Theories, Constraints, and Cognition

Douglas L. Medin and David M. Thau

On glancing at the Table of Contents and leafing through the chapters of this volume, readers will quickly realize that there exists a diversity of approaches to research in psychology, each offering a distinct perspective. We view this pluralism as healthy, and our goal is to add to it another exemplar. In this chapter, we describe four highly interrelated factors or strategies that have influenced our research in the study of concepts and classification learning: (a) ecological sensitivity, (b) functions, (c) constraints, and (d) formal models and theories. We also outline some interrelations among these influences using examples drawn from the area of categorization. Finally, contrary to our exemplar theorist natures, we conclude with an abstracted version of the methodology we would like to embrace.

It is important to note that the four factors we discuss act in a highly parallel fashion and that, to a certain extent, some act as checks and balances for others. Before discussing the interrelations among these factors, however, it is important to describe our particular bias on each factor considered individually.

This research was supported by National Science Foundation Grants BNS 89-18701 and BNS 89-18700. Edward Smith, Larry Barsalou, Edward Wisniewski, Woo-Kyoung Ahn, Frances Kuo, and Colleen Seifert provided guidance on this chapter.

http://dx.doi.org/10.1037/10564-007
Cognition: Conceptual and Methodological Issues, edited by H. L. Pick, Jr., P. W. van den Broek, and D. C. Knill

Ecological Sensitivity

Ecological Validity

We disagree with two extreme opinions concerning the role of ecological validity in psychology: (a) that nonnaturalistic experiments are flawed by their very nature and (b) that ecological validity should not even be a concern.

The first opinion is that if one wants to understand cognitive processes that operate in realistic, everyday situations, one should only conduct experiments that reflect the complexity of these situations. A problem with this argument is that the conclusion does not follow logically from the premise. Whatever ecological validity is, it cannot be equated with arguing against well-controlled experiments. We have never heard the claim that two confounded variables must ever more remain so because they happen to be correlated in realistic situations.

We also disagree with the premise of the argument. We do not think that our research agenda should be limited to the sort of practical questions that a layperson might find interesting. The most challenging questions about the mind typically involve processes that are so natural that we tend to take them for granted. For example, perception does not seem to be a problem because it does not occur to us to ask how a two-dimensional retinal projection gets converted into the experience of a three-dimensional world. Subjectively, the world is there and we see it, so there is nothing to be explained. Nor does it occur to us to ask just how we bring syntactic and pragmatic knowledge to bear on comprehending a sentence. People speak, and we understand them. We also know how children learn language—they imitate their parents. The mysteries arise only when we take a closer look, and our natural experience does not prompt us to do so. It is the different or unusual that catches our attention; shared cognitive abilities tend to be taken for granted.

The argument on the other side of the issue is that people have not evolved specialized cognitive modules that lead them to behave unusually in laboratory experiments. Therefore, artificiality is not a problem. Although we agree that introductory psychology has not yet been taught long enough for the natural selection of special-purpose survival strategies to take place, we think this position misses certain points.

First, any experimental situation will reveal some aspects of behavior and cognition and conceal others. For example, although one can learn something about schedules of reinforcement by putting pigeons in a Skinner box, this will not provide any information about how pigeons navigate. So, for openers, one needs to pay some attention to the world to determine which capabilities are in need of explanation.

Another important point is that ignoring real-world contexts increases the risk of failing to capture relevant information in analyses and at the risk of solving nonexistent problems. James J. Gibson argued that the view that we construct percepts by combining low-level sensory cues was a misguided consequence of elementaristic, impoverished psychological experiments (Gibson, 1979). Gibson's research program focused on an analysis of the information available in the environment, which he suggested was much more rich than people had assumed. By our reading, Gibson argued that inadequate analyses of the information available to the perceptual system may lead one to posit all sorts of complex computations to derive information that is actually already available. No analysis of perceptual processes can get very far without taking seriously the environment and the information it affords. Researchers in the area of artificial intelligence (AI), such as David Marr, were heavily influenced by Gibson's work, precisely because it addressed broad, computational-level questions (we use *computational* in Marr's, 1982, general sense, which is more abstract than its typical use in, for example, computational vision).

The Ecological and the Artificial

Although ecological sensitivity is important, it is clear that organisms are not driven by their environment alone. Because they have evolved along unique paths, different organisms will react differently to the same situation. Even at the level of sensory systems, some species are endowed with capabilities that others lack. Given that organisms have evolved mechanisms that process information from their environment, we must be concerned both with information in the environment and about the internal mechanisms that play a role in any given behavior.

Unfortunately, because organisms have evolved to cope with their environments, it is often difficult to determine whether an organism's behavior is due to cues in the environment or to some internal mechanism. It is in making this discrimination that the role of artificial situations becomes important. In our view, Shepard's (1984) evolutionary perspective on ecological constraints provides a clear example of this concern.

First of all, Shepard (1984) agreed with Gibson (1979) that organisms actively explore and manipulate their environment. He further argued that this exploration is not random but rather is guided by internal schemata. These schemata allow organisms to notice and anticipate vitally important events under conditions of impoverished information or time constraints.

