In T. Tighe & B. Shepp (Eds.)
Development: Interactional Analyses.

7 Structural Principles in Categorization

Douglas L. Medin
University of Illinois

INTRODUCTION

What Holds a Category Together?

What makes a category a category, and why are some categories better than others? Most people would agree that robin, sparrow, and eagle form a sensible grouping whereas elephant, rocket, and shoestring do not. And it seems obvious why—the members of the first set are fairly similar to each other, but those of the second set are not. Yet this example is quite deceptive—canary, banana, and the moon do not comprise a "good" category even though they share an attribute in common (yellowness). Questions concerning principles of category structure are nontrivial, and many of the "obvious" answers to these issues may not hold up well under closer scrutiny.

One way of addressing the issue of constraints on categories is to ask what makes a category sensible and what function is served by categorization. In laboratory studies involving artificially constructed categories, the answer is easy. The purpose of categorization (from the subjects' point of view) is to produce a correct response and a category is whatever the experimenter decides is going to be a category. The latter decision is usually based on the theories under consideration.

In real-world situations, categorization may serve multiple functions (as numerous people have suggested, among them Bruner, Gibson, and Rosch). First of all, categorization allows one to relate new experiences to old. We don't perceive, remember, and talk about each object and event as unique but rather as

an instance of a class or concept that we already know something about. When we find ourselves in a new situation, we are confronted not with an array of unique entities but rather objects that are members of classes like chairs, desks, and telephones. Once we have assigned an entity to a category on the basis of its perceptible attributes, we can infer some of its nonperceptible attributes. For example, having used perceptible properties like color, size, and shape to decide an object is an apple, we can infer that it is edible and that it has a core containing seeds. In short, a basic cognitive task is a segmentation of the environment by which nonidentical stimuli are treated as equivalent.

Because experiences might be partitioned in a limitless variety of ways, we are again led to ask what makes a good or useful category. Given that natural concepts evolve out of human experience, answers about the structure of categories may contain hints concerning fundamental cognitive processes operating on those categories. It is hard to credit the possibility that human culture has passed on categories having a structure not at all coordinated with constraints of human information processing.

Organization of this Paper

Other than intuition, what may be used to educe principles of category structure? Can we predict how people will organize categories naturally, or, given some organization, can we say what structural principles went into it? The present chapter draws on two sources of observations: categorization theories and (seemingly) atheoretical maxims. To varying degrees, categorization theories imply constraints on category structures, constraints that have rarely been evaluated experimentally. The maxims to be considered are aphorisms such as the idea that categories are well-structured to the extent that they maximize cue validity. These statements bear closer scrutiny in their own right, and there are some interesting links between them and particular classification theories.

This chapter is organized into four main sections. In the first section, I consider several maxims or guidelines for category structure. My aim is to flesh out these ideas and to point out some of their shortcomings and strengths. Then I examine relationships between category constraints and particular classification theories. Thirdly, I intend to combine observations from both the theories and the maxims to argue for one particular view of category structure. I conclude with some reservations and speculations concerning developmental changes in classification learning.

It should be noted from the outset that this discussion should be amply sprinkled with hedges. Doubtless the functions of categorization differ from person to person and from situation to situation. In some contexts, a large category such as *vegetables* may be efficient for some purpose. In other settings, much more refined categories such as *poisonous mushrooms* may act as fundamental. Although similarity will be treated as a fixed entity, it must be acknowledged that selective attention can serve to either increase or decrease the effective similarity between any two stimuli. Having hinted at these hedges, in what follows we largely ignore them, mainly because they seem to be refinements that await a better overall picture.

MAXIMS FOR CATEGORY STRUCTURE

The idea of this section is not so much to review research relevant to these maxims as it is to examine just what they imply about category structure. I argue that many of these implications are implausible and, therefore, that these maxims are unlikely to provide useful constraints. Nonetheless, many of these implications are testable and it is always possible that some cogent experimental evidence will provide support for one or more of these maxims, despite my reservations.

