#881

In D.L. Medin (Ed.), The psychology
of learning and motivation: Vol. 37.
San Diego, CA: Academic Press, 1997.

GOALS, REPRESENTATIONS, AND STRATEGIES IN A CONCEPT ATTAINMENT TASK: The EPAM Model

Fernand Gobet
Howard Richman
Jim Staszewski
Herbert A. Simon

Any explanation of human performance in cognitive tasks must take account of a broad range of factors. In memory or concept learning tasks these begin, on the stimulus side, with the structure of the task itself and the knowledge that is embedded in it, then move to the subject side—to the ways in which subjects interpret the task (their representation of the goal and of the task structure)—thence to the strategies they adopt for attacking it, and finally to the capabilities, some of them "built in," some of them acquired through previous learning, that they apply to it (Newell & Simon, 1972, ch. 14).

I. Inter-Subject Differences and Commonalities in Performing Cognitive Tasks

Subjects who represent the same task in an experiment differently, employ different strategies in approaching it, or have different capabilities in the form of memory and processing capacities and relevant knowledge can be expected to exhibit different behavior while performing it. If the differences

0079-7421/97 $25.00

among subjects are small, we may be satisfied in averaging over our sample in order to provide a general, and approximate, explanation of the behavior of "typical subjects." If the differences are large, averaging over subjects will provide only a rough and fuzzy picture of the behavior and a correspondingly incomplete account of it. Even if we are primarily interested in those mechanisms that are largely shared among subjects, we are more likely to get a clear view of them if we strip off the differences before trying to discern the similarities.

One approach to sorting out the individual differences and commonalities we can expect to encounter among subjects in a particular task is to undertake deliberately to magnify the differences by inducing different groups of subjects to adopt different problem representations or different strategies, or by selecting groups of subjects who can be expected to possess quite different bodies of knowledge about the task. In this way, the remaining human commonalities are revealed as the residual, so to speak, that is shared by the different experimental conditions.

At least three significant lines of experimental work in recent decades have pursued this strategy: one line focusing on the knowledge subjects have (comparing the performance of experts with that of novices) (e.g., Chase & Simon, 1973; de Groot, 1946; Ericsson & Staszewski, 1987); a second focusing on differences in problem representation, induced by presenting the task through distinct, but isomorphic, instructions (e.g., Hayes & Simon, 1974); a third focusing on differences in strategy, induced by including strategy recommendations as part of the problem instructions (e.g., Medin & Smith, 1981).

In this chapter, we discuss this third line of work, taking as an example for analysis the paper by Medin and Smith (1981) on concept attainment, and using the EPAM (Elementary Perceiver and Memorizer) theory, a familiar computer simulation model that has been used to simulate a wide and growing range of memory tasks. We describe how EPAM can model the different strategies subjects employed—differences induced by the three sets of task instructions. In doing this, we show how a formal simulation model can embody not only mechanisms that explain human commonalities ("invariant psychological laws"), but also the mechanisms that interpret different representations and different strategies for the same problem. The model thereby provides a way of unifying the theory of the behaviors of individuals who approach a task with wide differences in knowledge, previous experience, and task interpretation.

Changes in behavior, including those we describe as "learning" may result from changes in any one or more of these components. Differences among subjects in the representations and strategies they employ may reflect differences in the knowledge they bring to the task, or differences in what they learn from feedback provided by or during the task performance.

Needs, Goals, Task Representations, Strategies

When attention has been turned to a particular need or task, one or more goals may be evoked from memory that, if attained, would meet the need or complete the task. With the goal in place, a subgoal may then be set of representing the task in such a way that the cognitive processes can go to work on it. The task representation is usually called the *problem space,* and task activity can be thought of as a *search* through the problem space that, if successful, attains the goal. As it is seldom either possible or efficient to search the problem space in a routine, systematic way, the system must also generate a *strategy* to guide the search (Newell & Simon, 1972, pp. 788–789).

We can view the process as involving the sequence:

need → *attention* → *goal* → *problem space* → *strategy* → *search*

We should not be misled, however, into thinking that problem-solving efforts typically follow these steps in an inexorable linear order. "Linear thinking" (often contrasted with "creative thinking") is a much maligned phrase that is irrelevant to the picture we have just drawn. The process does not usually advance without many detours and retreats. To the diagram above, we must add numerous feedback loops. Ideas evoked while designing the problem space may lead back to reformulation of the goal; strategies often show the problem space to be incomplete or inappropriate and cause a return to restructuring it. The system that behaves in this way therefore requires a metastrategy to monitor its progress and decide when it should reconsider earlier steps.

Efforts toward structuring the problem (creating and modifying the problem space) usually predominate during the early stages of activity, and efforts toward searching a particular problem space predominate during the later stages, but with much intermingling of all the processes, particularly when a problem presents difficulty or novelty. In a difficult problem like the Mutilated Checkerboard Problem, which requires discovery of a nonobvious problem space for solution, subjects, who initially adopt an "obvious" problem space, generally return to searching for a new problem representation after an hour or more of unsuccessful and ultimately frustrating search in the "obvious" but inappropriate space; and if they then discover a more appropriate representation, solve the problem after one or two minutes of search in the new space (Kaplan & Simon, 1990). Failure-produced frustration followed by sudden success is characteristic of so-called insight problems.

Of course, failure is not always followed by success, nor does frustration always evoke a new problem space. Prolonged failure may simply lead to a brief, lengthy, or permanent abandonment of the task. Switches from one task to another are mediated by the whole structure of emotions, motives, and external demands, which must be included in any comprehensive model of the system and its behavior.

For example, Johannes Kepler, in one of his early works, announced a law stating that the periods of revolution of the planets increased with the square of their distances from the Sun; but, after about a decade, decided that the fit of the law to the data was unsatisfactory. Resuming his search, in about a month's time he found the law that we now regard as correct (Kepler's Third Law: the period varies as the 3/2 power of the distance). We know also that during the intervening decade, Kepler's attention was distracted by other pressing matters—not the least that his mother was being tried for witchcraft! He resumed his search shortly after she was found innocent.

We have laboratory notebooks for extended periods of a few scientists (e.g., Darwin, Faraday, Krebs, but not Kepler) that cast light on these attention-control processes, at least on a coarse time scale.

II. Architecture and Learning in Task Performance

In accounting for behavior in cognitive tasks, we need to deal with several channels of causation: the influences of the task domain, the problem space, and the subjects' strategies.