The general notion is that organisms are attuned to their environment and have internalized mechanisms for dealing with relevant structure. As an example, Shepard

(1984) drew an analogy between the perceptual system and biological or circadian rhythms. The activity pattern of many animals is guided by day–night cycles that could, in principle, be directly under the control of the sun. Researchers have found, however, that when animals such as hamsters are placed under conditions of constant illumination, a very artificial situation, they continue to show 24-hr activity cycles, plus or minus only a few minutes. In short, the periodicity has become internalized so that it continues in the absence of the external stimulus, allowing the animal to anticipate the future and freeing it from depending directly on the sun. The latter would be advantageous for animals on cloudy days or in environments (e.g., the safety of a burrow) in which cues from the sun are not directly available. These rhythms are not fully independent of illumination, however, and can be "entrained" by patterns of illumination produced in the laboratory. Circadian rhythms, then, behave very much like internalized schemata that are sensitive to relevant information in the organism's environment.

Shepard (1984) suggested that the same situation holds for the perceptual system. Basically, certain structures or constraints associated with the environment (more properly, the interaction of organisms with their environment) may be internalized or embodied in the perceptual system. To observe these constraints, and to evaluate their significance, researchers need to put organisms into artificial situations in which information underdetermines performance. In these ambiguous situations, one may see natural constraints or assumptions emerge, just as circadian rhythms are observed under the uninformative situation of constant illumination.

Shepard's (1984) framework will succeed or fail on its own merits in the area of perception. Our goal is not to defend this position in the domain of perception but rather to examine its viability for higher order cognitive processes such as categorization and reasoning. Later on, we will argue by example that it does provide an effective research strategy. Note that what we call *ecological sensitivity* involves a procedure for understanding the relation of cognitive systems to their environment by using artificial, underdetermined situations and is not a blanket endorsement of artificial situations of and by themselves. Indeed, worrying about real-world circumstances may be critical for interpreting results from these artificial situations.

In short, we believe that a concern with real-world circumstances is important to ensure that laboratory results will be generalizable. Perhaps even more important, ecological considerations are critical for understanding cognition, even cognition in the laboratory. Cognitive psychologists may profitably and explicitly violate ecological validity for certain purposes, but they cannot ignore ecological considerations. As we shall demonstrate, analyses of natural situations may also provide an important source of ideas about constraints that act to guide performance in complex cognitive tasks.

Functions

One way to incorporate ecological sensitivity into our methodology is by concerning ourselves with the functions of different behaviors and processes. Although it makes sense to raise questions about function in the life sciences (as opposed to the physical sciences), these questions have not been wildly popular in cognitive psychology. In some cases, function has been implicitly assumed, and in others, it has been considered to be nothing more than idle speculation. However, Anderson (1990), one of our discipline's preeminent cognitive modelers, has recently described a strategy based on asking questions about function and then formulating computational models that satisfy or optimize this function. One of the major contributions of research on so-called "everyday memory" is that it raises questions about function, the answers to which are leading investigators along some very promising lines of research (e.g., Neisser & Winograd, 1988).

Constraints and Naturalness

Underdetermination

As we have mentioned, the environment surrounding an organism is not the only factor bearing on that organism's behavior. In fact, almost any interesting cognitive task actually involves a shortage of information from the environment. In language acquisition, for example, any linguistic input has to be consistent with the correct grammar (we will ignore the fact that speech is not always grammatical) but will necessarily be consistent with an infinite set of alternative grammars. Further sentences will not rule out many of these incorrect grammars, but any finite set of sentences will always be consistent with an unlimited number of grammars. Indeed, there are formal proofs (e.g., Gold, 1967; Pinker, 1984; Wexler & Culicover, 1980) that in its general form, the language learning problem is insolvable.

Analogous problems in learning arise in a variety of contexts. Consider a rat that eats some contaminated food in the morning, wanders around its environment, sees a cat, hears thousands of sounds, is exposed to innumerable sights and smells, and then gets sick late in the day. To what should the rat attribute its illness? Seeing the cat? The sound of rattling garbage can lids? The water it drank in the early afternoon? There are limitless possibilities, but laboratory research suggests that the rat would associate illness with the smell and taste of the food eaten in the morning and acquire an aversion to it (e.g., Garcia, Ervin, & Koelling, 1966). In general, this bias toward associating tastes and smells with illness serves the rat quite well. However, when the true association conflicts with a bias, learning may be difficult. For example, it is very hard, if not impossible, for a rat to learn to associate illness with visual cues. The general point is that organisms do not often have the luxury of running factorial experiments to determine which correlations are valid and informative. Instead, they have certain assumptions or expectations that allow some things to be readily learned and others not.

Computational Complexity

Even when possibilities can be systematically enumerated, there may be too many of them to allow an exhaustive search. The search issue comes up again and again in AI. Many AI systems are computationally explosive; the time it takes to run the program increases exponentially as the problem size increases. To reduce this problem, AI systems use heuristics and biases to reduce the number of possible choices from which a system must decide.

Complexity problems have important implications for processing models. Consider the problem of category construction in children. Nelson (1974) argued that children learn natural object categories by first constructing their own categories and then learning which labels apply to them. A model of this type of category construction could either generate and then evaluate all possible category partitions or it could be biased to only generate a subset of the possible partitions. Given that the number of ways of partitioning unclassified objects is computationally explosive (one can partition 3 objects in 5 ways, 4 in 15 ways, 5 in 52 ways, and 10 objects in more than 100,000 ways), the latter process seems more likely. Focusing on complexity issues in this way highlights places in theories in which constraints may be necessary.

