be involved, and the level-of-analysis problem discussed previously with respect to learning variables is equally important here. In the present situation, however, some of the constraints placed on the analysis are defined a little more clearly. At some level the analysis must be stated in terms of those teaching behaviors or instructional modes that teachers and instructional designers find manageable from both a conceptual and a practical point of view. A lecture may be a convenient vehicle for teachers to think about the instructional process, but it may have extremely limited usefulness for understanding how instructional interventions are related to those cognitive processes necessary for learning to occur. A useful theory of instruction would specify the relationship among these various concerns. Thus, the more traditional modes of instruction, such as lecture and discussion, and/or the more specific teaching behaviors, such as asking questions, disseminating information, and explaining, should be incorporated into the analysis. These factors could then be related to more specific variables such as those outlined in Table 24.3. The best way to represent the relationship among these various factors may be hierarchical or orthogonal in nature, but it must be possible to relate them in one way or another to the types of psychological processes represented in Table 24.2.

On a very global level, the identification of various families or models of teaching, such as the analysis by Joyce and Weil (1972), may provide some helpful insights, but for the most part, such general taxonomies provide little help in identifying the specific instructional variables involved in such general models. This problem is really the same as the problem involved in specifying relevant treatment variables in the aptitude–treatment interaction literature. Little work has been done in this area so far, although Fleishman (1972, 1975) has made a beginning with respect to psychomotor learning. Other investigators (Cronbach & Snow, 1977; Frederiksen, 1972; Mischel, 1973) have discussed the importance of developing a taxonomy of situations or treatments, including

TABLE 24.3
Some Potentially Relevant Instructional Variables

Goal setting
Motivation inducing
Information presenting
Attention directing
Encoding inducing
Storage inducing
Retrieval inducing (recall cuing)
Hypothesis eliciting
Transformation generating
Rehearsal producing
Feedback providing
Organization inducing
Response eliciting

treatments involved in aptitude-treatment interactions, but the form that such a taxonomy should take is just beginning to come into focus.

THE ROLE OF INDIVIDUAL DIFFERENCES

Differences among individual learners are virtually limitless, and it is possible to define or describe these differences in a variety of ways. In considering the role of individual differences both in cognitive learning theory and in adapting instruction to the needs of individual students, some consideration must be given to the types of individual differences that are most appropriate for these concerns. Criteria must be developed that will permit us to determine which individual differences are important and which ones are trivial. Until recently, thinking about individual differences has been heavily influenced by the traditional psychometric approach to the problem and by a concern for those types of individual differences that are highly obvious, such as sex, race, and socioeconomic status, but that may be only tangentially relevant and of limited usefulness in helping us understand the role of individual differences in cognitive learning and instruction.

Three major sources of individual differences that seem to be particularly relevant to our present concerns are presented in Table 24.4. All three sources are important for an adequate understanding of individual differences, especially as they relate to instruction. It is not uncommon, however, for investigators to ignore one or another of these sources. Although it may be necessary for a given research project to focus on only a single source, we should be careful not to become so preoccupied with one source that we begin to feel that it is capable of explaining exclusively the role of individual differences important to our general

TABLE 24.4
Three Main Sources of Individual Differences

Knowledge

Different types of knowledge are involved.

Includes achievement-by-treatment interactions.

Use of task analyses are typically involved.

Learned Strategies

Strategies for processing information that have been learned but that are relatively stable.

Concern for trainable aptitudes.

Includes many of the cognitive-style variables.

Basic Processes

Processes involved in learning and cognition that cannot be changed by training.

Includes physiological mechanisms related to learning and cognition.

Examples would likely include such factors as channel capacity and reaction time.

concern. All three sources are important, and it seems likely that in many situations, two or even all three of them should be investigated simultaneously.

The first source is concerned with the learner's present knowledge that is relevant to what he or she is currently trying to learn. In that sense the concern is for state variables rather than process variables. It must be recognized, however, that several basically different types of knowledge are involved. Exactly what these different types of knowledge are is not completely clear at present, but the distinction between knowing what and knowing how (Ryle, 1949) and between semantic and episodic memory (Tulving, 1972) may be an appropriate place to begin. Likewise, Gagné's (1977) distinction among the various types of learning and learned capabilities and Bruner's (1964) distinction among enactive, iconic, and symbolic modes of representing knowledge may serve as useful starting points. The important thing to keep in mind is that there are different types of knowledge, and some concern needs to be given to how they relate to individual differences in learning and instruction. Tobias' (1976, 1978) concern for achievement-by-treatment interactions would clearly fit in this category. Nearly all attempts to individualize instruction have focused on this source of individual differences, often at the expense of the other sources.

The second source of individual differences can be conceptualized as either a state variable or a process variable, depending on the predilections of the investigator. This category is concerned with those differences in what probably is best referred to as strategies for processing information. These strategies are methods for processing information that are presumed to be learned by the individual but are relatively stable once they have been acquired. Various personality characteristics that affect learning, including many of the cognitive-style variables, would also be included here. It is possible for individuals to learn new strategies and to replace old strategies with more effective ones,

but accomplishing this acquisition process requires a relatively long time. The concern for aptitudes that are trainable (Glaser, 1972; Snow, 1976a) would fit in this category. The strategies that are involved in this source of individual differences may be either relatively general in scope, applying to a wide variety of tasks, or relatively narrow in scope, applying to a limited range of specific tasks.

The third category involves individual differences in basic cognitive processes that are probably, for at least all practical purposes, permanent. These might be individual differences in channel capacity, reaction time, and so forth. These differences are probably physiologically based and perhaps genetically determined. Individual differences in these process variables are not affected by training or experience; nevertheless, they must be taken into account in completing our understanding about individual differences in learning and cognition and in adapting instruction to meet the needs of individual students. In some cases, the second and third categories may overlap, such as when channel capacity appears to have been increased by the learner's use of encoding strategies that increase the size of a chunk of information—for example, reducing the number of chunks from 12 to 3 by encoding 177618121941 as 1776, 1812, and 1941.

Concern for individual differences in cognitive processes is relatively new, especially within the experimental tradition of psychology. A little over 10 years ago, Melton (1967) argued for the importance of describing individual differences in terms of process variables, saying: *"What is necessary is that we frame our hypotheses about individual differences variables in terms of the process constructs of contemporary theories of learning and performance* [p. 239, italics in the original]." Although a few investigators (Glaser, 1977; Hunt, Frost, & Lunneborg, 1973; Shuell, 1972; Snow, 1976b, 1976c, and Chap. 2, Vol. 1; Sternberg, 1977; Underwood, 1975) have either advocated or followed this approach, there has been little systematic research in this area. One purpose of the current volume, of course, is to explore the possibilities and limitations of such an approach.

The most systematic attempt to date to relate individual differences to underlying cognitive or information-processing variables is the work of Snow (1976c). He has suggested that there are four categories of process differences among individuals: (1) parameter differences; (2) sequence differences; (3) route differences; and (4) summation or strategic differences. The analysis is based on the typical information-processing model consisting of various stages involved in the processing of information. Parameter differences refer to those individual differences that exist within a given stage, such as differences in capacity of short-term memory or time required for stimulus encoding. Sequence differences refer to those differences that might exist between individuals in the order in which the various stages are involved in processing information; for example, one individual might generate hypotheses early in the sequence of stages, whereas another might wait until later in the sequence to generate hypotheses. Both individuals, however, would utilize all of the stages; only the order in which they

are involved would differ. Route differences, on the other hand, would involve qualitative differences in the stages actually used by various individuals; for example, one individual might use visual rotation or double checking, whereas another individual would not use these stages. These three categories, however, may not be adequate for describing individual differences in those complex types of learning and problem solving with which we are ultimately concerned. Thus, the summation or strategic differences category was included to handle those more molar aspects of information-processing models, and it includes gross differences in how individuals assemble and structure the program systems that they use, in contrast to route differences that represent differences within the same basic program.

A somewhat similar analysis is based on R. Sternberg's (1977) componential approach to the study of intelligence and analogical reasoning. (See Chapter 9, Volume 1, for his extension to deductive reasoning.) In componential analysis, a complex task (which might be a test item) is analyzed in terms of the components involved in performing the task and the rules used for combining the components. For example, in solving the analogy $A:B::C:D$, four components might be identified. An estimate of how much time is required to perform the last component of the task is obtained by allowing an individual as much time as desired to study the $A:B::C$ part of the analogy. When he or she indicates a full understanding of that part, the complete analogy is presented, and the time required to indicate the appropriate answer is recorded. Scores representing individual differences on each component can then be related to one another and to other batteries of individual-difference measures. Five sources of individual differences are suggested: (1) individual differences in number of components used in performing a task; (2) individual differences in the rules used for combining the components; (3) differences in the order in which the components are processed; (4) differences in the mode used for component processing; and (5) differences in component time or power. These sources of individual differences, along with more detailed explanations and examples, are presented in Table 24.5.

In trying to isolate sources of individual differences within various processes or stages, however, one must be careful of known relationships among the stages. For example, it is sometimes assumed that there are individual differences in memory corresponding to the substantial individual differences that are so obvious in learning. This assumption may appear to be supported by the individual differences obtained on tests with the word *memory* in their title. These tests, however, usually fail to consider the difference between the concepts of learning and memory (Shuell & Keppel, 1970; Underwood, 1964). If individual differences in how well the material is "learned" are taken into account, there is absolutely no indication of individual differences in memory corresponding to the individual differences obtained in learning (Shuell & Giglio, 1973; Shuell & Keppel, 1970; Shuell & Lee, 1976).

TABLE 24.5
Sources of Individual Differences in Componential Analysis[a]

Level	Source	Explanation	Example
Theory	Components	Some subjects use more components, fewer components, or different components than other subjects.	Subject 1 solves problem using components a, b, c; Subject 2 solves problem using components c, d, e.
	Combination rule for components	Some subjects combine components according to one rule, others according to a different rule.	Subject 1 combines components additively: $a + b + c$. Subject 2 combines them multiplicatively: $a \times b \times c$.
Model	Order of component processing	Some subjects order components in one sequence, others in a different sequence.	Subject 1 orders components a, then b, then c; Subject 2 orders them c, then b, then a.
	Mode of component processing	Some subjects process particular components in one mode, others in another mode.	Subject 1 processes *component a in self-terminating* mode; Subject 2 processes component a exhaustively.
Component	Component time or power	Some subjects process particular components more quickly or more powerfully than do other subjects.	Subject 1 is quicker in processing component a than is Subject 2.

[a] From Sternberg (1977, p. 69).