A. The Task Domain

First, the behavior will depend on the characteristics of the task domain (Newell & Simon, 1972). In the case of very simple tasks, the goal and task domain essentially determine the behavior: the actor takes the "obvious" action that achieves the goal. If we know that someone's goal is to reserve time on a parking meter, we readily predict that he or she will insert one or more quarters in the slot.

Notice that, even in this case, we must make a number of implicit assumptions: that the actor is familiar with parking meters, knows what denomination of coin is needed, knows where the slot is and how to insert the coin, or can read the appropriate instructions on the meter to obtain this information. Notice also that our prediction of the behavior depends on our own access to the knowledge we assume the actor is using and our ability to apply it appropriately: in particular, that we use the same "obvious" representation for the problem that the actor uses. The use of these assumptions to make predictions of behavior is what Newell (1990) has called "prediction at the knowledge level." A great many, perhaps most, of our predictions of behavior in everyday life are made in this way, that is, by emulation; and we are often quite unaware of the assumptions we are making about the actor's knowledge and ability (or inability) to reason.

Act - R

B. Representing the Task: the Problem Space

Second, in all cases beyond those in which the behavior can be predicted at the knowledge level, the behavior will depend on the way in which the actor represents the task: the problem space. Generating an appropriate problem space for a task may vary in difficulty from the trivial (as in the previous example) to the essentially impossible (a problem space for inferring Kepler's laws of planetary motion from gravitational attraction before the invention of the calculus). In the earlier example of the mutilated checkerboard problem, we have already shown how critical the selection of problem space can be in determining the subsequent behavior—and its failure or success.

C. Strategies

Third, because the problem space prescribes only the representation of the task and not the precise way in which it will be attacked, the actor's behavior will depend on the search strategies that are adopted for exploring the space for a solution. It is almost always inefficient, and usually infeasible, for people to search a problem space either exhaustively or randomly. Various procedures are adopted to guide the search in productive directions, so that the task can be accomplished after a very small part of the total space has been examined. In favorable circumstances (e.g., solving a linear equation in algebra), the actor may know a systematic procedure (algorithm) that is guaranteed to lead to the solution after a few steps. In most less formal domains, no such algorithms exist, and problem solvers must be satisfied with selective rules of thumb (heuristics) that often lead to solutions without excessive search, but are not guaranteed to do so.

Interwoven with the goals, the problem space, and the strategies is the knowledge the actor possesses that is relevant to understanding and formulating these elements. Some of this knowledge will be obtained from the task instructions (thereby involving learning processes); much of it will already be stored in memory (as the result of previous learning).

D. All Behavior Is Social

It follows from these characteristics of the situation that in the laboratory for cognitive psychology we are studying the behavior not of otherwise undefined specimens of *Homo sapiens,* but of particular sets of human beings who, as the result of both initial endowment and a mountain of experiences since birth, carry around in their heads a large body of knowledge and skills that is mostly social in origin and is enormously variable among subject populations as we move from one culture to another or from one time to another. In this important sense, all cognitive psychology is social psychology, and we have to look hard to discover invariants of behavior that remain stable over cultures and eras.

In cognitive research, we have typically used two means to finesse this problem of the social relativity of behavior. First, we summarize the knowledge and skills of subjects drawn from a relatively homogeneous population under brief, but informative, labels like "college sophomores at XYZ University." Second, we devote a large part of our research energy (or did so traditionally) to tasks that call largely for knowledge that all members of the subject population can be presumed to possess (e.g., puzzles like The Tower of Hanoi or Missionaries and Cannibals). In recent years, in contrast, we have also moved into research on knowledge-rich task domains by studying "expert" and "novice" populations—again assuming homogeneity within each population, but admitting the possibility of large differences between the populations.

More and more, especially as our research moves in the direction of knowledge-rich tasks, we will be obliged to learn, by pretesting and in other ways, what attitudes toward goals, what information about problem representations, and what information about strategies subjects bring to the tasks we set for them. We will also use the experimental instructions and various kinds of pretraining to alter the goals, problem spaces, and strategies that subjects have available.

E. The Model Must Incorporate Attention Control

Human bounded rationality requires people to perform their tasks using only the knowledge they have, however much that may depart from the reality it purports to describe. Moreover, the inferences they will draw from their knowledge will be severely limited by their computational capabilities. Finally, of the body of knowledge they have stored in memory, only a fraction—often a very small fraction—of the knowledge potentially relevant to a particular task will be evoked initially, or even in the course of time, by the presentation of the task and the instructions.

Under these circumstances, to understand and predict behavior requires us to understand and predict what part of the information in memory will actually be evoked and applied in the course of task performance. What will subjects attend to and when? Under what circumstances will shifts of attention occur and evoke new information or lead to the loss of information previously evoked? A theory of cognition must incorporate a theory of attention, and, as we saw earlier, attention is closely linked to motivation and even to emotion.

In the remainder of this chapter, we undertake to make these ideas more concrete by showing how they enter into the modeling of behavior observed in a well-known piece of experimental work on concept attainment and categorization. As the basis for our modeling, we use the EPAM system, which, first developed about 1959 to account for a number of the phenomena of rote verbal learning, has been extended in the succeeding 35 or more years to account for a progressively wider range of phenomena of perception and memory.

III. Strategy, Goals, Attention, and Task Representation in EPAM

Each EPAM simulation has three components: (1) an experimenter module representing the activities of the experimenter; (2) a subject module representing the activities of the subject; and (3) a manager model, which coordinates the two whenever time is added to the simulated clock.

The subject module requires problem spaces and strategies much as a mathematical model requires parameters. EPAM has not yet matured to the point where it determines its own goals and builds its own problem spaces. Currently, its problem spaces and strategies are programmed for each experiment and form an adjustable component of the model.

Just as mathematical model builders seek to keep the number of numerical parameters low and their values consistent across many tasks, the EPAM programmer seeks to keep the strategies as simple and constant as possible. Strategies are changed from one simulation to another only when necessary to correspond with differences in the experimental task or differences in the directions given to the subjects.

A. Strategy

Often the results produced by EPAM are direct outcomes of the strategies chosen, and are interpreted as such. For example, the very first simulation using EPAM predicted the invariant serial position curve as a consequence of an attention strategy, called the "anchor point" strategy, used by people when familiarizing themselves with an ordered list (Feigenbaum & Simon, 1962). EPAM was programmed to learn the list by working from the anchor points inward. For most lists this led EPAM to choose the first and last elements as the obvious anchor points. The program thus demonstrated that an anchor point strategy could match quantitatively the ubiquitous serial position curve. Introduction of other, attention-attracting anchors (e.g., a word printed in red) automatically produced a von Restorff effect.