Naturalness and Implicit Assumptions

What can organisms do in the face of these complexity and underdetermination problems? We see no alternative to the idea that organisms must be biased to learn some things rather than others, to draw some inferences rather than others, and in general to favor some possibilities at the expense of others. Because people cannot consider all the possibilities in a given situation, it should not be surprising that human cognition is interwoven with implicit assumptions about the world (assumptions that oversimplify but often work) and riddled with heuristics and strategies for dealing with problems and situations. In the case of categorization, one could say that people often rely on a "similarity heuristic," that is, the assumption that objects belonging to the same categories will tend to be more similar than objects belonging to different categories. Presumably, the human perceptual and conceptual system has evolved such that the similarity heuristic is usually correct.

More generally, these heuristics and implicit assumptions should render some tasks natural and easy (when people's biases fit the world) and other tasks difficult and unnatural (when their natural biases are not supported by data). Therefore, one can use people's performance in underdetermined situations to identify constraints or biases in learning. Naturalness can also serve as a guideline for evaluating theories of cognition. For example, categorization models make predictions about which kinds of partitionings will be hard for people to learn and which kinds will be easy. Theories may be judged by how well their predictions about naturalness correspond to data.

Theories and Formal Models

There are many distinct perspectives on the value of theories and formal models. Many researchers presuppose their value and importance and question the need for further discussion. On the other hand, we have heard one important branch of formal methods, mathematical psychology, talked about in the past tense (and with an air of "good riddance" at that). We believe that formal methods are of fundamental significance, and that the critical issue is how to use them properly. By formal methods we refer to any of a variety of procedures for developing, testing, and evaluating theories of cognition. These methods include at least logical and mathematical proofs and analyses, mathematical models, and models cast in the form of computer programs or simulations.

Why Formal Models Are Good

Because Intuition Is Bad

Argument by plausibility often drives our intuitions. Unfortunately, plausible argument is a rather blunt tool that often leads to mistakes. Consider, for example, some basic empirical findings from categorization research. In a line of work begun by Posner and Keele

(1968, 1970), investigators have studied the learning of ill-defined or "fuzzy" concepts. The modal procedure involves selecting some prototype (or best example) and then transforming the prototype in different ways to construct learning examples. After learning is complete, transfer tests are given that involve both old and new examples.

Three results from these procedures are very robust. First, typical examples (ones that vary little from the prototype) are more likely to be correctly categorized than are less typical examples. Second, the prototype, which is not presented with the test examples, may be classified more accurately than examples that do appear during training (Homa & Vosburgh, 1976; Medin & Schaffer, 1978; Posner & Keele, 1968). Finally, perhaps the most striking result is that over a delay interval, classification accuracy drops more rapidly for old examples than for the prototype or other new examples (Homa & Chambliss, 1975; Posner & Keele, 1970; Strange, Keeney, Kessel, & Jenkins, 1970).

These results have been interpreted as showing that (a) on the basis of experience with examples, people abstract out the central tendency or prototype for a category, and (b) classification decisions are based on similarity to this abstracted prototype. How else could prototypes be classified better than old examples, and how else could one explain differential retention?

This interpretation of the data stands in contrast to a less intuitive exemplar model. Exemplar models assume that learning involves storing examples and that classification is based on the similarity of the test items to the previously stored examples. Because it did not seem plausible that such a model could account for these data, exemplar models were either never considered or rejected by argument.

It turns out, however, that exemplar theories of categorization readily predict all these results (Hintzman & Ludlam, 1980; Medin & Schaffer, 1978). A prototype can be classified more accurately than an old example because the prototype of a category will tend to be very similar to many category examples and dissimilar to examples from contrasting categories. An old example will be maximally similar to itself but not necessarily similar to other examples from the same category and not necessarily dissimilar to examples from different categories. The same reasoning accounts for differential forgetting. Individual examples may be "on their own," whereas the prototype has many "friendly neighbors." Of course, the real test of this theory depends on whether or not mathematical or simulation models actually produce these results. They do, although not for all assumptions about forgetting (see Hintzman & Ludlam, 1980). If the exemplar model had been formalized and tested, it might not have been rejected so quickly and inappropriately.

Formal Models Indicate Where to Look for Information About Processes

As you can see, prototype and exemplar models of categorization often make similar predictions. Once these models are formalized, one can begin to ask about where to look to discover contrasts between models or, equally to the point, where not to look. For instance, there are some fairly broad conditions under which these two types of models make not just similar but identical predictions (Estes, 1986a; Nosofsky, 1990). Of course, there are other contexts in which the models make distinctive predictions, and these are the predictions to test experimentally. For instance, because prototypes only represent information about central tendencies, prototype models are insensitive to correlational information. Therefore, in a prototype model, knowledge about the average bird will not indicate that large birds are less likely to sing than small birds. Exemplar models, on the other hand, can account for this correlational information because most of the retrieved instances of large birds will not be song birds. The observation that people show sensitivity to correlation (Medin, Altom, Edelson, & Freko, 1982) provides evidence in favor of exemplar models.

Formal Models Help Conceptual Analyses

Models allow formal comparisons, and formal comparisons frequently yield rather surprising results. Consider, for example, the neural network model of category learning developed by Gluck and Bower (1988). This model is on the surface quite distinct from prior categorization models in its assumptions about representation and in its competitive learning rule. Nosofsky (in press), however, has proven that this network model is actually a special case of prototype models. This is a clear case in which formal analysis has illuminated deep underlying similarities that may have been concealed at the level of verbal description.