Maximizing Cue Validity

One has only to look at concepts reflected in our language such as poodle, dog, pet, mammal, and animal to verify that people use multiple, overlapping categories that may differ in their level of abstraction. Rosch and her associates (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976) found that one level of abstraction, which they call the basic level, is more fundamental than either the associated superordinate or subordinate level. For example, by their criteria, *chair* would be a basic level concept, but *furniture* and *rocking chair* would not be. Their claims are reinforced by a variety of empirical results (see Mervis & Rosch, 1981, for an up-to-date review).

It has been argued that the basic level is the level at which cue validity is maximized. Although there is an entire set of issues associated with determining the attributes or features of concepts that would be needed to compute cue validity, it would be convenient if cue validity could be used to determine the basic level. But, as we soon see, it cannot.

Cue validity of some feature i can be defined as the probability that an entity belongs to category j, given that feature i is present; that is,

$$\text{cue validity} = \frac{P(\text{category } j \text{ and feature } i)}{P(\text{category } j \text{ and feature } i) + \sum_{k \neq j} P(\text{category } k \text{ and feature } i)} \quad (1)$$

The denominator is equal to the probability of the feature.

Now consider the following levels of concepts: sparrow, bird, warm-blooded, and animal. It can be shown that cue validity will increase monotonically with the level of abstraction. For example, the cue validity for the attribute *egg laying*

associated with sparrow will be fairly low because other birds, fish, reptiles, insects, and some mammals also lay eggs. Cue validity will be increased with respect to the category bird, because other birds will be included in the numerator of Equation 1. Moving the level to warm-blooded will further increase cue validity (anteaters would be added) and cue validity would be highest for the category animal (which includes *all* egg layers).

The core of the matter is this: Cue validity at worst will stay the same and can never decrease as one moves to higher levels of abstraction. Consequently, cue validity cannot be used to determine basic level nor can it be used as a measure of category goodness, unless it should turn out that the best and most useful categories are the most abstract categories. This is counter to both common sense and a large body of experimental literature. This is not to say that cue validity is not important in categorization, only that it is unlikely to stand by itself as the basis for category structure.[1]

Maximizing Differentiation

Another approach to categorization is to argue that it is analogous to discrimination learning and that categories are most sensible and easiest to learn when categories are most differentiated. Maximizing differentiation is equivalent to minimizing average between category similarity. Although this idea initially seems plausible, it almost surely will not work. Minimizing average between category similarity will not work because one is always led by this formulation to prefer two contrasting categories to any other number.

That minimizing average between categories implies that subjects will always sort stimuli into exactly two categories is not intuitively obvious, but it is fairly easy to demonstrate. For example, suppose a subject is given a choice between partitioning a set of stimuli into three categories, A, B, and C, or lumping A and B of the categories together. Let $\bar{S}_{ij}$ be the average between category similarity of category i and category j. Then we can always change the labels on the categories such that A and B have greater between-category similarity than do either A and C or B and C ($\bar{S}_{ab} > \bar{S}_{ac}$, and $\bar{S}_{ab} > \bar{S}_{bc}$). According to our criterion we should lump A and B if

$$(\bar{S}_{ac} + \bar{S}_{bc})/2 < (\bar{S}_{ab} + \bar{S}_{bc} + \bar{S}_{ac})/3.$$

On can quickly determine that this will be true whenever $\bar{S}_{ac} + \bar{S}_{bc} < 2\bar{S}_{ab}$, which is equivalent to our initial assumption. This implies that subjects would always put two of the categories together (A and B) and never partition the stimuli into three distinct categories. It is extremely doubtful that this generally

[1]Murphy (1982) has independently developed this and related arguments against cue validity as the sole basis of category organization.

would be true, and, therefore, maximizing differentiation cannot be the sole principle of category organization.