Similarly, a 1967 simulation (Gregg & Simon, 1967) cast light upon a contradictory pattern of results where otherwise-similar experiments sometimes produced "one-trial" learning and sometimes incremental learning. When EPAM employed a strategy of rehearsing one syllable pair at a time until they were completely learned, it produced one-trial learning. Where it used an "all-at-once" strategy, stopping rehearsal of one pair whenever a new pair was presented, its learning was incremental. Postexperimental reports by subjects confirmed that this choice in rehearsal strategy corresponded to individual success with one-trial learning. EPAM explained how the strategy chosen by human subjects would determine whether they learned a syllable in a single trial or by increments over several trials.

B. Goals

The goal of the subject module in many EPAM simulations is to add information to long-term memory and then later retrieve that information. For example, in a rote learning task the system gradually creates a net for paired associates indexed by the stimuli, so that when a stimulus is presented, a chunk, comprising the stimulus together with the response, can be accessed and the system can give the correct response. Similarly, in a concept-formation task the system creates a net or nets such that the stimulus will elicit a chunk or chunks indicating the correct category, and the system can respond with that category. In both tasks the subject has to associate stimuli with correct responses. The main difference between the tasks is that, whereas each stimulus elicits a different response in the paired-associate task, several stimuli all elicit the same response in the categorization task. Subjects form generalizations that put many stimuli in the same class in the latter task.

Studies of subjects' strategies in concept attainment (e.g. Bruner, Goodnow, & Austin, 1956; Hunt & Hovland, 1961; Hunt, Marin, & Stone, 1966) have shown that people try to find a simple economical criterion involving only a few features that enables the system to categorize a great many stimuli. Such economy is impossible in the paired-associate task.

Within a system like EPAM, such economy can be achieved if the system finds efficient tests while building the net. For example, if all members of a "Category A" are red, and all members of a "Category B" are blue, an EPAM-like system would seek to put a test for color at the top of its discrimination net. Then all red items would sort to a single node and all blue items to a different node.

In earlier versions of the EPAM model (versions I through IV), the system could not replace an inefficient test after it had been added to the discrimination net. The most recent version, EPAM V, permits the insertion of effective tests above inefficient tests in the discrimination net. This can occur whenever a generalization is studied.

For example, a universe of simple geometric figures defined by three attributes, size, color, and shape, could include large red circles, small red circles, large blue squares, and so on. If all red squares, regardless of size, are in category A, then a generalization can be formed that "red squares regardless of size" elicit Category A.

Assume that EPAM V had created the inefficient discrimination net shown in Fig. 1. Although EPAM can recognize small red squares as members of Category A, it currently misrecognizes large red squares as members of Category C. EPAM V, at this point, needs to replace an inefficient test for size in its discrimination net with a test for color, which it does by discovering the appropriate generalization and then studying it.

There are many processes that concept formation systems could use to discover that red squares of any size are members of Category A. For example, the CLS (concept learning) system of Hunt et al. (1966) examined members and nonmembers of a category in order to discover a combination of features that appear together in all members but not together in any

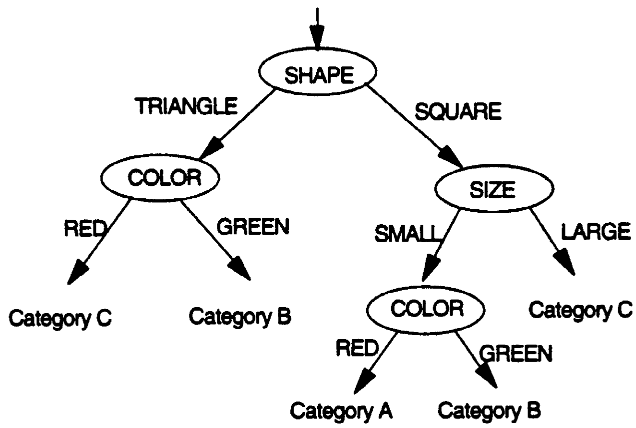

Fig. 1. An inefficient discrimination net.

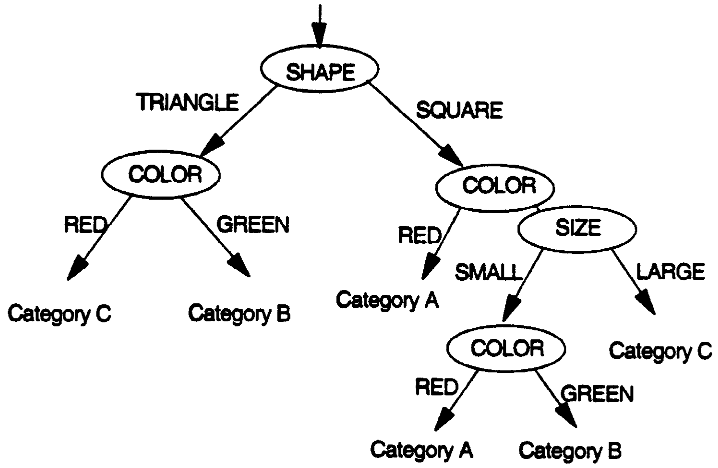

Fig. 2. EPAM's modified discrimination net.

nonmember. The RULEX (Rule-Plus-Exception) system of Nosofsky, Palmer & McKinley (1994) systematically tried single features or combinations of features in order to find a feature or a set of features that was sufficiently predictive.

A third process, that used by EPAM V in Medin and Smith's experiment, compares succeeding pairs of members of a category and forms a rule that classifies together two members (e.g., a large red square and a small red square) that are found to differ by only a single feature. We do not hold that this method is used most often by people, nor is it the only system that could be used by EPAM in order to discover rules; in fact, versions of the other systems programmed for EPAM work at least as well. The advantage of the minimal-pairs system is that it enables EPAM to simulate both the concept formation task and the rote learning task with the same default strategy.

Once a rule has been formed, EPAM studies it (see Fig. 2). When it studies "red square of any size" it sorts to the test for SIZE and inserts a test for the relevant attribute, COLOR, and a branch for red, producing the net shown in Fig. 2. Now all red squares will be correctly sorted to Category A.[1]

C. The Medin and Smith Task

In this section, we discuss the model's simulation of the experiment conducted by Medin and Smith (1981). Our main purpose is not to compare

[1] Implementational note: EPAM V permits multiple tests at a node. The new test for COLOR is simply added to the top of a list of tests at the node. The old test for SIZE remains at the same node in the net under the test for COLOR and is used when none of the branches for the test for COLOR applies.