There are numerous ways that individual differences might be organized into a taxonomy that would have some usefulness both for developing a theory of individual differences in learning and cognition and for designing adaptive instructional environments. Unfortunately, no such taxonomy currently exists. In addition to the various analyses already discussed, Snow (1976c) has suggested a *hierarchical organization of abilities in which general mental ability is pro-* gressively broken down into various abilities at lower levels. The implications of these different taxonomies for instructional purposes is not completely clear at present. We need to develop a theoretical framework for conceptualizing individual differences in learning and cognition that possesses some relevance for instructional theory and for the subsequent decisions that must be made in designing adaptive instructional environments.

ADAPTIVE INSTRUCTION

Before beginning any serious discussion of adaptive instruction, some concern should be given to exactly what it is that we are trying to accomplish. Is it academic equality? Do we want everyone to achieve the same outcomes or goals? Are we trying to eliminate individual differences among persons? Or are we trying to let all people maximize their accomplishments? If the latter, would we be willing actually to implement an instructional program that improved the performance of all students but increased the difference between the most capable and the least capable?

These are issues that are integrally related to any program of adaptive instruction, but they are seldom discussed. It is clearly beyond the scope of this chapter to discuss these issues at any length. Nevertheless, they provide an important context for a discussion of research on adaptive instruction. The present social milieu emphasizing equality of educational opportunity and the history in this country of social equality and mobility serve to influence our thinking and research on the topic, and they can even confuse the issue if we are not careful.

Investigators seldom articulate their assumptions in discussing these matters or discuss them in a rather vague fashion. The writings of a number of investigators imply that their goal is to help all people achieve the same goals—to eliminate or drastically reduce individual differences. Even when the investigators have a reasonably clear understanding of the issues involved, many readers bring their own assumptions with them and assume that the goal of adaptive instruction is to eliminate differences in achievement among individuals. Cronbach and Snow (1977) "urge the social planner to be concerned not with running a fair competition but with running a talent-development operation that will bring everyone somewhere near *his* or *her* highest level of contribution (with due regard to distributional requirements of the society) [p. 8, italics in original]." Glaser (1977) says: "An educational environment that is adaptive to the individual learner assumes different ways of succeeding and many goals

available from which to choose. It assumes further that no particular way of succeeding is greatly valued over the other [p. 17]." But are most researchers and educators willing to make these same assumptions? They probably are, although the picture is not all that clear. Over the years a number of authors, including Carroll (1967), have raised the question of whether or not we actually want to eliminate individual differences; wouldn't it be better in the long run to encourage diversity? We must also heed the warning of Bereiter (1969) and Carroll (1967) that adaptive instruction that is truly effective is likely to increase rather than decrease differences among individuals.

Designing adaptive learning environments that are optimal for individual students is a complex task. All of the concerns voiced earlier in this chapter must be taken into account and integrated in some meaningful fashion. These concerns include the relationship between psychological process variables responsible for learning (including corresponding affective components) and appropriate instructional variables, as well as the effect of various types of individual-difference variables. ATI research is in its infancy, and nearly all attempts that have been made to individualize instruction have been somewhat limited in scope and have yet to capture either the richness or the complexity that will ultimately characterize an effective program of adaptive instruction. One important issue that needs to be considered at this point, in order to improve the present situation, is how to conceptualize the various functions that are involved in adaptive instruction, and the interactions among them, in such a way that they can be integrated in some systematic and meaningful manner.

Effective individualization of instruction does not necessarily require the learner to work either independently or all alone. Group work and participation in teacher-led discussions and lectures are not contrary to the basic concepts or requirements of adaptive instruction. The basic tenet of adaptive instruction is that learning experiences provided for individual students should be tailor-made to their particular needs and requirements. In most practical situations, groups of students will be found who are similar enough to one another that at least part of their learning time can be spent in group situations, although the same group of students may not always be involved. Managerial effectiveness and economic factors must also be considered. In addition, as noted earlier, certain types of objectives either require group situations or are most effectively acquired in a group setting.

Consideration must also be given to the manner in which instructional decisions are made. In some systems of individualized instruction, the student makes the appropriate instructional decisions, whereas in other systems the decisions are made by the teacher for the student. Which approach is most effective very likely depends both on the objective being learned and the nature of the instructional decision. Atkinson (1972) compared several instructional strategies for determining the sequence of word pairs to study in learning a second-language vocabulary. A strategy based on a decision-theoretic analysis of the instructional task and a mathematical model of learning resulted in better performance than a

strategy in which the student was allowed to determine independently how best to sequence the material. The generality of this finding is not completely clear. If the objective of the instructional unit was to have the student learn how to make appropriate instructional decisions independently, then it would seem reasonable to expect that he or she should be allowed to make at least some of the decisions. The necessity for the teacher to make higher-order instructional decisions may still be involved, but who should make what decisions is not always clear at present.

There have been several attempts within the general framework being discussed in this chapter to characterize adaptive instruction. Cronbach (1967) outlined five different educational approaches to the general problem of adaptive education. Glaser (1976a, 1977) has developed flow diagrams of five different types of adaptive educational programs. These approaches range from a fixed educational system that adapts to differences among individuals by letting them continue in the system without modification until they are no longer successful (at which point they leave the system), to educational systems that accommodate individual differences by providing both different goals and different routes to those goals depending on the individual needs and aptitudes of the student.

There have been several major attempts to develop operational systems of adaptive instruction. These include Individually Guided Education (IGE) (Klausmeier, 1975); PLAN* (Program for Learning in Accordance with Needs) (Flanagan, Shanner, Brudner, & Marker, 1975); and the various curriculum programs of Individualized Prescribed Instruction (IPI) (Glaser & Rosner, 1975). Learning for Mastery (Block, 1971; Block & Burns, 1976; Bloom, 1976) and the Personalized System of Instruction (PSI or the Keller Plan) (Block & Burns, 1976; Keller & Sherman, 1974) must also be included. All of these programs have tried to individualize instruction by focusing primarily on individual differences in the knowledge students have at the beginning of an instructional unit or curriculum program. Although IGE, IPI, and PLAN* all explicitly acknowledge the desirability of adapting on the basis of differences in learning styles, strategies, and so forth, most discussions of this component are usually rather vague, or the adaptations are made in terms of student choices among activities such as listening to a taped story, reading story booklets, playing games, or working with other manipulable materials. Part of the difficulty is undoubtedly the present state of the art, but it is now time to move seriously in the direction of incorporating process differences as well as content differences into adaptive instructional programs. When process differences are considered, it is usually in terms of the rate at which different students learn, which at best is a very crude index of aptitude differences in the psychological processes responsible for learning.

The design or selection of learning environments that are most appropriate for the learning and instructional needs of individual students should be based on the following factors:

1. The type of knowledge or information that the learner is trying to acquire.
2. Psychological knowledge regarding the way individuals learn and process new information.
3. Individual differences of the learner, including all three sources of individual differences suggested in Table 24.4.
4. Information on how various instructional methods can be appropriately matched to the other factors in order to optimize the learning of the individual student.

There are a variety of ways that matches between the learner's characteristics and the optimal learning environment for that student can be made. Before any type of effective matching can be done, however, it must be possible to classify the characteristics of both students and learning environments in ways that will permit a meaningful match. Taxonomies of both relevant individual differences and appropriate task environments are required. The souces of individual differences presented in Table 24.4 and the taxonomies suggested by Snow (1976c) and by Sternberg (1977) are initial attempts at the former, but to date almost no work has been done on developing an appropriate taxonomy of instructional tasks and/or learning environments.

Three general ways of characterizing ATI matches between learner aptitudes and instructional methods have been suggested (Salomon, 1972; Snow, 1970). These are referred to as capitalization, compensation, and remediation. Capitalization is a match that builds on the strengths of the learner. For example, several studies (e.g., Domino, 1971) have indicated that students who "achieve via conformity" do best in courses where the teacher requires conformity, whereas students who "achieve via independence" do best in courses that encourage independence. A match made on this basis would capitalize somehow on the strengths of preferences of the learner. Compensation refers to a match in which the instructional treatment does something for the learner that he or she cannot do alone. For example, let's take the hypothetical case in which a teacher puts detailed notes on the chalkboard or distributes a mimeographed lecture outline for students who are low in memory ability. Finally, remediation refers to those situations in which the learner is provided with knowledge or skills that he or she is lacking but is capable of learning and that are prerequisites for the instructional unit being presented to the class.

It is likely, however, that an attempt to match on a unitary factor may prove to be impossible or undesirable. Combinations of the foregoing matches are possible (Cronbach & Snow, 1977), and there are likely to be situations when an apparent mismatch would be most appropriate (Messick, 1976). The objective that the learner is trying to achieve must be considered when making an appropriate match, and there may be times when the desired objective is antagonistic to the learner's preferred or optimal style of learning. This would be especially true when the objective has to do with the learner acquiring a particular type of

aptitude—for example, trying to improve the spatial ability of a high-verbal, low-spatial person. Because it is usually desirable to match on several different factors simultaneously, careful consideration of the various factors and the interactions among them is required. Multiple outcomes, as well as multiple sources of individual differences, must also be considered. Certain types of instructional treatments may maximize one type of outcome while minimizing another; a different treatment may do just the opposite (Mayer and Greeno, 1972; Olson, 1972; Walker & Schaffarzick, 1974).

IMPLICATIONS FOR FUTURE RESEARCH

Most of what has been presented in this chapter has been theoretical in nature and rather speculative. Much detail remains to be worked out, but the framework suggested does provide, in my opinion, a promising basis for guiding future research on individual differences in cognitive learning and instruction.

It should be noted that the general approach being taken in this chapter is rather different from the one suggesting that although the development of instructional theory incorporating individual differences and ATIs is feasible, the resultant theories will necessarily be local in nature and will consist to a very large extent of formative evaluations of instructional programs within a given school district or locale (Cronbach, 1975; Snow, 1977). That position argues that generalized scientific theorizing about instruction and individual differences is virtually impossible because of the complexity involved.

There is no question that the problem is a complex one, but that does not necessarily mean that it is impossible or undesirable. Perhaps part of the difficulty is that we have been designing our research hypotheses in terms of factors that are likely to be unproductive because they have been conceptualized with little concern for the psychological processes involved in the situation. Instead, there has been a tendency to hook onto variables that are highly visible, that have considerable face validity—at least at first blush—but that on closer analysis fail to take into account adequately those process variables most directly related to that in which we are ultimately most interested—in this case, learning from instruction.

The existentialistic approach advocated by Snow and Cronbach can make important contributions to both educational practice and our understanding of ATIs. Presumably, the approach would be guided by theory, but no attempt would be made to develop an integrated theory. The present chapter argues that such an integrated theory is both desirable and possible. If there is any regularity at all in the study of individual differences—and any attempt other than a totally idiographic case-study approach that tries to understand the nature of individual differences must make such an assumption—then it is worthwhile to attempt to articulate as explicitly as possible the interrelationship among the factors being

investigated. Many, many problems of a conceptual, theoretical, methodological, and practical nature are involved, but the development of an integrated theory and a corresponding base of empirical evidence regarding individual differences, cognitive processes, learning, and instruction is sufficiently important to give up for something easier and more practically expedient.