Maximizing Inferences from Category Membership

This guideline is consistent with the idea that categorization allows us to go beyond the information given to draw inferences (e.g., this object, an apple, has a core). With total number of features held constant, maximizing inferences is equivalent to maximizing the probability of a feature given a category, in some sense the converse of cue validity.

There are at least two problems with the criterion of maximizing the probability a feature is present given the category. The most glaring difficulty is that it implies that the most specific categories will provide the best categories. For example, singing, building nests in trees, and eating worms can be predicted more accurately from the knowledge that the animal in question is a robin than that the animal in question is a bird. Maximizing inferences is equivalent to maximizing within category similarity and the latter is maximized for the most specific categories. And the problem only becomes worse if number of features is taken into account, because more specific categories will have all the features of more general categories plus their own characteristic features. In its extreme form, the idea of having the most specific categories has the problem that they will rarely, if ever, be used. If every patient a doctor saw was totally unique, then the doctor's medical knowledge could never be brought to bear. At a minimum there must be some tradeoff between the frequency that a category can be applied and maximizing the probability that a feature is present given the category.

A second question that can be raised is why one should focus on inferring features from category membership to the exclusion of inferring features from other features. The features used to determine category membership could also be used to infer other features directly, without going through the two-step process of inferring membership and then drawing further inferences from membership. I would not want to go so far as to say that categorization serves no function in drawing inferences, only that, in some formal sense, inferences from features to features would serve as well as inferences from features to membership to other features.

Moreover, inferences from features to features might not have the same category structure as inferences from category membership to features and might permit a somewhat different category organization. Consider a situation where stimuli are comprised of binary-valued attributes from four stimulus dimensions. For example, color might be one dimension and we might use the value 1 to denote red and the value 0 to denote blue. Suppose we have two potential categories, X and X', each having four members as shown in Fig. 7.1.

In both cases the value 1 is typical for each of the dimensions and the probability of a feature given the category is the same for X and X'. Note,

	CATEGORY X DIMENSIONS					CATEGORY X' DIMENSIONS			
EXEMPLARS	A	B	C	D	EXEMPLARS	A	B	C	D
1	1	1	1	0	1	1	1	1	1
2	1	1	0	1	2	1	1	0	1
3	1	0	1	1	3	1	1	1	0
4	0	1	1	1	4	0	0	1	1

FIG. 7.1. Two hypothetical categories having four members, each comprised of values on each of four dimensions.

however, that dimensions A and B for category X' have correlated attributes such that, knowing the value on one dimension, one could be certain of the value on the other dimension.

Considering only the criterion of maximizing inferences from category membership, there is no basis for claiming that one of these categories is better organized than the other. But there may be important structural differences between the two categories that are being ignored. Rosch and Mervis have argued that real-world categories are organized to take advantage of correlated attribute clusters (Mervis & Rosch, 1981; Rosch, 1975, 1978; Rosch & Mervis, 1975), a claim that is consistent with the idea that relations between features play an important role in categorization processes.

Maximizing Within Category Similarity and Minimizing Between Category Similarity

A plausible conjecture at this point is that neither within-category similarity nor between-category similarity by itself determines category membership, but some joint function does (Mervis & Rosch, 1981). Although it is intuitively appealing, there are good reasons to be cautious in endorsing this view.

First of all, one should remember that in general one cannot simultaneously maximize within-category similarity and minimize between-category similarity. Maximizing within-category similarity calls for the most specific categories and minimizing between-category similarity calls for the most general categories (this point is not original with me—see, e.g., Tversky, 1977). It seems then that one can only maximize some function of within- and between-category similarity.

To my knowledge the only attempts to be specific about this function have used either a difference or a ratio of average similarity (Homa, Rhoads, & Chambliss, 1979; Rosch & Mervis, 1975). The conjecture to be evaluated is that categories are good to the extent that they maximize within-category similarity relative to between-category similarity.

Because maximizing within-category similarity and minimizing between category similarity is the most popular aphorism for category structure, it is considered in some detail. First, we take up the case of a preexisting category to which

one might add members or set up a new category and then we turn to situations where contrasting categories are already present.