EPAM with other models of categorization in their goodness of fit to the data, but to illustrate how directions given to subjects may be incorporated into the EPAM subject module as different strategies. These diverse strategies produce different patterns of response.

A sequence of stylized (Brunswick & Reiter, 1938) faces (see Fig. 3) are presented to subjects, the faces varying with respect to eye height (EH), eye separation (ES), nose length (NL), and mouth height (MH), and subjects are instructed to assign them to Category A or Category B. In the five stimuli that the experimenter assigned to Category A (see Table I), four have high eyes and four have long noses, while three have wide eye separation and three have high mouths. Thus, positive values (1s) on these traits tend to indicate membership in A; negative values (0s), membership in B, with high eyes and long noses being the more reliable indicators. Four of the five stimuli have three of the positive traits, and one (Face 7) has only two. Faces 4, 7, and 15 all have both of the "reliable" traits; Faces 5 and 13 have only one each.

In the four stimuli that the experimenter assigned to Category B, three have low eyes, three have short noses, three have low mouths, and two have narrow eye separation. Hence, the first three of these characteristics are criterial for B. Two of the Category B faces have two traits each that are criterial for A, and one for B (Faces 12, 2); one face (14) has one trait that is criterial for A, and two for B; one (10) has no traits criterial for A, and three for B.

We might expect that almost any scheme for learning to assign the faces to these categories will place especial weight on the two reliable traits for A and on the relative number of A traits and B traits a stimulus possesses. By whatever mechanisms these criteria are implemented, one could therefore predict that 4, 7, and 15 among the A faces, and 10 and 14 among the B

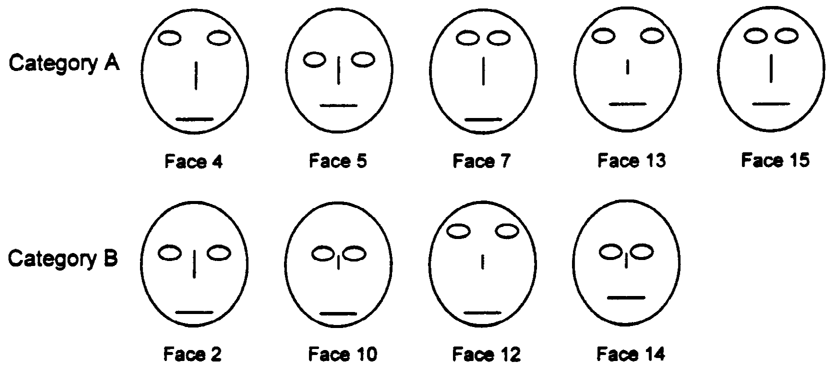

Fig. 3. Sequence of stylized (Brunswick) faces assigned to categories based on eye, nose, and mouth characteristics.

TABLE I

ATTRIBUTE STRUCTURE OF CATEGORIES USED IN THE EXPERIMENT

Face number	Attribute value			
	EH	ES	NL	MH
A exemplars				
4	1	1	1	0
7	1	0	1	0
15	1	0	1	1
13	1	1	0	1
5	0	1	1	1
B exemplars				
12	1	1	0	0
2	0	1	1	0
14	0	0	0	1
10	0	0	0	0
New transfer items				
1	1	0	0	1
3	1	0	0	0
6	1	1	1	1
8	0	0	1	0
9	0	1	0	1
11	0	0	1	1
16	0	1	0	0

Note. EH = eye height; ES = eye separation; NL = nose length; MH = mouth height. See text for explanation of binary notation. From Medin & Smith (1981).

faces would be easy to learn; while 5 and 13 among the A faces and 2 and 12 among the B faces would be hard to learn. We would not expect experimental instructions or other similar interventions to change very much this division of stimuli between "easy" and "hard." We will see that, in general, these predictions hold up well, but that we can refine them a bit further if we infer from specific experimental instructions the corresponding learning strategies the subjects will use.

D. EPAM SIMULATION OF MEDIN AND SMITH (1981)

Medin and Smith presented three groups of subjects with separate instructions for the categorization task: (1) "standard instructions," (2) "rules-plus-exception instructions," and (3) "prototype instructions." We have translated each set of instructions into a strategy programmed into the EPAM module. Each strategy has two interrelated components: (1) a recognition strategy and (2) a learning strategy. First, we look at each of these instructions and our interpretation of it by a program for the EPAM subject module. Then we compare the results of employing each of the three strategies to simulate the experimental behavior with the behavior of Medin and Smith's subjects in the corresponding experimental condition.

In proceeding in this way, we are adding additional degrees of freedom to the EPAM structure (the strategies we assume), thereby reducing its parsimony and predictive force. We have to assume that the strategies we construct are veridical interpretations of the task instructions, and a plausible case can usually be made for more than one strategy as compatible with the instructions. Selecting a particular strategy from this set is like selecting particular parameter values to fit a theory that contains parameters. We will be interested in the sensitivity or insensitivity of the predictions to postulated strategy differences.

A more powerful theory, a future EPAM, would generate the strategies automatically from the task instructions, in a manner similar to the way that the UNDERSTAND program generates problem representations from verbal task instructions (Hayes & Simon, 1974). Such a theory would have substantially fewer degrees of freedom, leaving no room for the modeler's judgment in associating strategies with instructions. But this more complete version of the EPAM theory has yet to be constructed.

The Medin and Smith experiment had three phases: a *learning phase* with feedback, which continued for 32 trials or until the subject had a single perfect trial; a *transfer phase* without feedback, which continued for 2 trials; and a *speeded-classification phase,* with feedback, for 16 trials during which time the subject's response latencies were measured.

The subject routine responds with the same algorithm for every stimulus or stimulus–response pair presented:

1. Subject waits until the next stimulus appears in its visual sensory store.
2. If the stimulus that appears states that the experiment is done, subject exits this loop.
3. Subject responds to the stimulus, using its find-category routine. This routine can vary with different instructions.
4. Subject waits for the next stimulus or response to be presented in the visual sensory store.
5. If the stimulus that appears states that the experiment is done, subject exits this loop.
6. If the next stimulus or response is a stimulus rather than a response, subject returns to Step 3. (This occurs during the transfer phase, in which feedback is not given.)
7. Subject studies the contents of the visual sensory store and the visual imagery store using the study-category routine. This routine can vary with the instructions.

8. Subject replaces the current contents of the visual-imagery store with a copy of the current contents of the visual-sensory store.

Step 9. Subject returns to Step 1.

Different find-category routines and study-category routines were used by the subject in the different experimental conditions.