What, then, are some of the issues that need to be tackled? Certainly one important issue is the development of a more detailed taxonomy of information-processing/learning variables, instructional variables, and individual differences relevant to these concerns. The interrelationship among these various domains is extremely important. Determination of the level of analysis most appropriate for both research activities and practical application must be accomplished. The interrelationships among variables may be hierarchical in nature, multidimensional, or even a combination of the two. The development of a viable taxonomy, however, must not be viewed as an end in itself. Its main purpose would be to guide research efforts designed to further our understanding of the factors represented in the taxonomy and their application to ongoing educational practices.

The need to develop an appropriate taxonomy of instructional environments is especially important, for it would help to guide attempts to develop new aptitudes based on those factors in the instructional environment perceived to be most important. Glaser (1972) has pointed out the need to develop new types of aptitudes, a point also made by Cronbach and Snow (1977). Carroll's (1976) analysis of psychometric tests in terms of the cognitive processes required to perform well on the test is an important step in this direction. The need to move toward aptitudes defined in terms of learning variables and cognitive processes is clear. This will require, in many cases, developing completely new psychometric instruments designed to discriminate among individuals who perform well differentially in various instructional environments.

The need to develop a viable instructional theory should not detract from the equally important tasks of developing a theory of teaching and especially a theory or technology of design. An instructional theory is only the first step in developing and implementing a science of design.

The work outlined is formidable, to say the least. I believe, however, that the task is both worthwhile and ultimately feasible, and I remain optimistic that it will prove to be a promising undertaking.

REFERENCES

Anderson, J. R., Kline, P. J., & Beasley, C. M., Jr. *A theory of the acquisition of cognitive skills* (Tech. Rep. No. 77-1). New Haven, Conn.: Yale University, 1978.

Anderson, R. C., Spiro, R. J., & Montague, W. E. (Eds.). *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Atkinson, R. C. Ingredients for a theory of instruction. *American Psychologist*, 1972, 27, 921-931.

Atkinson, R. C., Herrmann, D. J., & Wescourt, K. T. Search processes in recognition memory. In

R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Bereiter, C. The future of individual differences. *Harvard Educational Review,* 1969, *39,* 310-318.

Block, J. H. (Ed.). *Mastery learning: Theory and practice.* New York: Holt, Rinehart & Winston, 1971.

Block, J. H., & Burns, R. B. Mastery learning. In L. S. Shulman (Ed.), *Review of research in education* (Vol. 4). Itasca, Ill.: Peacock, 1976.

Bloom, B. S. *Human characteristics and school learning.* New York: McGraw-Hill, 1976.

Bower, G. H. Cognitive psychology: An introduction. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 1). New York: Wiley, 1975.

Bransford, J. D., & Franks, J. J. Toward a framework for understanding learning. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 10). New York: Academic Press, 1976.

Bruner, J. S. The course of cognitive growth. *American Psychologist,* 1964, *19,* 1-15.

Bruner, J. S. *Toward a theory of instruction.* New York: Norton, 1966.

Carroll, J. B. Instructional methods and individual differences. In R. M. Gagné (Ed.), *Learning and individual differences.* Columbus, Ohio: Merrill, 1967.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "structure of intellect." In L. B. Resnick (Ed.), *The nature of intelligence.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist,* 1957, *12,* 671-684.

Cronbach, L. J. How can instruction be adapted to individual differences? In R. M. Gagné (Ed.), *Learning and individual differences.* Columbus, Ohio: Merrill, 1967.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. *American Psychologist,* 1975, *30,* 116-127.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods.* New York: Irvington, 1977.

Domino, G. Interactive effects of achievement orientation and teaching style on academic achievement. *Journal of Educational Psychology,* 1971, *62,* 427-431.

Dunkin, M. J., & Biddle, B. J. *The study of teaching.* New York: Holt, Rinehart & Winston, 1974.

Fiedler, M. L. Bidirectionality of influence in classroom interaction. *Journal of Educational Psychology,* 1975, *67,* 735-744.

Flanagan, J. C., Shanner, W. M., Brudner, H. J., & Marker, R. W. An individualized instructional system: PLAN*. In H. Talmage (Ed.), *Systems of individualized education.* Berkeley, Calif.: McCutchan, 1975.

Fleishman, E. A. On the relation between abilities, learning, and human performance. *American Psychologist,* 1972, *27,* 1017-1031.

Fleishman, E. A. Toward a taxonomy of human performance. *American Psychologist,* 1975, *30,* 1127-1149.

Frederiksen, N. Toward a taxonomy of situations. *American Psychologist,* 1972, *27,* 114-123.

Gage, N. L. Paradigms for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

Gage, N. L. Theories of teaching. In E. R. Hilgard (Ed.), *Theories of learning and instruction* (Sixty-third yearbook, Part I, National Society for the Study of Education). Chicago: University of Chicago Press, 1964.

Gagné, R. M. Military training and principles of learning. *American Psychologist,* 1962, *17,* 83-91.

Gagné, R. M. *The conditions of learning* (3rd ed.). New York: Holt, Rinehart & Winston, 1977.

Glaser, R. Individuals and learning: The new aptitudes. *Educational Researcher,* 1972, *1*(6), 5-13.

Glaser, R. Cognitive psychology and instructional design. In D. Klahr (Ed.), *Cognition and instruction.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976. (a)

Glaser, R. Components of a psychology of instruction: Toward a science of design. *Review of Educational Research,* 1976, *46,* 1-24. (b)

Glaser, R. *Adaptive education: Individual diversity and learning*. New York: Holt, Rinehart & Winston, 1977.

Glaser, R., & Rosner, J. Adaptive environments for learning: Curriculum aspects. In H. Talmage (Ed.), *Systems of individualized education*. Berkeley, Calif.: McCutchan, 1975.

Greeno, J. G. Processes of learning and comprehension. In L. W. Gregg (Ed.), *Knowledge and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Hunt, E., Frost, N., & Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 7). New York: Academic Press, 1973.

Inhelder, B., Sinclair, H., & Bovet, M. *Learning and the development of cognition*. Cambridge, Mass.: Harvard University Press, 1974.

Jamison, D., Suppes, P., & Wells, S. The effectiveness of alternative instructional media: A survey. *Review of Educational Research*, 1974, *44*, 1-67.

Joyce, B., & Weil, M. *Models of teaching*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Keller, F. S., & Sherman, J. G. *The Keller Plan handbook*. Menlo Park, Calif.: Benjamin, 1974.

Klahr, D. (Ed.) *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Klausmeier, H. J. IGE: An alternative form of schooling. In H. Talmadge (Ed.), *Systems of individualized education*. Berkeley, Calif.: McCutchan, 1975.

Klein, S. S. Student influences on teacher behavior. *American Educational Research Journal*, 1971, *8*, 403-421.

Mayer, R. E., & Greeno, J. G. Structural differences between learning outcomes produced by different instructional methods. *Journal of Educational Psychology*, 1972, *63*, 165-173.

McKeachie, W. J. Instructional psychology. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology*, 1974, *25*, 161-193.

Melton, A. W. Individual differences and theoretical process variables: General comments on the conference. In R. M. Gagné (Ed.), *Learning and individual differences*. Columbus, Ohio: Merrill, 1967.

Messick, S. Personal styles and educational options. In S. Messick (Ed.), *Individuality in learning*. San Francisco: Jossey-Bass, 1976.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, *80*, 252-283.

Newell, A., & Simon, H. A. *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Noble, C. G., & Nolan, J. D. Effect of student verbal behavior on classroom teacher behavior. *Journal of Educational Psychology*, 1976, *68*, 342-346.

Norman, D. A., Gentner, D. R., & Stevens, A. L. Comments on learning schemata and memory representation. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Olson, D. R. On a theory of instruction: Why different forms of instruction result in similar knowledge. *Interchange*, 1972, *3*(1), 9-24.

Rosenshine, B. Teaching behaviors related to pupil achievement: A review of research. In I. Westbury & A. A. Bellack (Eds.), *Research into classroom processes*. New York: Teachers College Press, 1971. (a)

Rosenshine, B. *Teaching behaviours and student achievement*. Slough, England: National Foundation for Educational Research in England and Wales, 1971. (b)

Rosenshine, B., & Furst, N. F. Research on teacher performance criteria. In B. O. Smith (Ed.), *Research in teacher education: A symposium*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Rosenshine, B., & Furst, N. The use of direct observation to study teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally, 1973.

Ryle, G. *The concept of mind*. London: Hutchinson's University Library, 1949.

Salomon, G. Heuristic models for the generation of aptitude-treatment interaction hypotheses. *Review of Educational Research*, 1972, *42*, 327-343.

Samuels, S. J. Attentional process in reading: The effect of pictures on the acquisition of reading responses. *Journal of Educational Psychology*, 1967, *58*, 337-342.

Shuell, T. J. *Individual differences in learning and retention* [Final Report: U.S.O.E. Grant No. OEG-2-700023(509)]. Buffalo, N.Y.: State University of New York, 1972.

Shuell, T. J., & Giglio, J. Learning ability and short-term memory. *Journal of Educational Psychology*, 1973, *64*, 261-266.

Shuell, T. J., & Keppel, G. Learning ability and retention. *Journal of Educational Psychology*, 1970, *61*, 59-65.

Shuell, T. J., & Lee, C. Z. *Learning and instruction*. Monterey, Calif.: Brooks/Cole, 1976.

Snow, R. E. Research on media and aptitudes. In G. Salomon & R. E. Snow (Eds.), Commentaries on research in instructional media: An examination of conceptual schemes. *Viewpoints: Bulletin of the School of Education, Indiana University*, 1970, *46*(5), 63-89.

Snow, R. E. Aptitude-treatment interactions and individualized alternatives in higher education. In S. Messick (Ed.), *Individuality in learning*. San Francisco: Jossey-Bass, 1976. (a)

Snow, R. E. Research on aptitude for learning: A progress report. In L. S. Shulman (Ed.), *Review of research in education* (Vol. 4). Itasca, Ill.: Peacock, 1976. (b)

Snow, R. E. *Theory and method for research on aptitide processes: A prospectus* (Tech. Rep. No. 2). Stanford, Calif.: Stanford University, School of Education, Aptitude Research Project, 1976. (c)

Snow, R. E. Individual differences and instructional theory. *Educational Researcher*, 1977, *6*(10), 11-15.

Sternberg, R. J. *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Taba, H. *Teacher's handbook for elementary social studies*. Reading, Mass.: Addison-Wesley, 1967.