Formal Models Force Researchers to Be Concrete

Formal models have a built-in safeguard against vagueness. This is especially true for computational models; unless one's assumptions can be translated into steps in a program, the program will not run. Writing a program forces one to face issues and assumptions that otherwise might be hidden in informal verbal descriptions and mathematic formulas.

Formal Models Can Show Where Constraints Are Needed

As discussed in the section on constraints, computational models also serve to heighten our awareness of computational complexity problems. A program might run but take forever to come up with an answer because of the number of possibilities to be considered. In AI programming, a standard question is whether a program will "scale up," that is, continue to perform efficiently when given a larger problem or more knowledge. Programs that do scale up need heuristics (i.e., constraints) to limit the amount of possibilities that they consider.

Why Formal Models May Be Bad

Formal Models May Invite Too Narrow a Focus

To keep models from getting too complex, one may have to simplify the experimental situation so severely that what is truly of interest gets left out. As a consequence, one may end up constructing models that describe the constraints of the situation rather than the constraints of the human mind. In the study of decision making, for example, focusing on how people make choices between alternatives may ignore the problem of how people generate choices to begin with (e.g., Hogarth, 1981). To avoid modeling only task-specific strategies, recent mathematical models have placed a premium on breadth and applicability to a variety of situations (e.g., Hintzman, 1988; Raaijmakers & Shiffrin, 1981).

Even within a domain of analysis, formal models tend to focus on some phenomena at the expense of others. Mathematical models often contain (free) parameters that are estimated as part of the application of the model to data. For example, consider Bower's (1961) one element model for paired-associate learning. This model describes in great detail the consequences of assuming that learning takes place in an all-or-none manner and that the learning probability does not change across trials. At the same time, however, the theory says nothing at all about what would make the learning probability large or small. That is, even successful models are like a slice through a sphere—they reveal some things and conceal others.

Comparisons between models may engender another type of narrow focus. Because one model is often proposed as an alternative to another, the assumptions shared by both may not be questioned. In the flurry of interest in contrasting prototype and exemplar models, for example, one should not lose sight of the fact that both models invoke similarity (rather than other types of knowledge) as the basis for classification.

Formal Models May Be Taken Too Literally

When a model successfully describes a set of data, one should not confuse the explanatory "success" of the formalization with the ideas that led to the formalization. One must recognize that a variety of other ideas, perhaps quite different in character, might have led to the same formalization. For example, the decision rule associated with exemplar models of classification often takes the form of the sum of the similarities of the probe to the examples of the category of interest divided by the sum of the similarities of the probe to all stored examples. Fried and Holyoak (1984) have proposed a different model in which the learner stores information in memory that will allow him or her to make estimates of the likelihood that some probe was generated from the alternative exemplar distributions. It turns out that exemplar models of categorization are closely related to likelihood ratio models and under some conditions are not distinguishable from them (Nosofsky, 1988). In short, these two descriptions of the categorization process produced very similar formalizations. (For further illustrations and recommendations associated with the level at which models are evaluated, see Palmer, 1978.)

Formal Models May Become Opaque

When moving from mathematical models to computational models, one encounters what Smith (1978) referred to as the *sufficiency/transparency trade-off.* Any computational model comprises both a theory and some amount of extra programming necessary to make the model run. The general principles of the theory, however, may be buried in the mass of detail needed to make the model sufficient. Consequently, it may be difficult to isolate individual assumptions or components of a model and evaluate them separately. There may be an unavoidable trade-off between the clarity of a theory and its ability to handle complex problems.

We include connectionist or parallel distributed processing models as a specific type of computational model. For these models in particular, it is often difficult to identify the reason for success and failure (for a clear counterexample to this generalization, see Dell, 1986). Consequently, this area has had to use techniques (such as factor analysis) aimed at figuring out exactly what the model has learned.

Interactions and Examples

So far, we have treated the reader to a pretty abstract diet. In the following section, we link our arguments, observations, and conjectures to some specific examples.

Constraints and Models

Formal Models and Constraints on What Is Learnable

We have described the computational complexity problems that confront any computing device, including algorithms and simulation models. To our knowledge, every model for category learning has constraints or biases associated with it in the sense that the models predict that some kinds of classification problems should be easier to master than others.

One way to evaluate alternative learning models is to see whether the problems that they predict should be easy or difficult are, in fact, easy or difficult for people to master.

One constraint of interest is linear separability. A number of models, including prototype models, imply that categories must be linearly separable to be learnable. Classifying examples on the basis of similarity to a prototype basically involves summing evidence against a criterion. For example, if an instance shows a criterial number of "bird" features, it will be classified as a bird. The key is that there must be some weighted additive combination of properties that can be used to assign instances as members or nonmembers. This means that a prototype process requires that all bird examples be more similar to the bird prototype than to alternative prototypes and that nonbirds must be more similar to their respective prototypes than to the bird prototype. If a bat were more similar to the bird prototype than to the mammal prototype, it would be incorrectly classified.