E. Standard Instructions

We have identified the standard instructions in the Medin and Smith experiment with the default concept attainment strategy of the EPAM model. The standard instructions simply tell the subjects to guess at first, but then to pay attention to the feedback so that they can assign each face to its appropriate category.

EPAM's default find-category routine is to sort the stimulus in the net and report the category at the node it reaches. If it cannot find a category associated with the stimulus, then it guesses.

EPAM's default study-category routine is the following algorithm:

1. If the find-category routine responds correctly, do nothing.
2. Pick a random number from 0 to 99, if it is over 17, then do nothing.
3. If a previous study-category routine is busy transferring information to long-term memory (it takes EPAM 5 s of background learning time to add a new chunk such as a new node or response to its discrimination net), do nothing.
4. If the last stimulus (currently in the visual-imagery store) and the present stimulus (currently in the visual-sensory store) share the same response and have at least three features in common, then form and memorize a generalization consisting of the features that are on both stimuli and associate that generalization with the correct response.
5. Otherwise associate the present stimulus with the correct response.

As an example, the discrimination net produced by a particular run of EPAM in the standard instructions condition is illustrated in Fig. 4. The net is shown as it was at the conclusion of the learning phase of the experiment.

This net has a test EH for eye height at its top node and additional tests for nose length and mouth height within the net. Examination of the net reveals the following:

1. Faces 4, 7, and 15, which have EH = 1 but do not have NL = 0, are sorted to a node that states that they belong to Category A. EPAM must apply two tests in order to reach this terminal. Assuming that it takes EPAM about 10 ms to notice that a new face has been presented, 100 ms to enter a discrimination net, and 250 ms to find each of the attribute

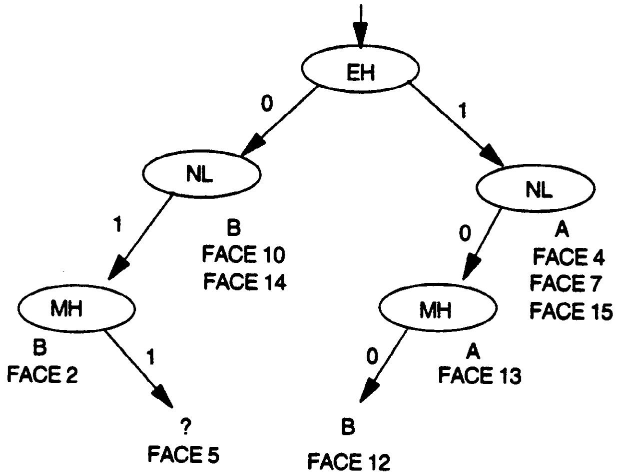

Fig. 4. Discrimination net produced by EPAM in the standard instructions condition in the Medin and Smith experiment.

values (EH and NL) using tests that require separate eye fixations, with the discrimination net of Fig. 4, it takes EPAM 610 ms to categorize Face 4, 7, or 15 as Category A.

2. Face 13 has EH = 1 and NL = 0 but not MH = 0. It is sorted to a terminal that is labeled as belonging to Category A. It takes EPAM about 860 ms to categorize this face: 10 ms to notice that a new face has been presented, 100 ms to enter the net, and 750 ms to perform the three tests.

3. Face 12 has EH = 1, NL = 0, and MH = 0. It takes EPAM 860 ms to categorize it as a member of Category B.

4. Faces 10 and 14 (EH = 0 and NL = 0) are each sorted in 610 ms to a terminal labeled for Category B.

5. Face 2 (EH = 0, NL = 1, and MH = 0) is sorted in 860 ms to a terminal labeled for Category B.

6. Face 5 (EH = 0, NL = 1, and MH = 1) sorts to an ambiguous node that is not labeled with a category. During the 26th trial of the learning phase, EPAM guessed correctly that Face 5 was a member of Category A and as a result, EPAM was able to categorize all of the faces correctly during that trial, even though it had not yet learned the correct categorization of Face 5. During both of the transfer phase trials, EPAM guessed incorrectly that Face 5 was a member of Category B. At the beginning of the 16-trial speeded-classification phase of the experiment, it took EPAM 860 ms to sort Face 5 to the node for its face, and an additional 1000 ms to guess the value of a category, for a total of 1860 ms. During this phase of the experi-

ment, however, EPAM learned the correct category for Face 5, and then only 922.5 ms was taken by EPAM to categorize Face 5.

F. Rules-Plus-Exception Instructions

The rules-plus-exceptions instructions that Medin and Smith (1981) gave their second group of subjects were much more complex than the standard instructions. The former described a two-stage learning process and a complex recognition process.

During the first stage of the learning process the subjects were told to create a rule, based on nose length, find which category long noses usually belong to, and then associate short noses with the other category. In this first stage, EPAM's subject module creates a rule net and as the first stimuli come in, determines which category is most often associated with short noses and which category with long noses. At the completion of this stage, the subject module creates a net to cache the results of this rule: If NL = 1, the net sorts to a node labeled for Category A. If NL = 0, the net sorts to a node labeled for B.

During the second learning stage subjects were instructed to memorize exceptions to the rule. EPAM stores exceptions in a second net for exceptions. The terminals in this net reports "Yes" if the item that sorts to the node is an exception and "No" if it is not.

Medin and Smith describe the rules-plus-exceptions strategy to subjects in the following general fashion:

> When you have mastered the task, you will be doing something like looking to see if the face is one of the exceptions, if so, make the correct response, if not, apply the rule for short and long noses. (p. 247)

EPAM's find-category routine uses a recognition strategy that corresponds closely to these instructions. First, it looks in the exceptions net to see whether the stimulus is an exception. Then it sorts in the rule net to find out whether the rule classifies the face as A or B. If the item is not an exception, EPAM categorizes the item according to the decision of the rule net. On the other hand, if the item is an exception, EPAM reverses this decision. If the rule has not yet been learned, EPAM accesses the hypothesis currently being held in short-term memory and categorizes the item according to the prediction that would be made by that hypothesis.

EPAM's study-category routine for the rules-plus-exceptions condition is also a two-stage algorithm. In the first stage, if the subject has not yet learned the rule, it forms a hypothesis, and studies the rule following a strategy very much like that outlined in the instructions. Specifically, it accepts a nose-length rule, such as "short nose predicts Category A," when its excess of correct over incorrect predictions exceeds 3, and accepts the opposite nose-length rule (i.e., "short nose predicts Category B") when the excess of correct over incorrect predictions dips below zero.