Tobias, S. Achievement treatment interactions. *Review of Educational Research*, 1976, *46*, 61-74.

Tobias, S. *Interaction between achievement and instructional method*. Paper presented as part of a symposium entitled "Achievement, Aptitude, Arousal: Three approaches to ATI" at the annual meeting of the American Educational Research Association, Toronto, 1978.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

Underwood, B. J. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1964, *3*, 112-129.

Underwood, B. J. Individual differences as a crucible in theory construction. *American Psychologist*, 1975, *30*, 128-140.

Walker, D. F., & Schaffarzick, J. Comparing curricula. *Review of Educational Research*, 1974, *44*, 83-111.

25 Discussion: Coordinating Research Topics With Instructional Requirements

William E. Montague
Navy Personnel Research and Development Center

A major intent of the conference was to focus attention on instructional science and to begin to suggest the role(s) cognitive research and theory may play in the enterprise (cf. Federico, Chapter 1, Vol. 1). Specifically, regarding means for accommodating instruction practices to learner aptitudes, the requirement is to understand how fundamental abilities, acquired knowledges, and procedures relate to learning new tasks and control performance.

It is interesting to note that this is not the first time that psychologists have attempted to provide a theoretical base for the development of instructional science. At the turn of the century, Binet and Cattell tried, with only limited success, to measure intelligence in terms of underlying capabilities estimated from performance on simple tests like reaction time and memory span. At the same time, Thorndike and Dewey not only were concerned about components of intelligence but also assumed that application of the knowledge produced by research would be found in instructional settings. They often selected, for research, tasks from among those important in schooling. However, the bulk of psychologists interested in acquisition withdrew to the laboratory in order to build a science. In so doing, they created a gap between what they did, what they knew, and instructional practice. For the most part they were unconcerned with application of their knowledge and procedures. Except for a few attempts to remedy this (see Glaser, 1976, for a detailed discussion), the schism between what goes on in instruction in the schools and psychological models of learning and cognition continued late into the 1970s. Instructional practice is affected by research outcomes and psychological theory only very indirectly (Clifford, 1973).

The recognition that subjects can adopt many radically different methods to accomplish a task, depending on their background knowledge and minor task

303

details. has changed research perspectives and provided knowledge and tech-
niques of possibly greater relevance. The result is a shift in attention from the
deceptively "simple" laboratory tasks using large groups of subjects to the
discovery and verification of the specific methods individual subjects use to
perform complex tasks of various kinds. This has raised the hope that a substan-
tial body of knowledge and theory will result and become the basis for develop-
ing instructional theory and practice.

As we have been reminded several times during the conference, research and
the development of process theories of human cognition may provide a means for
reuniting experimental and measurement psychologies. It is apparent that the
conferees generally agree that process models are a useful way to represent the
cognitive events that underlie performance on various tasks, including standard
aptitude and achievement tests. Therefore, by analyzing test requirements and
the tasks with which they correlate, the hope is to develop a superior psychomet-
ric theory. The current emphasis on the detailed analysis of complex task re-
quirements and subjects' use of procedures and processes in response to those
requirements seems to be more representative of actual testing conditions and
may have more potential for success than the older approach. It seems likely that
this change in perspective will have a significant impact on differential psychol-
ogy and theories of individual differences.

The four chapters to be discussed exemplify the shift in emphasis and research
concerns. I do not discuss the papers in detail. Rather, I try to provide a perspec-
tive that I think is necessary for bringing instructional design considerations to
the attention of cognitive theorists. Comments on the chapters are made in that
frame of reference. Shuell (Chapter 24, this volume) describes in some detail
how process models may provide important conceptual and procedural knowl-
edge about how students learn to perform tasks that may be translatable into
instructional practice. He lays out many of the functional requirements of the task
of designing instruction and describes the need for appropriate information about
individuals in order to make it adaptive. The rules for creating adaptive instruc-
tion are unclear, however, and I doubt that their development will be very rapid.
In contrast, Millward (Chapter 23, this volume) focuses on a traditional labora-
tory research area and provides an interesting description of the transition from
simple stimulus-feature theories to a more acceptable information-processing
theory of concept induction. He outlines the older approaches where stimulus
factors were paramount in relatively simple, arbitrary tasks and describes the
transition to one where in relatively complex tasks, subjects' schemata are pre-
ferred representations of conceptual knowledge and serve as programs for action.
Although this representation is more currently acceptable, it provides no basis for
instruction. Little attention is paid to differences in schemata that might result
from differences in task context, to methods for assessing them, to how they
might differ among individuals, to the conditions that foster their acquisition and
form, or to their use in learning and performing new tasks. If we are to develop a
knowledge base as a source of prescriptions for guiding instruction, explicit

attention to these issues is necessary. Although a useful summary and position statement on the state of the art is provided, the chapter suggests only a very global basis for deriving the taxonomy of instructional variables that Shuell calls for to assist in arranging for concept learning.

As Millward indicates, the major concentration in the research area uses tasks requiring concept induction or discovery. This emphasis, regardless of its historical roots, may be inappropriate as far as providing information for improving instruction. Discovery may not be the usual or most important means of having students learn concepts. Expository methods are very common in education and are effective. I think that there is an important issue to be raised because of this observation. If cognitive research and theory is to provide the knowledge base for developing a theory of instruction, then attention must be paid to what instructional theory must do. Recently, Gagné (1978) asserted that the research community has a responsibility for attending to instructional problems and for disseminating information of instructional relevance. I would add that there is a responsibility to determine what knowledge is needed and systematically to supply it. For example, what, if any, different knowledge structures or processes are involved in discovery and expository arrangements for concept learning? If we knew, we could begin to prescribe instruction.

The bringing together of the interests of cognitive psychologists and educators that Shuell envisions probably calls for more than a meeting of the minds. As early as 1899, John Dewey (1900) called for a separate enterprise that he referred to as a "linking science." However, it has only been recently that serious attention has begun to focus on what the discipline would be like, who might be involved, and how it might operate. Glaser (1976) and Reigeluth, Bunderson, and Merrill (1978) contrast it with descriptive sciences such as cognitive science. Along with Shuell, they suggest that the primary task would be to design effective and efficient instruction by deriving prescriptions from the descriptive knowledge base.

It is apparent to me, as Shuell's discussion also suggests, that the discipline would be involved in considerably more than just developing instructional theory. Diagnostic analysis of problems of the instructional system and determining what and/or whether modifications in instruction are necessary and possible within the system's resource constraints would be most important functions. Also apparent is an implementation function. Once decisions are reached about what the problems are and how they can be alleviated, the implementation must be arranged. Lack of attention to this function can negate the effectiveness of instructional development and make it impossible to evaluate properly (Cooley, 1978). Therefore, we need both an organized body of principles for how to instruct efficiently and another to manage implementations, along with an organization to carry out the process.

I raise the issue of a design science for several reasons. It seems reasonable to expect some change in the content and methods of studies if researchers are oriented toward providing a base to be used for prescribing instruction. Probably

we would all agree that cognitive process analysis has increased our understanding about the mental components that underlay complex task performance and about how people might differ regarding these components. The chapter by Calfee and Hedges (Chapter 12, Vol. 1) provides a very clear description of what such an analysis might be and how tests might be developed to detect defects in processing components. It provides a good example of how a research situation may be used to further our understanding of cognitive processes *and* to provide information important for designing instruction. Furthermore, it suggests that remedial action can be undertaken once defects in component processes are/ detected. Therefore, further refinement of this approach may assist in determining means of adapting instruction to an individual's processing capabilities.

Many innovations in education have directly or indirectly been attempts to adjust instructional materials or procedures to student capabilities and characteristics in the hope of producing gains in learning efficiency and effectiveness. Examples range from attempts to group students into classes of relatively homogeneous aptitude levels to the individualized instruction programs enumerated by Shuell. Although these latter programs have been shown to be effective and to save time when compared with standard group instruction, the means of accommodation have been crude and have not increased our understanding of the process. In individualized programs, materials are organized into modules, and progress is monitored by module tests usually covering a limited number of objectives. Failure to pass (master) the test usually results in a recommendation to recycle through the material relevant to missed items. Sometimes elaborations of the material are presented, or it is presented in another form. In some cases, the sequencing of modules and perhaps the presentation medium are left to student choice.

Although existing programs are as effective and more efficient than regular classroom instruction, substantial improvement in effectiveness could result from testing designed to detect the processing defects that cause errors in performance. As the programs are structured now, the efficiency gains are due primarily to having well-defined goals for learning and self-pacing. Error correction relies on the student's ability to detect errors in his or her performance by reviewing the material. In complicated tasks requiring the use of complex procedures, students may be unable to determine what they are doing wrong. Rigney (Chapter 13, Vol. 1) points out that processes that mediate new learning may be unconscious. We would not expect students deficient in such processing resources to be able to detect this. Similarly, children lack many of the control processes Rigney discusses and should be less able to diagnose their own errors and correct them. Also, Shuell refers to Atkinson's work where a model was designed to determine which vocabulary items were not learned and to present them. This method produced performance superior to that when students were allowed to select items on their own. At least in some cases, then, the analysis of cognitive processes and models of their operation can assist in designing more effective

adaptive procedures. The attention that Rigney pays to cognitive learning strategies (control processes) seems well founded. As they are important in the acquisition of new tasks, studying the complexity of their operation should reveal procedures that can be used to enhance their effectiveness.

The methodologies applied to an intensive study of specific mental processes used by subjects in performing complex tasks brings the interests of psychologists and educators closer together again. There is now reason to hope that the information provided in cognitive research may provide a source for prescriptive application. Although a substantial instructional design technology has existed for some time, there is concern that it may be imprecise and in need of refinement. Unfortunately, in all likelihood, a rapid blending of cognitive psychology's knowledge and theory will not be undertaken by educators alone. Cognitive psychologists must pay specific attention to problems of arranging instruction and teaching. However, it seems to me, from the discussion at this conference and the content of the literature, that such considerations are not an intrinsic part of the context in which most cognitive research is done.

It is vitally important to know about cognitive learning strategies and their role in orienting the approach subjects use in learning new tasks. Also, schemata as representations of what is learned and how they are used in transfer are well established and important conceptions. But how are these things to be addressed? Can the knowledge structures possessed by a student be assessed at the start of learning, and what changes in them result with practice? How does a preexisting schema influence what is learned in the new task? Where is the information from research presented in such a way to answer these questions and make the task of an instructional designer easier?