Figure 1 gives a more intuitive description of linear separability. For examples that have values on two dimensions, the categories they form are linearly separable if there is a straight line that perfectly partitions them (Figure 1a). If no straight line will partition the objects (Figure 1b), then there is no way to construct prototypes such that all examples are closer to their own category prototype than to the prototype for the contrasting category.

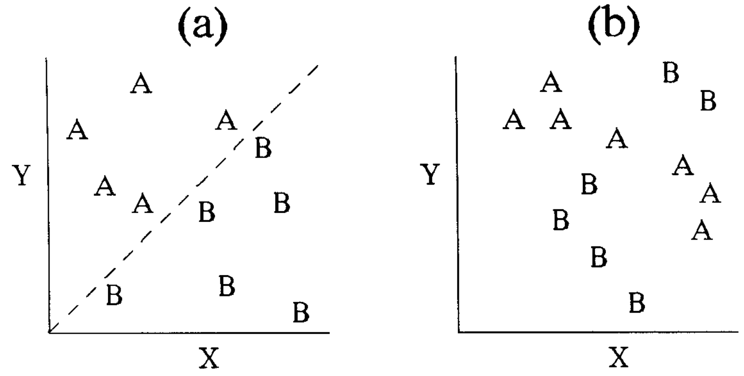


FIGURE 1 Two-dimensional example of a linearly separable category (Panel A) and a nonlinearly separable category (Panel B). (In each graph, members in Categories A and B are denoted by A and B, respectively.)

If linear separability acts as a constraint on human categorization, people should find it easier to learn categories that are linearly separable than categories that are not linearly separable. To make a long story short, studies using a variety of stimulus materials, categories, subject populations, and instructions have failed to find any evidence that linearly separable categories are learned more easily than are nonlinearly separable categories (e.g., Kemler-Nelson, 1984; Medin & Schwanenflugel, 1981).

Many network models are also constrained by linear separability. Like prototype models, single-layered network models involve a weighted integration of input units and, in a sense, add up the evidence favoring a classification decision (Minsky & Papert, 1988). Although more sophisticated network models, those that have "hidden units," can learn categories that are not linearly separable, Gluck (1991) has found that they consistently predict that linearly separable categories will be mastered more easily than nonlinearly separable categories.

We hasten to add that these results are not a problem for network models as an entire class. One can make alternative assumptions about how the input is encoded (e.g., Gluck & Bower, 1988) or how examples are represented (e.g., Kruschke, 1990), neither of which lead to a linear separability bias. The point of our example is that one can compare human performance and theories at the broad level of constraints. Our models of category learning should be as unbiased by linear separability as are people.

Formal Models and Constraints on What Is Learned

In addition to focusing on what people can learn, constraints may also be found by studying what people do learn.

Medin and Ross (1989) have argued that induction should be conservative. By *conservative* they meant that abstractions should preserve more than the minimal information necessary to perform a task. Exemplar models fare better than prototype models mainly because prototype models do not conform to conservative induction (see Nosofsky, in press, for a review of the comparisons made between prototype and exemplar models). For example, a prototype representation discards information concerning category size, variability of examples, and within-category correlations of properties. There is good evidence that people are sensitive to all three of these types of information (e.g., Estes, 1986b; Flannagan, Fried, & Holyoak, 1986; Fried & Holyoak, 1984; Medin & Schaffer, 1978; Medin & Shoben, 1988).

It is important to note, however, that exemplar models do produce abstract information. The key difference between abstraction in exemplar and prototype models is that exemplar models integrate information at the time of retrieval rather than at the time of storage. During a new–old recognition task, for example, an old cue may access both its own representation in memory and those of similar stored exemplars. A new–old judgment for such a cue will be based on a conglomeration of several exemplars, and thus may be incorrect. Indeed, some of the strongest support for exemplar models comes from classification experiments in which new-old recognition is barely above chance (Smith & Medin, 1981).

The moral of this story is not that exemplar models are better than prototype models. Instead, the point is that attention to the information that is preserved or lost in classification tasks provides clues about what constraints exist in categorization processes.

Environmental Sensitivity and Function

Cognitive psychologists often ask people to make similarity judgments, and the standard assumption is that these judgments reflect computations in terms of matching and mismatching features. Goodman (1972), however, has argued that this notion of similarity is too unconstrained to be useful because one always needs to specify the respects in which two things are similar. Indeed, the most prominent theory of similarity to date, Tversky's (1977) contrast model, describes how selected features are evaluated but says nothing about how these features are selected in the first place. An important clue to "establishing respects" may be provided by an analysis of the functions that similarity comparisons serve for organisms in their natural contexts.

Glucksberg and Keysar (1990) suggested that similarity comparisons may act like similes in important ways. Similes are directional comparisons that involve assertions. For example, saying that butchers are like surgeons asserts something very different from saying that surgeons are like butchers. Our recent research on similarity judgments is motivated by the idea that similarity is less a computation across a predefined set of features than a comparison in which the goal is to determine the relevant respects. These respects are an essential part of a speaker's message when he or she asserts that one thing is like another.

Like similes, similarity comparisons may be directional. For example, people rate the similarity of crayons to pencils to be greater than the similarity of pencils to crayons. We interpret this as an example in which the "respects" considered vary with the direction of the comparisons and modify one's impression of similarity. Specifically, in evaluating the similarity of crayons to pencils, one focuses on salient properties of pencils and considers whether these properties are also true of crayons. A salient property of pencils is that one writes with them. One can also write with crayons. In the reverse comparison, a salient property of crayons is that one colors with them, and it is not clear that one can color with pencils. With respect to coloring, then, pencils and crayons are not very similar.