The second stage is very much like the find-category routine except that the system is determining whether or not the face is a member of the exceptions net. Specifically:

1. If the find-category routine responds correctly, do nothing.
2. Pick a random number from 0 to 99, if it is over 17, do nothing.
3. If a previous study-category routine is busy transferring information to long-term memory, do nothing.
4. If the last stimulus (which is currently in the visual-imagery store) and the present stimulus (which is currently in the visual-sensory store) share the same response and at least three features in common, form and completely memorize a generalization consisting of the features that are on both stimuli and label that generalization as exception or not, consistently with the present stimulus.
5. Otherwise label the present stimulus as an exception or not, as the case may be.

The two discrimination nets after completion of EPAM's learning stage in a run of the rules-plus-exceptions condition are illustrated in Fig. 5. The net on the left is a rules net with a single test for nose length. The net on the right is an exceptions net with a top test for mouth height. In this case,

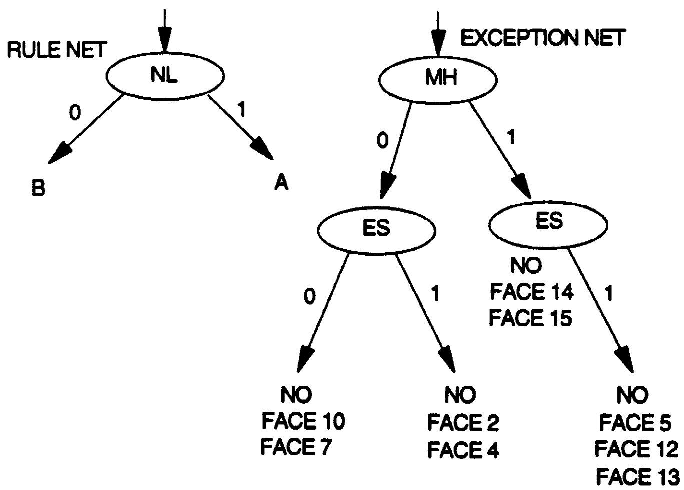

Fig. 5. Two discrimination nets after completion of EPAM's learning stage in a run of the rules-plus-exceptions condition.

EPAM went through all 32 learning trials without a perfect trial, and, indeed, using this net, EPAM currently misclassifies both of the exceptions to the rule, Faces 13 and 2.

It currently takes EPAM about 960 ms to categorize each face: 10 ms to react to the stimulus, 100 ms to enter each of the two nets, and 250 ms to sort through each of the three tests (the test for NL in the Rule net and the tests for MH and ES in the Exception net.

During the speeded-classification phase of the experiment, the system adds additional tests to the net, and these permit it to discriminate the exceptions. As a result, the average categorization latencies for Faces 4, 5, 13, 2, and 12 are higher than 960 ms and with the guessing that occurs before the categories for the new nodes are learned, the average latencies for the two exceptions, Faces 13 and 2, are over 1350 ms each.

As Faces 13 and 2 are the "exceptions" in this condition, we would expect them, with almost any strategy consistent with the instructions, to be more difficult to learn than the others. In all three experimental conditions these faces are among the three most difficult for EPAM, but they are especially difficult in the Rules and Exceptions condition. On the other hand, Face 12, which is not an exception to the long-nose rule, but is first or second in difficulty in the other two conditions, is fourth (and very much easier than Faces 13 and 2) in this condition.

G. Prototype Instructions

Medin and Smith's instructions for their "prototype" subjects were to memorize what A faces look like and what B faces look like. They were told that they later would have to answer questions about the characteristics of each type of face. EPAM memorizes types of faces by making a separate net for each type.

Medin and Smith's instructions were: "we want you to use these general impressions to help you classify these faces." EPAM's find-category routine for the prototype condition does this by sorting each stimulus in both nets. If a face is found to be a member of one category but not the other, EPAM responds with the former category. If it is found to be a member of both or is not found to be a member of either, the subject module guesses the category. If it guesses wrong, it elaborates the net for the correct category in which to include this stimulus.

EPAM's study-category routine for the prototype condition follows almost the identical strategy as the study-category routine for the standard condition, except that there is the additional step: the study-category routine must determine which net to use for studying. Specifically:

1. If the find-category routine responds correctly, do nothing.
2. Pick a random number from 0 to 99, if it is over 17, do nothing.

3. If a previous study-category routine is busy transferring information to long-term memory, do nothing.
4. If the last stimulus (which is currently in the visual-imagery store) and the present stimulus (which is currently in the visual-sensory store) share the same response and at least three features in common, form a generalization consisting of the features common to both stimuli. If the stimulus was not identified as a member of the correct net, then memorize the generalized stimulus completely in the correct net and associate it with "Yes" in that net. On the other hand, if the stimulus was identified as a member of the correct net, then it must have been misidentified as a member in the other net, so memorize the generalized stimulus completely in the other net and associate it with "No" in that net.
5. If the stimulus was not identified as a member of the correct net then associate it with "Yes" in that net. If the stimulus was identified as a member of the correct net, then memorize the stimulus completely in the other net and associate it with "No" in that net.

The two discrimination nets that resulted from one particular run of EPAM in the prototype-instructions condition are illustrated in Fig. 6.

1. Face 10 sorts to a node that identifies members of the B Net and to another node that does not identify it as a member of the A Net. The subject responds with category B in 1210 ms: 10 ms to react to the stimulus, 200 ms to enter the two nets, and 1000 ms to sort through the four tests.
2. Faces 4, 7, 13, and 15 are members of the A Net but are ambiguous in the B Net. They are correctly identified as members of Category A.

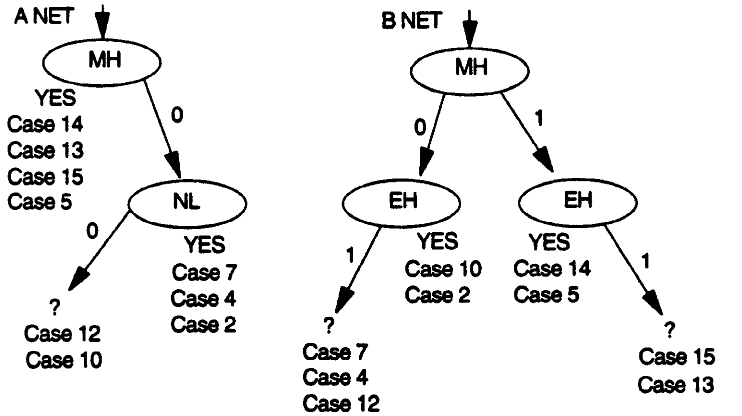

Fig. 6. Two discrimination nets resulting from one run of EPAM in the prototype-instructions condition.

3. Face 12 is ambiguous in both nets, so it is categorized by guessing. It was correctly categorized by guessing on the 32nd learning trial.