The prevailing theme expressed in these last chapters is that once a good process analysis of a task has been done, the differences among individuals in their use will be ascertainable. At that time, individual differences can be examined in relation to task requirements and task arrangements in order to optimize learning or performance. There is obviously a considerable amount of research to be done. It appears to me that if these questions were a part of the research context, more rapid progress would be made toward obtaining relevant information. This would be a step in the direction of providing for an instructional science. A few years ago, Glaser (1976) voiced concern that scientific researchers often assume that questions about applications are of no interest to them and that simply reporting their experimental results or describing their theories is all that is r   red of them. The implication is that if application is to be made, then others (e.g., teachers, educators) will find the information, see its relevance, and figure out how to use it to instruct. As I indicated earlier, this approach doesn't lead to much research impact on teaching practice. For this reason, Glaser suggests that an instructional science of design is necessary. This is a call for coordinating cognitive research interests with those of designing instruction.

Recently, Greeno (1978) made a similar point. In discussing the idea that significant basic research questions and applied questions were the same, he asserted that if fundamental understanding is provided by research activity, practical implications are automatic. He exemplified this by describing his research on children's understanding of different kinds of quantitative relationships in school mathematics, and the development of a model of the processes involved. The research contributes to the development of theory about comprehension and problem solving *and* provides a basis for prescriptive design of instruction in elementary arithmetic. Obviously, from this point of view, the choice of tasks is important.

The advances in cognitive theory that seem to be most important to instructional concerns have focused on complex tasks. Because the major issue concerns how people organize information and use processes to solve problems and learn, complicated tasks (reading, story comprehension, chess problem solving, and so on) that are interesting have been selected for studying the mental processes involved. This research provides knowledge about processes, about differences among individuals, and about task structure in performing tasks—knowledge that has usefulness for instructional design. Why not study tasks of more direct interest and relevance to education? By doing so, theory development in cognitive science benefits, as well as our understanding of cognitive processes necessary for developing adequate instructional theory. If this is combined with a general understanding of the functional requirements of teaching and designing instruction to guide data gathering and reporting, a proper foundation for a design science of instruction is assured.

REFERENCES

Clifford, G. J. A history of the impact of research on teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally, 1973.

Cooley, W. W. Exploratory observational studies. *Educational Researcher*, 1978, *7*, 9-15.

Dewey, J. Psychology and social practice. *Psychological Review*, 1900, *7*, 105-124.

Gagné, R. M. Educational research and development: Past and future. In R. Glaser (Ed.), *Research and development and school change*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Glaser, R. Components of a psychology of instruction: Toward a science of design. *Review of Educational Research*, 1976, *46*, 1-24.

Greeno, J. G. *Significant basic research questions and significant applied research questions are the same questions*. Paper presented at the American Educational Research Association meeting, Toronto, Ontario, Canada, March 1978.

Reigeluth, C. M., Bunderson, C. V., & Merrill, M. D. What is the design science of instruction? *Journal of Instructional Development*, 1978, *1*, 11-16.

26 General Discussion: Relationships Between Aptitude, Learning, and Instruction

Robert Glaser
University of Pittsburgh

In the work of psychologists, the concepts of aptitude, learning, and instruction have been kept at a distance from one another. Attempts at integrating studies of aptitude and learning theory have had to overcome the long-standing division between the two, orginating in the different approaches taken by 19th century British and continental psychologists. As a result, mental-test technology and the experimental psychology of learning have been nurtured in separate contexts. Learning and instruction also have been divorced for some time. At the beginning of this century, the mutual benefits of jointly pursuing the experimental psychology of learning and educational psychology were recognized in the work and writings of the great psychologists of the time. This early marriage, however, was followed by the need to establish independence; experimental psychology moved into the laboratory to prove itself as a theoretical, experimental science, and educational psychology addressed the needs of educators for principles and methods of educational practice. It now appears that each partner is strong enough to attempt a reunion, and there is a blurring of the boundaries between basic theoretical research and research concerned with understanding educational phenomena.

The chapters in this book are representative of current work devoted to understanding the relationships between concepts of aptitude, learning, and instruction. The basis for this integration resides in the development of theories of cognition that provide a common set of explanatory constructs to describe the behavior involved in each of the three components. In this chapter, I comment on this integration as attempted by the authors in this volume, considering three general themes: the application of process theories to understanding the nature of

aptitude, the study of learning as it relates to fostering transitions in competence, and a general definition of adaptive instruction.

APPROACHES TO PROCESS THEORIES OF APTITUDE

A number of the chapters in this book provide a sampling of the ways in which concepts of cognitive psychology are being employed to analyze individual differences as measured on tests of aptitude and intelligence. The general aim of these investigations is to develop process models for understanding the differences in performance that appear on classes of test tasks constructed for psychometric assessment. Individual differences are examined in terms of parameters of a model, or in terms of the adequacy with which variants of the model account for the performance of different individuals. In this sense, the search for individual differences is a secondary enterprise that follows from the availability of theories of cognitive performance for the tasks used to assess aptitude and intelligence. In the work reported, three research tactics are apparent: (1) Performance on an aptitude test is correlated with performance on measures of cognitive processing derived from laboratory experiments (Hunt, Chapter 4, Vol. 1); (2) models of test-task performance are intuitively or rationally derived, and experimental work is carried out to establish their validity (e.g., Sternberg, Guyote, & Turner, Chapter 9, Vol. 1); and (3) initial experimental work is carried out that results in a tentative model or some notions about the kind of model that might evolve (Pellegrino & Glaser, Chapter 8, Vol. 1; Cooper, Chapter 7, Vol. 1). The several approaches are considered here.

Cognitive Correlates of Aptitude Test Performance

The methodology employed by Hunt and his associates essentially involves correlating aptitude test scores with performance on tasks used in the study of memory and cognition. Individuals with high and low scores on verbal aptitude tests are characterized in terms of their differences on cognitive process measures which are carefully defined in experiments on human memory. These tasks are assumed to assess the speed of accessing codes in short- and long-term memory, the duration of information in STM, STM capacity as a function of developmental level, and so forth. The directing hypothesis of this work is that these properties of cognitive performance are important differential aspects of intellectual functioning, and evidence has been assembled to show that over a wide range of intellectual ability, such processes appear to differentiate between brain-damaged, retarded, and average individuals. The findings of Hunt and his colleagues also lead them to suggest that properties of memory may differentiate between high and low verbal aptitude in the upper range of cognitive ability.

An admitted shortcoming of this research is that it minimizes the involvement of acquired knowledge. Verbal aptitude consists of information about language and the processing and manipulation of this information. Spatial aptitude consists of knowledge of familiar figures and configurations and the processing of this knowledge. In contrast, in Hunt's approach so far, the processing parameters used to characterize individual differences in aptitude are derived from laboratory tasks involving minimal and trivial content knowledge—minimal in that they do not tap the complexity of organized knowledge in long-term memory and trivial in that they utilize such highly overlearned knowledge as the recognition of the relationships between letters, numbers, and simple pictures that comprise the content of laboratory tasks. The work reported in Chapter 8 (Vol. 1) by Pellegrino and Glaser forces the conclusion that it is necessary to consider the ways in which the content and structure of knowledge influence cognitive processing in order to arrive at an adequate understanding of individual differences in cognitive abilities. Given this caveat, the work of Hunt and his colleagues does suggest that particular processing components of information handling can contribute to individual differences in verbal ability. These differences include automatic and relatively inflexible processes involved in decoding and short-term memory capacity, and more flexible processes involved in allocating attention and selecting strategic styles.

Cognitive Component Analysis of Aptitude Tasks

Sternberg, Guyote, and Turner (Chapter 9, Vol. 1) present a general theory of the component processes involved in the solution of deductive and inductive reasoning tasks. Their model is rooted in a prior rational analysis of task performance, and they collect experimental data to prove the validity of this analysis. This work on inductive reasoning tasks is of particular interest, because such tasks appear generally as items on aptitude and intelligence tests—in solution of such items as analogy, series completion, and classification and matrix problems. Inductive reasoning—the ability to formulate rules and relations based upon event instances—is obviously taken to be a significant aspect of performance in the psychometric assessment of intelligence (e.g., Spearman, 1923).

In Sternberg's work, component stages of information processing and the order of their execution are hypothesized for the various tasks, and possible models for task performance are presented. In order to test these models, quantitative estimates are obtained of the involvement of each component stage, using error and latency data. Individual differences are described in terms of parameter estimates of the components of a model or in terms of different models for different types of individuals. Stringent criteria are applied for accepting or rejecting a model in the pursuit of a hierarchical theory for various human reasoning tasks. As in the work of Hunt and his colleagues, the emphasis is on

component processes and their organization, and the involvement of content knowledge is minimized.

This approach has the general appeal of rational and quantitative analysis, but its strong top-down form of analysis may preclude both theoretical correction as well as further discovery of the psychological processes that humans actually employ in performing the tasks that are studied. The theoretical framework is overly constrained, and the model proposed will need to take more advantage of existing theories of perception, memory structure, and problem solving in order to understand more fully the postulated component stages of task performance. The details of the processes involved require a more fine-grained analysis. For example, an interesting finding requiring elaboration is Sternberg's report that the encoding component was the most time-consuming process in analogy problems, and that the encoding parameter is positively correlated with scores on standardized reasoning tests, meaning that longer encoding times are associated with higher reasoning scores. "The positive correlations suggest that better reasoners may follow a strategy whereby they encode the terms of the analogy more carefully and completely than do poor reasoners, thereby facilitating subsequent component processes on these encodings [Chapter 9, Vol. 1]."

A particular constraint may be inherent in the fact that Sternberg's models are additive, linear descriptions of performance. In contrast, the chapter by Pellegrino and Glaser indicates that this linear property may not reflect the complexities involved as tasks become more difficult for individuals to perform. Encoding ambiguities and semantic search strategies, which differ as a function of the difficulty of a task for an individual, influence performance in various ways. As task complexity increases, performance becomes less algorithmic than Sternberg's models would imply. The solution of difficult tasks takes on a heuristic problem-solving character. In problem solving, interactions occur between component processes that involve recursions through these components as a result of changing hypotheses, recoding of task features, and searching for routines that will yield results to satisfy subgoal criteria. The organization and sequencing of these cognitive activities probably depend on higher-level executive processes that have not yet been investigated in this work.

The general flexibility of task performance is reflected in a number of other chapters. Such evidence is presented by Cooper (Chapter 7, Vol. 1) in her discussion of diverse strategies for dealing with spatial information. She suggests that if we are to discover the nature of the processes used to solve items on tests of spatial abilities, we need to investigate the task-dependent nature of spatial information processing. Individuals can generate many representations for a given stimulus display, but the constraints of the particular task involved make particular sorts of representations and processing activities more or less optimal. The flexibility of humans in dealing with these constraints is conceivably a significant aspect of individual differences in test-item performance. Shifts in performance as a function of task format also are described by Kogan (Chapter

10, Vol. 1) in his study of metaphoric thinking as a cognitive style. Hunt also presents evidence showing the flexibility of processing style in the study he reports on the ability of individuals to switch from verbal to visual processing strategies in the Clark and Chase sentence comprehension task. Flexibility of performance is apparent in Frederiksen's component process model for reading (Chapter 5, Vol. 1). The patterns of intercorrelations among component cognitive skills and reading test measures suggest that individuals compensate for low efficiencies in lower-level processes—for example, encoding multiletter units—by reducing the depth of processing when visually familiar words are encountered.