Single-Category Case

To Add, Not to Add, or to Delete. Suppose that we are given a category comprised of n members and then are presented with one or more candidates and asked to either accept or reject them as members of a category. According to the principle of maximizing within-category similarity, we should accept a candidate if it increases the average within-category similarity.

If $n = 1$, this algorithm will never get off the ground because any new member will decrease average within-category similarity. If $n = 2$, then a candidate could be accepted but a new issue would be raised, as pointed out to me by a University of Illinois graduate student, Gerald Dewey. The new issue is this: If one can add members to categories when they increase within-category similarity, it seems that one also might delete members from a category when that would increase within-category similarity. Were deletions allowed, the category would remain at size 2 because there would always be one of three members whose deletion would lead to increased within-category similarity (except in the trivial case where all members were equally similar to each other).

Lumping Versus Setting up a New Category. Consider a modification of the task such that we start with a category of n members having an average similarity $\bar{S}_W$ and present two new stimuli having similarity to each other of $\bar{S}_H$ and similarity $\bar{S}_B$ to the preexisting category. The instructions are either to create a new category with these two new stimuli or to lump them into the old category. The potential situations are summarized in Table 7.1 where the different possibilities for within- and between-category similarity are laid out.

If one were maximizing the ratio of within- to between-category similarity, one would never set up a new category, because the ratio is always highest when the between-category similarity is zero. Using the difference in within- and between-category similarity, one would set a second category whenever

$$\frac{\binom{n}{2}\bar{S}_W + \bar{S}_H}{\binom{n}{2} + 1} - \bar{S}_B > \frac{\binom{n}{2}\bar{S}_W + 2n\bar{S}_B}{\binom{n}{2} + 2n} - 0.$$

Some algebraic readjustments lead to the only slightly less complicated inequality, implying that one would set up a new category when

$$2\,[\bar{S}_W(2n^2 - 3n + 1) + \bar{S}_H)n + 3)] > \bar{S}_B(n^3 + 6n^2 - 5n + 14).$$

What does this inequality tell us? It suggests a problem. If n is very large, one would almost never set up a new category even if $\bar{S}_H$ were much larger than $\bar{S}_W$ and $\bar{S}_B$ were quite small. To see this, assume for the moment that similarity can

TABLE 7.1
Patterns of Similarity When either a New Category Is Developed or a Single Category Remains[a]

	Within-Category Similarity	*Between-Category Similarity*
Single category (size $n + 2$)	$\frac{\binom{q}{2}\bar{S}_W + 2n\bar{S}_B}{\binom{q}{2} + 2n}$	0
Two categories	$\frac{\binom{q}{2}\bar{S}_W + \bar{S}_H}{\binom{q}{2} + 1}$	$\bar{S}_B$

[a]The notation, $\binom{\ }{\ }$, refers to combinations that is, $\binom{q}{2}$ is n items taken two at a time.

range between 0 and 1 with 0 representing no similarity and 1 representing maximal similarity. Then if $\bar{S}_W = .80$, $\bar{S}_H = .99$, and n were 20, $\bar{S}_B$ would have to be approximately .10 or less before a new category would be set up. For what it's worth, my intuition is that, to the contrary, new categories should be more likely to be set up when n is large (and one has some idea of within-category variability) than when n is small.

Contrasting Categories

Adding Members. Assume that two categories each of size n have been set up and that their average between-category similarity is $\bar{S}_B$ and in both cases the average within-category similarity is $\bar{S}_W$. A stimulus X is presented that has similarity $\bar{S}_1$ to the first category and similarity $\bar{S}_2$ to the alternative category. Suppose further that the stimulus is more similar to the first category than the second ($\bar{S}_1 > \bar{S}_2$) and that the subjects' task is to either assign X to category 1 or put it in neither category (reject it).