"Discovering" respects may also influence similarity judgments. Medin and Goldstone (1991) recently asked people to rate the similarity of terms on a 9-point scale, with 9 being the highest similarity. Judgments were either made in two separate contexts or in a common context. In isolation, people rated the similarity of skin and hair to be 4.71 and the similarity of skin and bark to be 6.58. In a combined context, however, people rated the similarity of skin and hair to be greater than the similarity of skin and bark. These results support the idea that the comparison of skin and bark in isolation yields a sensible type of respects, leading to correspondingly high ratings. In the combined context, subjects were led to consider a different set of respects, for which skin and bark were less similar than skin and hair.

What do these results tell us about the way people make similarity judgments? We think they mean that people often answer a different question than the one being asked. When the experimenter asks, "How similar are A and B?" the subjects appear to base their answers on "how A and B are similar." That is, similarity judgments are primarily comparisons for establishing respects, not computations over respects that are predefined. If this conjecture is correct, the moral is clear. What people do in normal, more naturalistic contexts may intrude on and heavily influence performance in laboratory contexts. This appears to be as true for similarity judgments by people as it is for the activity patterns of hamsters under conditions of constant illumination.

Constraints, Functions, and Models

In principle, anything might be a constraint. For example, rats could be biased to associate illness with sounds instead of tastes. To guide one's search for constraints, it is often useful to think in terms of natural environments and what functions might be served by a particular bias. Although thinking in terms of adaptation may in principle only shift the problem of constraints from the subject of inquiry to the researcher, we nonetheless believe that questions about adaptation provide a useful heuristic.

Inference

An instructive example of this approach is Anderson's (1990) rational theory of categorization. As we have mentioned, Anderson argued for the general strategy of constructing models on the basis of what would be rational (or even optimal) given an analysis of an organism's goals. The purpose of categorization, according to Anderson, is to maximize the inference potential or predictability of information. The key information for inferences is category validity, the probability of some feature being present in a given category. For example, if one knows that some entity is a bird, one can predict that it has two legs, has wings, may sing, and is unlikely to be dangerous. Anderson's model provides a surprisingly good account of a number of categorization effects, and we take this success as evidence that the inference function of categories is indeed important.

If categories serve to maximize inference potential, people may be biased to produce certain kinds of categories over others. Consider the results from rule induction experiments in which people are asked to develop a rule for categorization based on preclassified examples. In this situation, there are many rules that would correctly partition the examples. Among these rules, however, Medin, Wattenmaker, and Michalski (1987) found that conjunctive rules appeared more than five times as often as disjunctive rules. Conjunctive rules reveal a concern for, or sensitivity to, category validity. For example, a rule of the form *A and B* allows one to infer from knowledge of category membership that both A and B will be present. In contrast, a rule of the form *A or B* does not allow any inference to be made with certainty. Because category validity is relevant to drawing inferences from category membership, the preponderance of conjunctive rules is consistent with the inference function of categories.

Communication

In contrast to Anderson (1990) and others who assume that inference is the sole function of categories, we believe that categories serve multiple functions (Matheus, Rendell, Medin, & Goldstone, 1989). For example, concepts play a role in communication. Almost all clustering algorithms, including Anderson's rational theory, perform computations on the properties of examples. The resulting categories tend to maximize either inference potential or some relation of within- and between-category similarity. These methods, however, ignore the form of the category descriptions. If categories serve to ease communication, people should prefer partitions that have straightforward descriptions.