General Comment

Given the foregoing approaches to the development of process theories of individual differences, my preference is neither for Hunt's cognitive correlates approach nor for Sternberg's top-down theorizing of the components of task performance. What seems most profitable is the gradual development of performance models as illustrated in the Pellegrino and Glaser chapter and the Cooper chapter. This research tactic has the advantage of allowing enough freedom in model construction so that a level of analysis can be determined at which theories of individual differences are most effectively formulated. Different levels of analysis appear to be appropriate for different purposes.

The suggestion of Hunt and his colleagues is that very basic information-processing mechanisms, the "mechanics of thought" such as speed of accessing short-term memory and the nature of memory search, may underlie important differences that exist between extremes of intellectual functioning. In contrast, in the more restricted ranges of average intellectual functioning, individual differences may reside in other aspects of performance such as problem-solving strategies, differences in knowledge structure, and the use of executive processes or metacognitive activities.

It also seems likely that the interaction of content knowledge and process will determine cognitive strategies that differentiate individual performance. Individuals develop strategies for processing a given body of knowledge—chunking and categorizing it for efficient representation, searching for appropriate knowledge-based concepts, and identifying subgoals and algorithms related to problem solution. Representational and strategic knowledge of this kind appears to characterize differences between high and low levels of performance in a number of the chapters presented (e.g., Frederiksen, Chapter 5; Greeno, Chapter 14; Kogan, Chapter 10; Pellegrino & Glaser, Chapter 8). A reasonable prediction is that individual differences in the cognitive components of aptitude measures will be more effectively analyzed as the result of variations in higher-level strategies than as the result of the more molecular aspects of elementary processes such as speed of retrieval from short-term memory. These higher-level

strategies will interact with knowledge-based declarative and procedural information to yield the cognitive basis of individual differences in cognitive competence and style.

Explicating Psychometric Findings and Aptitude-Treatment Interaction

The chapter by Snow (Chapter 2) describes various psychometric findings that he suggests a cognitive theory of aptitude must eventually explain. Snow refers to aptitude-aptitude relationships and to hierarchical theories of aptitude that have resulted from the study of patterns of test intercorrelations through factor analysis and multidimensional scaling. The persistent character of these results provides a significant body of data that might eventually be explained by a theory of cognition and that can be used for guiding process studies of individual differences. The hierarchical structures resulting from factor analytic studies provide an initial scheme for specifying sets of tasks that should manifest some commonality of cognitive processes.

Another noteworthy fact about aptitude tests is their consistent validity for predicting academic achievement in conventional instructional environments. An information-processing analysis of these relationships in terms of the skills and knowledge involved can be a significant undertaking. The explanation and recovery of factor analytic and psychometric findings represent a challenge for cognitive process theories. It is clear that the structural descriptions derived from correlational results offer little specification of the underlying cognitive mechanisms. If it is possible to establish connections between process explanations and psychometrically identified variance, then accounting for these findings could provide a confirmatory step for cognitive theories of individual differences.

Snow further reminds us that the relationships among aptitudes, instructional variables, and learning outcomes offer another set of data that challenges a process analysis explanation. A major reason for the lack of strong ATI findings in the literature has been our ignorance of the processes that relate the three components of ATI experiments—namely, the tasks measured by aptitude, the instructional activities presented, and the criterion tasks to be learned. In ATI studies, the dimensions of performance that relate these three aspects have not been carefully analyzed for a connecting set of constructs. A fallacy in ATI experimentation has been that aptitude tests are accepted solely on the basis of their names—for example, spatial orientation, spatial visualization, memory span, inductive reasoning, associational fluency, and so on. These labels are based on *surface* features of test items, with little detailed analysis of the mental processes and cognitive strategies used. As a result, in much of the ATI work, experimenters assign idiosyncratic meanings to these test labels, and use these meanings to interpret their findings. Thus, we see researchers pairing high performance on the Hidden Figures test and Thurstone's cubes with procedures that deemphasize

verbal content in instruction. But the mere absence of words in instruction (using diagrams, for example) by no means implies the presence of the abilities required on these "nonverbal" tests, and the forms of nonverbal instruction used do not show an enhancement of learning outcomes. In contrast, it is of interest to note that in studies of "cognitive style" where ready-made tests are not available, the investigator has been required to be more analytical about the processes common to both the test, the instructional treatment, and the task to be learned; as a consequence, more promising results have been obtained (cf. Bond & Glaser, 1979; Cronbach & Snow, 1977).

LEARNING AND INSTRUCTION

Fostering Transitions in Competence

A key concept in the psychology of instruction as indicated in the chapters by Rigney and Shuell (Chapters 13 and 24, respectively) is the design of conditions for learning and performance that lead to transitions from one level of competence to the next. Given an initial state of performance capability and a state of competence to be acquired, the instructional problem becomes one of designing conditions to facilitate transitions in knowledge and skill that approach a desired state of competence (Glaser, 1976a). Conditions that foster or retard the development of knowledge and skill are present whether the conditions of instruction are deliberately designed or whether the decision is made not to intervene and to let things develop "naturally and spontaneously." But even in the latter case, an instructional setting is designed by default. In any event, the task of instruction is the deliberate design of conditions for the acquisition of performance based on some theory of learning—intuitive theories built up over the years by an experienced teacher or an experienced self-learner, notions of instruction designed into a teaching device, or theories of learning constructed by psychological scientists.

Various psychological theories have suggested, directly or indirectly, how conditions might be implemented to foster the transition of states of performance to higher stages of competence. Attempting to map these various attempts onto the ideas considered in these volumes, I will describe and classify them here very briefly. This survey of theories of learning and related instructional efforts will run somewhat chronologically from the 1950s up to the present day. The concepts presented derive from attempts that have either been made directly in optimization studies or implied in training and instructional experiments.

Behavioral Theory: Statistical Learning Models. Stimulus sampling and Markov models of learning have led to optimization studies on paired-associate learning, including beginning reading and foreign language vocabulary. Tran-

sitions between states of learning are assessed by changes in response probability. The postulation of a continuous model of these changes prescribes a different instructional procedure than an all-or-none model or a mixed-state model. These alternate hypotheses about how transitions occur imply different techniques of optimization, and the resulting instructional procedures display a range of possibilities: (1) response-insensitive strategies that consider the number of learning trials that have occurred without taking account of the learner's response history; (2) response-sensitive strategies that use learner performance information to make decisions about a subsequent instructional condition; and (3) more complex instructional routines that employ learner performance data to update parameters that estimate the student's ability and the difficulty of the items to be learned (Atkinson, 1972; Atkinson & Paulson, 1972; Groen & Atkinson, 1966).

With respect to learning and the subject-matter domain involved, these statistical models make few assumptions about process or knowledge structure. Probabilities of response and rates of learning provide the information used to make decisions about the presentation of instructional experiences. States (or stages) of competence are defined by response probability, but little is said about qualitative principles of transition; any event is just as likely as any other event to drive one cycle of the process. Subject-matter constraints exist only in the sense that these models are typically applied to paired-associate types of learning.

Behavioral Theory: Programmed Instruction.   The programmed instruction paradigm attempts to optimize performance by direct use of the principles of operant conditioning, using techniques of successive approximations accompanied by contingent feedback and reinforcement. These ideas are well known and need not be described any further here (cf. Glaser, 1978; Lumsdaine & Glaser, 1960; Skinner, 1958). In general, programmed instruction, like instructional design based on statistical learning models, makes minimal assumptions about cognitive processes and minimal assumptions about the structure of subject-matter knowledge.

Behavioral Theory: Transfer Assumptions.   The theoretical concept most prevalent in this category is Gagné's learning hierarchy model, where a curriculum structure is analyzed into ordered skills and the acquisition of a subordinate skill bears a transfer relationship to a superordinate skill. The resulting optimization procedure involves learning a lower-order skill that facilitates the learning of higher-order skills. Individual differences are manifested in terms of the number of subskills that are learned at any one time—that is, the size of the learning step. A curriculum structure can be ordered into a treelike sequence of events where prerequisite knowledge and skills are specified as components integrated into higher-order performance (Gagné, 1968, 1977; Gagné & Paradise, 1961).

Transfer relationships have also been made explicit in computer-assisted instruction procedures where the simultaneous study of two or more areas of

knowledge mutually facilitate performance. An example is the Stanford Reading Program (Atkinson, 1974; Chant & Atkinson, 1973), where instruction is ordered around two basic curriculum strands—one devoted to sight-word identification and the other to phonics. Empirically, it is known that the learning rate on one strand depends on how far along the student is on the other strand, and tne optimization model assumes that the learning rate for each of the two areas depends on the difference between achievement level in the two knowledge areas. The optimization procedure in this case involves maximizing the level of achievement for some weighted average of performance on the two strands over a fixed time period for instruction. Time is controlled on each strand to derive a maximal average-learning-rate path through the subject matter. In contrast to statistical learning theory models and programmed instruction, the transfer models make somewhat deeper assumptions about learning processes and about subject-matter structure.

Cognitive Process Models.   With the opening up of the black box, further knowledge and postulations of cognitive processes and mediational variables were available to influence instructional attempts. The use of mental imagery was suggested as a means of optimizing paired-associate memorization in learning a foreign language vocabulary. In the keyword method described by Atkinson and his associates (Atkinson & Raugh, 1975; Raugh & Atkinson, 1975), the recommended procedure has a learner associate the sound of the new foreign word with a given keyword and then generate a mental image relating the keyword to the English translation.

Transition strategies are also suggested by developmental studies identifying hierarchical stages of declarative and procedural knowledge. These successive stages characterize levels of task performance during development or learning. Once the knowledge stage of a learner is identified, then cognitive activities are introduced that foster the acquisition of successively higher levels of performance. A first step in the optimization procedure involved here is to conduct a rational task analysis of how a task is performed at the most sophisticated level of competence and, also, to derive either rationally or from empirical findings the rules that govern less sophisticated stages of performance. The second step is to identify the declarative and procedural knowledge necessary for individuals to progress from one level of functioning to the use of a more advanced rule. An example of this procedure is Siegler's work on balance scale problems (Siegler, 1976, 1978). His analysis of performance on these tasks shows how individuals differ in the ability to encode and represent particular features of the problem situation. Given this information, instruction is given on where to focus attention and how to encode problem features. These learned abilities then enable the individual to detect and use higher-order rules that facilitate the transition to higher levels of performance.