4. Faces 2, 5, and 14 are identified as members of both nets, so they are categorized by guessing. All three were correctly categorized by guessing on the 32nd learning trial.

H. Comparison with Human Data

A single free parameter called the "study parameter" was adjusted so that approximately the same proportion of EPAM-simulated subjects as real subjects would attain the same overall result in the learning condition. Medin and Smith reported that 14 of 32 people met the criterion of a perfect trial in both the standard and rules-plus-exceptions conditions, while only 8 of 32 people met the criterion in the prototype condition. With the study parameter set at 18, so that EPAM studied only 18% of the time when studying was possible, 151 of 320 simulated subjects met the criterion in the standard condition, 129 of 320 met the criterion in the rules-plus-exceptions condition, and 63 of 320 met the criterion in the prototype condition, closely matching the ratios for the human subjects. Both EPAM and people found it easier to meet the criterion in the standard and rules-plus-exceptions condition than in the prototype condition.

Table II compares Medin and Smith's human subject error data with the results from 100 runs of the EPAM model for each condition.

There was no special adjustment of parameters for this simulation, and simple and straightforward interpretations of the instructions were used for EPAM. For all three experimental conditions, the Pearsonian correlation between human subjects and EPAM of the numbers of errors for the various faces is very high: .93, .93, and .77 for the three conditions.

In EPAM as in the human experiments, Faces 13, 2, and 12 (except 12 in the Rules and Exceptions condition) produced by far the largest number of errors. There was little difference in rank order among the several conditions for either the human subjects or EPAM. In this important respect, the instructions had little effect on the outcomes, affecting only the overall level of difficulty of the task as a whole. The relative reduction in difficulty of Face 12 in the Rules and Exceptions condition was reflected in the performance of both subjects and EPAM.

Face 7 produced fewer errors than Face 4 in all conditions, which, as Medin and Smith point out, is consistent with context models but not with independent-cue models, where the net effect of cues is additive. As EPAM is, in many respects, highly nonlinear and nonadditive in its operation, hence a "context" model in the sense of Medin and Smith, we would predict this result. On average, both the subjects and EPAM found the prototype

TABLE II

MEAN NUMBER OF ERRORS FOR EACH FACE DURING INITIAL LEARNING AS A FUNCTION OF INSTRUCTIONS

	Instruction					
	Standard		Rules and Exceptions		Prototype	
Face number	People	EPAM	People	EPAM	People	EPAM
4	4.5	9.0	3.9	6.5	7.7	11.2
5	8.2	10.8	5.9	8.5	9.2	11.9
7	4.2	6.6	3.3	4.0	6.7	9.9
13	11.9	11.3	10.7[a]	18.8[a]	13.7	12.8
15	2.8	5.6	2.8	4.5	4.9	8.6
2	12.9	14.0	13.8[a]	18.5[a]	10.3	16.0
10	4.4	6.6	3.8	4.9	4.2	10.4
12	15.2	12.8	6.3	7.7	17.4	15.1
14	6.6	8.4	6.8	5.3	8.7	12.4
M	7.9	9.5	6.3	8.7	9.2	12.0
s	4.5	2.9	3.7	5.8	4.2	2.4
Pearson's r	.93		.93		.77	

[a] Face was an exception in rules-plus-exceptions condition.

condition hardest, the standard condition next hardest, and the rules-and-exceptions condition easiest.

The range of errors from the easiest to the most difficult faces was smaller for EPAM than for the human subjects in the standard and prototype conditions, but not in the rules-plus-exception conditions.

The simulation of the prototype condition is arguably the least satisfactory of the three. The errors for subjects in the standard and prototype conditions were closely similar (r = .92), suggesting that some subjects in the prototype condition may have ignored the instructions and followed a strategy much like that used by subjects in the standard condition. The correlation between the standard EPAM condition and the prototype subject condition is .81, higher than the correlation between the EPAM and subject prototype conditions (.78).

To obtain a closer fit of EPAM to the human data in the standard and prototype conditions would require EPAM to respond less promptly to the need to add new branches to the net. A change in strategy in this direction would probably also increase the relative difficulty of the harder over the easier faces. However, we have preferred to show the quite good results obtained with a strategy that was not specially "tuned" to the data.

I. Transfer Task

After completing their initial learning of the classification of the nine faces, subjects were given a transfer task, in which they were asked to categorize the same faces again, intermingled with examples of seven new faces. The results of the transfer experiment are shown in Table III.

Again, there is a close relation between the subjects' data and the EPAM simulations on the transfer test, the relation being somewhat closer for the old than for the new faces. EPAM tends to move closer to chance (50%) on the new faces, which is consistent with the fact that it sometimes guessed

TABLE III

Observed and Predicted Proportions of Correct Categorizations for Each Face during Transfer

	Instruction					
	Standard		Rules and Exceptions		Prototype	
Face number	People	EPAM	People	EPAM	People	EPAM
Old faces						
4A	.97	.83[a]	.89	.87	.77	.75
7A	.97	.96	.94	.98	.97	.85[a]
15A	.92	.96	.94	.94	.88	.90
13A	.81	.78	.72	.64	.70	.68
5A	.72	.76	.78	.77	.60	.68
12B	.67	.71	.73	.80	.45	.57[a]
2B	.72	.68	.70	.67	.72	.57[a]
14B	.97	.89	.91	.94	.83	.75
10B	.95	.96	.95	.96	.87	.89
M	.86	.84	.84	.84	.77	.74
M abs. diff.	.05		.03		.08	
New faces						
1A	.72	.54[a]	.45	.26[a]	.73	.56[a]
6A	.98	.93	.88	.74[a]	.87	.88
9A	.27	.52[a]	.08	.33[a]	.28	.50[a]
11A	.39	.58[a]	.75	.88[a]	.52	.61
3B	.44	.63[a]	.80	.91[a]	.35	.55[a]
8B	.77	.55[a]	.42	.28[a]	.78	.53[a]
16B	.91	.83	.88	.66[a]	.88	.74[a]
M abs. diff.	.17		.17		.15	

Note. In EPAM, the current hypothesis is lost from short-term memory before the transfer test. Thus, both the items that were correctly identified via the hypothesis and those correctly identified via guesses often produce errors on the transfer test.

[a] Difference between subjects and model exceeds .10.

the categories correctly in the learning experiment without extending the differentiation net to classify them unambiguously. Nevertheless, the correlation over all conditions combined between errors made by EPAM and errors made by subjects for all the different faces was .82, but .88 for the faces seen previously and .73 for the seven new faces.