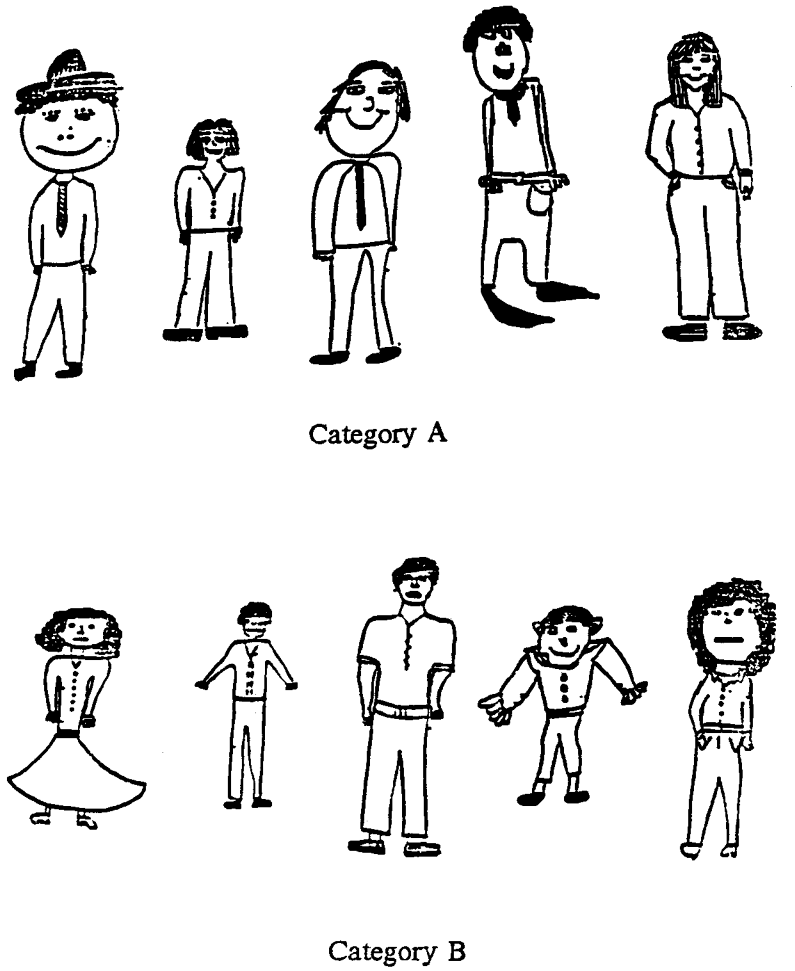
Unfortunately, some distributions of examples do not permit simple descriptions. In these situations, it seems likely that a bias toward easily described categories will lead people to represent their concepts by a simple rule plus a series of exceptions. Models that classify by predictiveness, on the other hand, will tend to develop categories based on overall similarity ("family resemblance"). On the basis of data from their rule induction task, Medin et al. (1987) suggested that subjects develop categories of the latter sort. They found that subjects constructed simple rules, and when these rules did not work perfectly, they "patched the rules up" instead of dropping them. In another experiment, Ahn and Medin (1989) asked subjects to sort examples that were structured so as to make one-rule descriptions impossible. They were able to predict sorting behavior in terms of a two-stage model in which the first stage corresponded to the development of a simple rule and the second stage consisted of strategies for dealing with examples that did not conform to the rule. In short, subjects in these two experiments preferred categories with simple descriptions over the family resemblance categories predicted by inference-based classification models.

Explanation

Categories also play a role in theories and in explanatory structures. Some researchers acknowledge the importance of the explanatory function of concepts but suggest that only the similarity-based aspect of categorization is tractable. This attitude entails a commitment to the view that similarity-based categorization is an isolatable component or module in categorization.

We believe that the explanatory function of categories needs to be integrated with similarity-based categorization (see Medin & Ortony, 1989, for one approach). The relation between similarity-based and explanation-based categorization has recently become a central focus in our research. In one line of work, we have used children's drawings as stimuli (see Figure 2) and introduced knowledge by varying the category labels (Wisniewski & Medin, 1991). For example, in a control condition the categories may be labeled *A versus B* and in the knowledge conditions (drawn by) *high IQ versus low IQ* (children). The task is to induce a rule that will successfully partition the categories.

If similarity can be treated as a separate module, one might account for the influence of knowledge by suggesting that induction occurs in two stages. In Stage 1, knowledge or explanations select and weight features of drawings. Stage 2 involves a similarity-based system that uses these weighted features to induce a rule. Our results, however, undermine the idea that similarity is an isolatable module. When the categories are labeled *A versus B*, people develop rules of the form that Michalski's (1983) similarity-based induction system would produce. For example, a typical rule for Figure 2 might be, "Category A drawings have buttons or stripes on their shirts and dark, thick hair." In the knowledge conditions, the rules were either more abstract (e.g., "The high IQ drawings are more relaxed and free flowing") or consisted of abstract generalizations linked to more specific, supporting predicates (e.g., "The high IQ drawings are more detailed, showing, for example, teeth, extensive shading, and drawing the body underneath the clothes").



Category A

Category B

FIGURE 2 Example of stimuli from Wisniewski and Medin (1991). (Subjects were to decide which group of pictures was drawn by children with high IQs and which was drawn by children with low IQs. Used by permission.)

To condense things quite a bit (see Wisniewski & Medin, 1991, for details), these results lead to two primary conclusions: (a) Knowledge-based rule induction involves developing links between abstract, explanatorily-relevant properties and more specific perceptual features, and (b) knowledge and similarity are tightly coupled and interact in a manner not captured by separate modules. In short, knowledge influences do more than simply select and weight perceptual features. These results, should they generalize, preclude the notion of a distinct similarity-based induction module. To better understand the processes of category learning, the explanatory function of categories must inform the models.

Summary and Conclusions

A true exemplar theorist following his or her model would be much more comfortable with case-based reasoning than with making abstract generalizations. Therefore, we will focus more on summary than on drawing conclusions. Nevertheless, we feel the need to insert a caution or two. The examples we have given are clearly not ideals. Many of our studies could be criticized on the standards that we have just outlined. In several cases, we may have learned something useful despite the fact that our stimuli and procedures may have worked to undermine success.

With this disclaimer in place, we conclude with a summary of the 12 pairwise connections between the four perspectives discussed. Although we do not cover every connection, many of the examples touch on one or more of them.

Beginning with the interaction between environmental sensitivity and function, we note that attention to the environment provides information about the problems that an organism needs to solve. Knowing the obstacles that an organism is likely to meet provides insight into what functions the organism is likely to have. Environmental sensitivity can also highlight the need for constraints. For example, if the language-learning problem is insolvable given the information available, there must be language-learning constraints. Finally, no formal model can adequately describe a process without representing the input to that process.

The interactions between environmental sensitivity and the other perspectives are not simply unidirectional. The environment is a vast place, and observations need to be driven by theory. This theory can be derived by concentrating on the functions that different processes might serve and the constraints that may be inferred from experimental data.