Instructional possibilities, that may be neglected in traditional teaching, are suggested by analysis of the strategic knowledge required for solving problems in

a subject-matter domain. Such knowledge involves schemata and higher-order rules that assist in setting goals and forming plans. Strategic knowledge of this kind guides the search for the solution of an instructional exercise that leads to new knowledge. Greeno's study of problem solving in geometry (Chapter 14) is an illustration of this. He points out that in the usual geometry text and in classroom instruction, certain kinds of knowledge are explicitly taught, namely, (1) perceptual concepts used in recognizing figural features, patterns, and relationships; and (2) propositions used in making inferences. But a third kind of knowledge—used in setting goals and planning in the solving of problems—is assumed either to exist as a function of the student's intelligence or as a general capability gained through experience. The optimization procedure suggested by this work is to identify the nature of these strategies for problem solving in a knowledge domain, and to provide explicit instruction in their use. This kind of analysis, which attempts to detail higher-order rules for problem solving and to understand the acquisition of procedural skills, is also the focus of Chapter 18 by VanLehn and Brown.

The interesting question raised is whether procedural knowledge of this kind can become a more delineated part of instruction; the answer is not readily forthcoming. Are problem-solving strategies best taught directly, or are they to be induced by the learner in the course of a carefully designed set of examples? This has been a persistent question related to teaching the techniques of problem solving (cf. Polya, 1962; Wickelgren, 1974). Strong evidence of the success of attempts to teach general problem-solving strategies is not available, and concern continues to be expressed that while rote knowledge can be taught well, the processes of planning and problem solving are usually not. Present lines of research in cognitive psychology and artificial intelligence are investigating the possibility that these cognitive procedures can be made explicit and teachable (cf. VanLehn & Brown).

Semantic Structure Models. The analysis of information structures in the form of networks of facts, concepts, and procedures is providing another approach to instruction. The theory and techniques involved come from work in artificial intelligence concerned with knowledge structures, semantic information networks, and question-answering systems using natural language communication with computers. From a psychological point of view, this approach to instruction begins with an ideal model of the organization of knowledge as it might exist in human memory. Assuming that memory is organized in the form of a semantic network, then such a network specified in advance provides the type of organization of knowledge that is to be learned by the student (Carbonell, 1970).

Starting with a model of the ideal structure, instruction proceeds by the student's interrogation of this structure and by providing information about his or her errors that reflect a difference between the student's semantic struc-

ture and the postulated ideal structure. Diagnostic and remedial techniques are employed that eventually enable the student, when interrogated, to give the same answers that would be forthcoming from interrogating the ideal model. With this form of tutorial instruction, it is assumed that the computer model of knowledge organization and the acquired knowledge organization of the student result in essentially the same output, even though an exact match of the similarity of the two memory organizations may not be implied. Additionally, the semantic network that was first rationally imposed can be redesigned to approximate more closely the student's memory organization, and a pedagogical procedure can be determined that most effectively facilitates acquisition of a desired knowledge structure.

The tutorial interactions studied by Stevens and Collins are an example ·f this approach. The instructional tactic involved is to proceed like a human tutor who takes into account the properties of a particular subject-matter structure, increasing experience with student performance, and the effectiveness of certain instructional exercises. As instruction proceeds, more and more information is obtained about each of these aspects, and these data are used to improve the representation of the knowledge structure, to investigate the effectiveness and efficiency of diagnostic and error-correction procedures, and to generate a theory of tutorial interaction.

The emphasis in this work on changes in the representation of knowledge structure as learning proceeds is a major concept emphasized in the chapters of this book in one way or another. The development of "representational ability" is referred to in such cognitive performances as imaging, elaborating, and identifying prototypic structure. Novices in a knowledge domain differ from experts in this regard (Chase & Simon, 1973a, 1973b; Chi & Glaser, 1980; Larkin, McDermott, Simon and Simon, 1980; Simon & Simon, 1978). Expert performers appear to have powerful means for representing a task or a problem. Once an adequate representation is available, expert performance takes place with automaticity and in terms of chunked and organized routines; there appears to be less need for extensive search of the problem space and for the planning of solution strategies. Only when very difficult or novel problems occur are the latter activities significantly displayed in the performance of experts. Given this observation, it should be noted that most theoretical work in problem solving has emphasized the kind of search procedures, planning, and means–end analyses that may be more characteristic of novice than of expert performance. The initial critical aspect of expert performance, involving representation of the task situation, appears to be more difficult to study.

Future Work.   An approach not discussed so far is one that combines memory organization and cognitive process. It is likely that a combination of these two will take place in the near future, but at present, there are no reasonably well developed examples. The tactics of the two approaches should result in models of

knowledge and skill acquisition that incorporate structural aspects of memory, processes of cognition, and environmental variables that facilitate learning.

Related efforts now on the horizon are developmental studies of qualitative changes in structure that come about with higher levels of knowledge acquisition. This work reflects Piaget's general emphasis on structural changes; these are now being investigated with more precision—for example, the work by Klahr and Wallace (1976) on the emergent properties of production systems as a model of cognitive development and the work on learning mechanisms by John Anderson. That qualitative changes in information structures occur as a result of acquiring more knowledge appears to be an important area to investigate at this time. Of particular significance is the increasing amount of work that attempts to characterize the differences in performance and the underlying cognitive strategies and knowledge organizations that distinguish novices from experts in specific areas of knowledge and skill.

Another important form of instructional theorizing is the design of interactive models of *both* the tutor and the tutored. Work so far has emphasized primarily one or the other, the learner or the instructional system. A combined description of the changing relationships between these two (even when the learner is his own tutor) should bring us closer to theories of learning with important implications for instructional design.

Cognitive Processes and Learning Theory

Learning theory has been considered directly in the chapters by Anderson, Kline, and Beasley (21) and by Rigney (13). Anderson and his colleagues continue efforts to understand how humans improve their cognitive capabilities through learning. They construct a theory that is embodied as a computer simulation program, ACT, that attempts to learn the same cognitive skills as a human. The model has been applied to examples of language acquisition, the acquisition of problem-solving skills in mathematics and computer programming, and to study skills for social science texts. The underlying theoretical structure involves a propositional network representation of declarative knowledge and a production system representation of procedural knowledge. The learning of a skill mainly involves the addition and modification of the productions that take place through learning mechanisms. At least three of these mechanisms—generalization, discrimination, and strengthening—are the pervasive processes of classical learning theory. In the context of ACT, generalization is described as the process by which productions extend their range of application beyond the domain for which they were originally designated; discrimination is a corrective mechanism by which overgeneralized productions are restricted; and strengthening is the process by which successful productions gradually acquire control of processing resources and facilitate the automatization of a skilled performance. Predictions from this learning model are tested against existing and new data, and the results

obtained lead to improvements in the theory. In comparison to classical learning theory, the hope for ACT is that computer implementation will provide strong tests of the predictions as well as internal consistency of the new theory. Such rigorous measures were not available in older theories.

Rigney, based on his significant experience in the design of training aids and the conduct of training studies, is impressed with the apparent change in the relative involvement of conscious and unconscious processing as an individual acquires knowledge and skill and develops increasing proficiency. There is a transition of processing from conscious to unconscious levels that is important when accounting for the differences between novice and expert. Conscious processing proceeds in a relatively slow, constrained fashion and is supported by faster, more automatic, unconscious processes that can constitute the bulk of processing resources.

Rigney urges us to consider the idea that failures in training and in the attainment of competent performance result from a breakdown in the linkages between conscious and unconscious processing that are well established in the expert. He suggests that a general training objective for the attainment of proficiency is instruction in the use of (or task design that forces the use of) ''cognitive learning (processing) strategies'' such as imagery and elaborative strategies (see Chapter 15 by Rohwer), mnemonic techniques, orienting and self-direction strategies, and specific problem-solving strategies that can circumvent deficiencies in processing resources. Training that incorporates the use of such processing strategies facilitates the transition and continuity between conscious and unconscious processing characteristic of expertise.

Unlike Anderson et al. and Rigney, most of the authors in this book, and cognitive psychologists in general, do not directly address the problem of how behavior is acquired. Research focuses on understanding and describing the nature of performance based on prior learning and on identifying cognitive stages in the progression to higher levels of performance, but the transition mechanisms that account for changes between these levels have been little studied. One reason for this neglect indicated by Anderson et al. is that ''learning'' is usually defined in a negative exclusionary way that does not encourage the postulation of mechanisms for investigation by current cognitive theory. The classic example is the definition in Hilgard and Bower (1975). (I cite the definition from the 1975 edition, which differs very little from the 1956 definition.)

> Learning refers to the change in a subject's behavior to a given situation brought about by his repeated experiences in that situation, provided that the behavior change cannot be explained on the basis of native response tendencies, maturation, or temporary states of the subject (e.g., fatigue, drugs, etc.).
>
> The definition has the import of allowing an inference regarding ''learning'' only when a case cannot be made for another explanation. It does not state sufficient conditions for learning, since some cases of repeated experience with a situation do not produce much in the way of observable changes in responses [p. 17].

At the present time, our interpretation of the nature of learning is being changed by cognitive theories of performance and human development. As a result, the study of learning appears to be taking on the characteristics of a developmental psychology of performance changes—the study of changes that occur as different knowledge structures and complex cognitive strategies are acquired, and the study of conditions that affect these transitions in competence. Developmental psychologists studying the mechanisms of transition between developmental stages and psychologists interested in learning who are studying the mechanisms involved in acquiring knowledge and skill should begin to find certain common concepts and methodologies.

ADAPTIVE INSTRUCTION

Putting all three of the words in the title of this book together—*aptitude, learning,* and *instruction*—brings us to the notion of adaptive instruction (cf. Federico, Chapter 1, Vol. 1). If the theory and methodology involved in these three areas could take on a common conceptual basis that relates them, then systems of adaptive instruction could be realized more effectively in practice. The general concept of adaptive instruction is that the actions taken in an instructional setting (by a teacher, a student, or a teaching device) vary as a function of past and present information about a student. In order to define this enterprise, I have, in previous writings (Glaser, 1976b, 1977), described some general models illustrating different ways that instructional systems might be adaptive to student performance—particularly the extent to which a system provides different instructional programs based on assessments of the student's initial entering state and on continued updating of student performance. Five models, ordered by increasing adaptability to student performance, are briefly mentioned here.

Model 1, which can be called a selective model with a fixed instructional path, optimizes educational outcomes by selecting students whose entering ability levels indicate a high probability of attaining particular competencies in a relatively fixed instructional environment. The adaptive decision is to select or reject individuals for an instructional program on some measure that predicts their success through the program and achieving the competencies it teaches.