Medin and Smith used regression models (a "context" or multiplicative model, and an "independent cue" or additive model) to fit the data from the transfer experiment, obtaining average absolute deviations about a third as large as EPAM's in the first case, and about half as large in the second. However, each of these models had 4 free parameters that were used in fitting the data, and were estimated separately for the three experimental conditions—a total of 12 parameters. Hence, it is hard to conclude that the regression models did a better job than EPAM of fitting the facts. Medin and Smith remark that the results were very sensitive to the exact values of the parameters, which suggests that the parameters were doing much of the work. This kind of flexibility was not available to EPAM, for the same interpretations of the instructions were used to model both the learning and transfer experiments.

J. Speeded Classification

Finally, after the subjects had completed the transfer task, they were asked to perform the classification task again with the original nine faces, but respond as rapidly as they could. In Table IV, we show the average reaction times of the subjects in responding to each face for each set of instructions, and compare these with EPAM's reaction times, without modifying any of EPAM's time parameters from their usual values.

In the three conditions, the subjects took, on average, 31, 28, and 24% longer than the EPAM simulation. Hence, the times predicted with parameters obtained from earlier studies of rote verbal learning provided a reasonable fit to the data. There is a high rank-order correlation between the times, averaged over subjects, taken to respond to the different faces and the numbers of errors they had made while learning the faces. In the three experimental conditions, there are rank-order correlations between EPAM's times on individual faces and the subjects' times of .71, .50, and .13, respectively. Thus, the speeded-classification task shows much the same pattern of findings as the two previous tasks.

K. Discussion

In this chapter we have described how EPAM, a program originally constructed to predict the behavior of human subjects in verbal learning experiments, can be used to predict behavior in categorization experiments, with-

TABLE IV

MEAN REACTION TIMES FOR CORRECT RESPONSES FOR EACH OLD FACE DURING SPEEDED CLASSIFICATION AS A FUNCTION OF INSTRUCTIONS[a]

	Instruction					
	Standard		Rules and Exceptions		Prototype	
Face number	People	EPAM	People	EPAM	People	EPAM
4	1.11	.96	1.27	1.23	1.92	1.65
5	1.34	1.02	1.61	1.33	2.13	1.68
7	1.08	.70	1.21	.93	1.69	1.39
13	1.27	1.03	1.87[b]	1.44[b]	2.12	1.64
15	1.07	.64	1.31	.95	1.54	1.22
2	1.30	1.14	1.97[b]	1.53[b]	1.91	1.78
10	1.08	.65	1.42	.96	1.64	1.25
12	1.13	1.10	1.58	1.30	2.29	1.75
14	1.19	.77	1.34	.95	1.85	1.44
M	1.17	.89	1.51	1.18	1.90	1.53

Note. Reaction times are calculated using 250 ms/test node traversed plus 250 ms/net utilized, plus 250 ms if the system guesses the category.

[a] Mean reaction times in seconds.

[b] Face was an exception in the rules-plus-exceptions condition.

out the need to modify substantially the basic learning and performance mechanisms of the system or the time parameters that predict rate of learning and speed of response. To illustrate how EPAM accomplished this, we took as an example a task that had been studied by Medin and Smith under three different conditions that corresponded to three different sets of task instructions.

This task is of special interest because it requires subjects to use a different strategy for each of the experimental conditions. Hence, the data of the subjects' performance reflect not only their own learning and response capabilities, but also differences of difficulty in categorizing the individual stimuli (characteristics of the task domain) and differences in the strategies they adopt. Although the importance for task difficulty of the task domain, the subjects' representation of the task domain (the problem space), and the strategies employed by subjects has been known for a long time (see, e.g., Newell & Simon, 1972, especially Chapter 14), there are still relatively few published experiments in which these variables have been manipulated, or in which the subjects' behavior on these dimensions have been recorded and reported.

The complex, and sometimes apparently conflicting, results that appear in the literature on concept attainment and categorization underscore the importance of understanding in as much detail as possible the processes that subjects use to formulate and attack the problems presented to them, and the differences in performance that can be produced by different choices of problem space and strategy. But in addition, as this particular set of experiments shows, the detailed structure of the task domain can itself show through into the behavior of subjects: here, largely determining the relative difficulties of the different stimulus items in a way that is predictable from the structure of the stimuli.

In our reexamination of the Medin and Smith data, we have also addressed the issues that must be faced in applying models like EPAM that, while they simulate processes in some detail, have considerable generality enabling them to model behavior over a wide range of laboratory tasks. Before any model possessing substantial generality can be employed in a particular task, a component must be added to the model to represent the task definition—its goals and constraints—and another component to represent the subjects' strategies.

In traditional mathematical modeling, these adaptations are achieved by manipulating parameters that are built into the model structure. In modeling symbolic processes, they are achieved by constructing and inserting subroutines corresponding to these components of the task. Of course, the degrees of freedom available for shaping the components cause a loss of parsimony in the theory. In the long run, the added components should not be built on an ad hoc basis for each task, but should emerge from the workings of learning processes that constitute a permanent part of the model itself. In the absence of such learning processes, effort must be taken to obtain direct evidence from the behavior of the human subjects of the problem representations and strategies they are actually using to perform the task.

References

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking.* New York: Wiley.

Brunswik, E., & Reiter, L. (1938). Eindruckscharaktere schematisierter Gesichter. *Zeitschrift fur Psychologie, 142,* 67–134.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (chap. 5). New York: Academic Press.

de Groot, A. D. (1946). *Het denken van den schaker.* Amsterdam: North-Holland.

Ericsson, K. A., & Staszewski, J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235–267). Hillsdale, NJ: Erlbaum.

Feigenbaum, E. A., & Simon, H. A. (1962). A theory of the serial position effect. *British Journal of Psychology, 53,* 307–320.

Gregg, L. W., & Simon, H. A. (1967). An information-processing explanation of one-trial and incremental learning. *Journal of Verbal Learning and Verbal Behavior, 6,* 780–787.

Hayes, J. R., & Simon, H. A. (1974). Understanding written problem instructions. In L. W. Gregg (Ed.), *Knowledge and cognition* (chap. 8). Potomac, MD: Erlbaum.

Hunt, E. B., & Hovland, C. I. (1961). Programming a model of human concept formulation. *Proceedings of the Western Joint Computer Conference, 1961,* pp. 145–155.

Hunt, E. B., Marin, J., & Stone, P. J. (1966). *Experiments in induction.* New York: Academic Press.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*(3), 374–419.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7*(4), 241–253.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review,* 101(1), 53–79.