Constraints also affect, and are affected by, formal models. As we have mentioned, formal modeling can indicate which processes will be combinatorially explosive and hence need constraints. In turn, knowledge of what constraints exist in a given process must be included in any model of that process. For example, models of category learning should be as unconstrained by linear separability as are people.

Another interaction exists between constraints and functions. Medin et al.'s (1987) finding that people produce conjunctive rules more often than disjunctive rules provides both a constraint and a clue to the functions that rule induction might serve.

Finally, models interact with all three of the other perspectives. Function, environmental information, and constraints must be included in any model. The model can, in turn, determine whether the information given is sufficient. There may be too few constraints, too little information, or the model may not generalize to account for many functions.

Although we have only discussed pairwise interactions, three-way connections also exist. For example, we have seen how the difference between conjunctive rules and disjunctive rules implies function (predictiveness of concepts is important) and constraints (people produce more conjunctive than disjunctive rules) and bears on models (models that maximize category validity do a good job in predicting subjects' performance in certain tasks).

Abstraction theorists would probably tell us that we are trying to say something like the following: For at least the domain of classification learning and concept formation, a useful research strategy involves a mutually reinforcing interaction of sensitivity to ecological considerations, a constraints framework, questions about function, and formal and computational models. Ecological sensitivity guides and interprets experimental findings. Constraints address computational complexity. Function assists in understanding preferences or biases in underdetermined situations. Formal models help to avoid a host of mistakes and to construct mechanisms that could give rise to observed behavior. We think these four approaches to research, considered individually and in combination, provide an effective framework for thinking about and studying cognition.

References

Ahn, W. K., & Medin, D. L. (1989). *A two-stage categorization model of family resemblance sorting.* Paper presented at the 11th Annual Conference of the Cognitive Science Society, Ann Arbor, MI.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Bower, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika, 26,* 255–280.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93,* 183–321.

Estes, W. K. (1986a). Array models for category learning. *Cognitive Psychology, 18,* 500–549.

Estes, W. K. (1986b). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General, 115,* 155–175.

Flannagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 241–256.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distribution: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234–257.

Garcia, J., Ervin, F. R., & Koelling, R. A. (1966). Learning with prolonged delay of reinforcement. *Psychonomic Science, 5,* 121–122.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghton-Mifflin.

Gluck, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science, 2,* 50–55.

Gluck, M. A., & Bower, G. H. (1988). Evaluating an adaptive network model of human learning. *Journal of Memory and Language, 27,* 166–195.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97,* 3–18.

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 10,* 447–478.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437–447). New York: Bobbs-Merrill.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition, 8,* 378–382.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin, 90,* 197–217.

Homa, D., & Chambliss, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 351–359.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 322–330.

Kemler-Nelson, D. G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior, 23,* 734–759.

Krushscke, J. K. (1990). *A connectionist model of category learning.* Unpublished doctoral dissertation, University of California, Berkeley.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Matheus, C. J., Rendell, L. R., Medin, D. L., & Goldstone, R. L. (1989). *Purpose and conceptual functions: A framework for concept representation and learning in humans and machines.* Paper presented at the Seventh Conference of the Society for the Study of Artificial Intelligence and Simulation of Behavior. Sussex, England.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical diagnosis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 37–50.

Medin, D. L., & Goldstone, R. (1991). *Respects for similarity.* Manuscript submitted for publication.

Medin, D. L., & Ortony, A. (1989). Psychology essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge, England: Cambridge University Press.

Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem-solving, and induction. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189–223). Hillsdale, NJ: Erlbaum.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 355–368.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20,* 158–190.

Medin, D. L., Wattenmaker, W. D., & Michalski, R. S. (1987). Constraints and preferences in inductive learning: An experimental study of human and machine performance. *Cognitive Science, 11,* 299–339.

Michalski, R. S. (1983). A theory and methodology of inductive learning. *Artificial Intelligence, 20,* 111–161.

Minsky, M. L., & Papert, S. A. (1988). *Perceptrons.* Cambridge, MA: MIT Press.

Neisser, U., & Winograd, E. (Eds.). (1988). *Remembering reconsidered: Ecological and traditional approaches to the study of memory.* Cambridge, England: Cambridge University Press.

Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review, 81,* 267–285.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700–708.

Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of categorization. *Journal of Mathematical Psychology, 34,* 393–418.

Nosofsky, R. M. (in press). Exemplars, prototypes and similarity rules. In S. Kosslyn, R. Shiffrin, & A. Healy (Eds.), *Festschrift for William K. Estes.* Hillsdale, NJ: Earlbaum.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Erlbaum.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Shepard, R. H. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 19,* 417–447.

Smith, E. E. (1978). Theories of semantic memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 6, pp. 1–52). Hillsdale, NJ: Erlbaum.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Strange, W., Keeney, T., Kessel, F. S., & Jenkins, J. J. (1970). Abstraction over time of prototypes from distractions of random dot patterns: A replication. *Journal of Experimental Psychology, 83,* 508–510.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.

Wisniewski, E. J., & Medin, D. L. (1991). Harpoons and long sticks: The interaction of theory and similarity in rule induction. In D. Fisher & M. Pazzani (Eds.), *Computational approaches to concept formation* (pp. 237–278). San Mateo, CA: Morgan Kaufman.