Model 2 is less selective than Model 1 and focuses on the development of initial competence. In this model, performance is optimized by strengthening initial ability so the individual can achieve the entering skills required by a fixed instructional program and its established competence goal. In this case, individuals are not only assessed with respect to the presence or absence of abilities that allow them to profit from the instructional program but some diagnosis also is made of the nature of these abilities. Adaptation takes place through an attempt to develop these abilities (prerequisite knowledges and aptitudes) so that an indi-

vidual's probability of success in the program of instruction provided is increased. Thus, this second model essentially attempts to improve initial competence.

Model 3 focuses on accommodating to individuals as a function of their ways of learning and the nature of their achievement. This model, like those already mentioned, holds goals constant, but it modifies instruction on the basis of entering skills. Again, assessment is made of an individual's entering competence, but in this case, the attempt is made to match abilities to different and appropriate instructional programs. The model assumes that alternative means of learning can be matched to the abilities and levels of competence of different individuals. This matching is a more or less continuous process that occurs throughout the course of learning. As information is obtained about student performance, this information is used to make decisions about instruction that will enhance the probability of a student's success in achieving the goals of the program. The goals of the program are not altered for different individuals, and the attempt is made to allow different individuals to attain generally recognized achievements through different learning experiences. Model 3 essentially accommodates to different styles, readiness for learning, and progress in attaining the goals of instruction.

Model 4 is a combination of Models 2 and 3. The probability of attainment is increased both by improving the abilities required for profiting from the instructional programs available and by providing flexible environments in these programs by which matching can occur. In this model, both initial state and continuous adaptation to the progress of learning modify the instructional program.

Model 5 is like Model 4 but different from the other models in the nature of the achievement attained at the end of the instructional program. In this model, optimization of performance considers all three aspects of instruction—entering ability, learning skills, and differential goal attainment (or qualitatively different competencies). In contrast to Model 5, Models 1 through 4 assume common goals of instruction. For example, all individuals attain certain fundamental literacies—the literacies of elementary school or particular job performance. However, instruction over the long term produces different constellations of abilities, different forms of achievement, and different goal aspirations and interests. Instructional programs, then, vary to the extent that they attempt to optimize similar (singular) or different (multiple) attainments among individuals. Some degree of each of these aspects is present in all instructional systems, but it is apparent that instructional programs change from singular to multiple as one moves from elementary to more advanced schooling. In general, singular and multiple attainment systems also represent changes in advanced education as one proceeds from learning general fundamentals to attaining high levels of individual specialization. Adaptation to individual differences, in this context, can refer to the extent to which a program of instruction encourages eventual differential achievement—that is, adapts instruction so that individuals can discover

their interests, talents, and specializations. [Flowcharts that describe in detail the various models just mentioned are presented elsewhere (Glaser, 1976b, 1977).]

The development of adaptive instructional systems requires movement away from fixed-track programs like Model 1 toward the more flexible programs outlined in the other models. Progress in this direction will rely on two kinds of work: (1) field research with experimental school programs; and (2) research and theory construction on individual differences, learning, and cognitive performance as these relate to the acquisition of complex knowledge and skill. Work of the first kind should result in the development of global models of instructional systems that link population characteristics, curriculum organization, classroom activities, and student progress (Bloom, 1976; Carroll, 1963; Cooley & Leinhardt, 1975; Suppes, Macken, & Zanotti, 1978). Techniques of causal analysis that apply more directly to field studies than to controlled laboratory situations will be useful for this purpose. The outcome to be anticipated is a macrotheory of teaching and instruction—"macro-" in the sense that it is concerned with the large practical variables dealt with in schools, such as the allocation and efficient use of time, the structure of the curriculum, the nature of feedback and reinforcement to the student, the pattern of teacher–student interaction the relationship between what is taught and what is assessed, the degree of classroom flexibility required for adapting to learner background, and the details of curriculum materials. Such variables need to be part of a theory of instruction in the same way the large variables of economic theory are applied to economic change. As theory at this level develops, it will be undergirded by the more micro-studies of human intelligence, problem solving, and learning such as fill these volumes.

ACKNOWLEDGMENTS

REFERENCES

Atkinson, R. C. Optimizing the learning of a second language vocabulary. *Journal of Experimental Psychology*, 1972, 96, 124-129.

Atkinson, R. C. Teaching children to read using a computer. *American Psychologist*, 1974, 29, 169-178.

Atkinson, R. C., & Paulson, J. A. An approach to the psychology of instruction. *Psychological Bulletin*, 1972, 78, 49-61.

Atkinson, R. C., & Raugh, M. R. An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *104*, 126-133.

Bloom, B. *Human characteristics and school learning*. New York: McGraw-Hill, 1976.

Bond, L., & Glaser, R. ATI, but mostly A and T with not much of I. (Review of *Aptitudes and instructional methods* by L. J. Cronbach & R. E. Snow). *Applied Psychological Measurement*, 1979, *1*, 137-140.

Carbonell, J. R. AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, 1970, *11*, 190-202.

Carroll, J. B. A model of school learning. *Teachers College Record*, 1963, *64*, 723-733.

Chant, V. G., & Atkinson, R. C. Optimal allocation of instructional effort to interrelated learning strands. *Journal of Mathematical Psychology*, 1973, *10*, 1-25.

Chase, W. G., & Simon, H. A. The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973. (a)

Chase, W. G., & Simon, H. A. Perception in chess. *Cognitive Psychology*, 1973, *4*, 55-81. (b)

Chi, M. T. H., & Glaser, R. The measurement of expertise: Analysis of the development of knowledge and skill as a basis for assessing achievement. In E. L. Baker & E. S. Quellmalz (Eds.), *Design analysis and policy in testing and evaluation*. Beverly Hills, Calif.: Sage Publications, Inc., in press.

Cooley, W. W., & Leinhardt, G. *The application of a model for investigating classroom processes*. Pittsburgh, Pa.: University of Pittsburgh, Learning Research and Development Center, 1975.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods*. New York: Irvington, 1977.

Gagné, R. M. Learning hierarchies. *Educational Psychologist*, 1968, *6*, 1-9.

Gagné, R. M. *The conditions of learning* (3rd ed.). New York: Holt, Rinehart & Winston, 1977.

Gagné, R. M., & Paradise, N. E. Abilities and learning sets in knowledge acquisition. *Psychological Monographs*, 1961, *75*(Whole No. 519).

Glaser, R. Components of a psychology of instruction: Toward a science of design. *Review of Educational Research*, 1976, *46*, 1-24. (a)

Glaser R. The processes of intelligence and education. In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976. (b)

Glaser, R. *Adaptive education: Individual diversity and learning*. New York: Holt, Rinehart & Winston, 1977.

Glaser, R. The contributions of B. F. Skinner to education and some counterinfluences. In P. Suppes (Ed.), *Impact of research on education: Some case studies*. Washington, D.C.: National Academy of Education, 1978.

Groen, G. J., & Atkinson, R. C. Models for optimizing the learning process. *Psychological Bulletin*, 1966, *66*, 309-320.

Hilgard, E. R., & Bower, G. H. *Theories of learning* (4th ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Klahr, D., & Wallace, J. G. *Cognitive development: An information-processing view*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. Expert and novice performance in solving physics problems. *Science*, 1980.

Lumsdaine, A. A., & Glaser, R. *Teaching machines and programmed learning: A source book*. Washington, D.C.: National Education Association, 1960.

Polya, G. *Mathematical discovery: On understanding, learning, and teaching problem solving* (Vol. 1). New York: Wiley, 1962.

Raugh, M. R., & Atkinson, R. C. A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology*, 1975, *67*, 1-16.

Siegler, R. S. Three aspects of cognitive development. *Cognitive Psychology*, 1976, *8*, 481-520.

Siegler, R. S. Cognition, instruction, development, and individual differences. In A. M. Lesgold, J.

W. Pellegrino, S. D. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction*. New York: Plenum, 1978.

Simon, D. P., & Simon, H. A. Individual differences in solving physics problems. In R. S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Skinner, B. F. Teaching machines. *Science*, 1958, *128*, 969-977.

Spearman, C. *The nature of intelligence and the principles of cognition*. London: Macmillan, 1923.

Suppes, P., Macken, E., & Zanotti, M. The role of global psychological models in instructional technology. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Wickelgren, W. A. *How to solve problems*. San Francisco: Freeman, 1974.

Author Index

Numbers in italics denote pages with complete bibliographical information.

A

Abelson, R. P., 13, *21,* 87, *94,* 193, *197*
Adams, J. C., 56, *78*
Allen, F., 123, *136*
Alper, S. N., 141, 147, *174*
Ammon, M. S., 44, *46*
Anderson, J. R., 13, *20,* 200, 201, 202, 209,
  222, *234,* 265, *273,* 280, *299*
Anderson, R. C., 280, *299*
Angell, J. R., 139, 141, *173*
Ashton, R., 139, *175*
Atkinson, R. C., 44, *45,* 86, *93,* 286, 295,
  299, 316, 317, *324, 325*
Attneave, F., 246, *273*
Austin, G. A., 258, *273*

B

Bailey, N., 215, 234
Ball, T. M., 145, *174*
Baribeau-Braun, J., 54, *81*
Barr, A., 92, *94*
Bartlett, F. C., 246, 262, *273*
Beach, L. R., 69, *78*
Bean, J. P., 24, 29, *46*
Beard, M., 92, *94*
Beasley, C. M., Jr., 280, *299*
Belmont, J. M., 25, 44, *45*

Bemesderfer, M. E., 155, 156, 157, 158, 159,
  160, *174*
Benson, D. A., 56, *78*
Bereiter, C., 295, *300*
Betts, G. H., 140, *174*
Biddle, B. J., 279, *300*
Block, J. H., 296, *300*
Bloom, B., 296, *300,* 324, *325*
Bloom, B. S., 296, *300*
Bobrow, D. G., 8, *20,* 58, *81,* 261, *273*
Boden, M., 170, *174*
Boies, S. J., 58, *81*
Bond, L., 315, *325*
Bostock, H., 72, *79*
Bott, R., 242, *244*
Bovet, M., 280, *301*
Bower, G. H., 229, *234,* 255, 272, *274,* 286,
  300, 321, *325*
Boyes-Broem, P., 256, *275*
Bragdon, H. R., 49, *79*
Bransford, J. D., 247, 248, 249, 272, *273,*
  274, 280, *300*
Braren, M., 52, *82*
Brauth, S. E., 72, *80*
Brooks, L., 252, *273*
Brown, A. L., 25, 44, *45*
Brown, J. S., 19, *21,* 105, 129, *136,* 182,
  186, 187, *196, 197,* 232, *234*
Brown, R., 225, 227, *234*
Brudner, H. J., 296, *300*

327

Subject Index

333

END

DATE
FILMED

6-81

DTIC