According to the rule under consideration, X will be placed into category 1 whenever

$$\frac{\binom{n}{2}\bar{S}_W + \binom{n}{2}\bar{S}_W + n\bar{S}_1}{\binom{n}{2} + \binom{n}{2} + n} - \frac{n^2\bar{S}_B + n\bar{S}_2}{n^2 + n} > \bar{S}_W - \bar{S}_B.$$

This inequality can be reduced to

$$\bar{S}_1 + n(\bar{S}_1 - \bar{S}_2) > \bar{S}_W + n(\bar{S}_W - \bar{S}_B).$$

This inequality is comprehensible in that it implies roughly that X will be accepted into category 1 if the difference between its similarity to category 1 and category 2 is as large as the difference between the within-category and the between-category similarity. Even so, there are some concerns that can be raised. If $\bar{S}_1 - \bar{S}_2 = \bar{S}_W - \bar{S}_B$, then X will be accepted if $\bar{S}_1 > \bar{S}_W$ and rejected if $\bar{S}_1 < \bar{S}_W$. One might have thought that the decision should be based only on $\bar{S}_1 - \bar{S}_2$ versus $\bar{S}_W - \bar{S}_B$. Related to this is the observation that if $\bar{S}_1 > \bar{S}_W$ but $\bar{S}_1 - \bar{S}_2 < \bar{S}_W - \bar{S}_B$, then whether or not X is accepted will depend on category size (n). When n is small, X will be accepted; but, for larger values of n, X will be rejected. Conversely, if $\bar{S}_1 < \bar{S}_W$ but $\bar{S}_1 - \bar{S}_2 > \bar{S}_W - \bar{S}_B$, then X will be rejected for small n and accepted for large n. This may be a minor problem but I can see no rationale for why the judgment should depend on category size in the manner predicted.

Forced-Choice Categorization. If we force subjects to put a new item into one of two preexisting categories, then maximizing within-category similarity relative to between-category similarity boils down to putting the item into the category to which it has the greater average similarity. I mention this primarily because a large set of categorization models, such as prototype models, would have precisely this expectation. But even here there may be a problem. Suppose that an item falls exactly in between two categories but that one of the two categories has much higher within-category variability (smaller average within-category similarity) than the other. Intuition suggests that the new item would be sorted into the high variability category; the choice rule under discussion predicts no preferences. Indeed, if one were to apply an analysis in terms of likelihood ratios, an item closer to the low variance category actually might be much more likely to have been drawn from the high variance category. (See Fried, 1979, for related evidence.)

Lumping Versus Setting up a Third Category. Again suppose we have two categories of size n with average within-category similarity $\bar{S}_W$ and average between-category similarity $\bar{S}_B$. Two new stimuli are presented with similarity $\bar{S}_H$ to each other, similarity $\bar{S}_W$ to category 1, and similarity $\bar{S}_B$ to category 2 (we use two prospective members to avoid the issue of whether or not a category can have only one member). The task is either to put the new stimuli in category 1 or to set up a new category.

The algebra associated with the given situation is tedious and largely unilluminating. Suffice to say that for some cases lumping is predicted to increase with n, and for others lumping is predicted to decrease with n. There is no clear rationale for why this should be so.

Summary

I have been trying to make two points concerning the stratagem of maximizing within-category similarity and minimizing between-category similarity. One is that, because it is based on *average* similarity, it is far from theoretically neutral. Rather it is a natural extension of what Medin and Schaffer (1978) refer to as independent cue models. Specifically, these models assume that categorization judgments are based on an additive, independent summation of component information. Data inconsistent with independent cue models (Medin & Smith, 1981) indirectly undermine this maxim.

The other point is that when examined in detail, this maxim has certain implications that are far from obvious. It is conceivable that these counter-intuitive predictions are correct. Without empirical support, however, their intuitive implausibility suggests that the criterion itself is incorrect. The idea that categories are cohesive to the extent that they maximize within-category similarity and minimize between-category similarity, by itself, does not seem to constitute or resolve the issue of structural principles in categorization.[2] We turn now to specific categorization theories to see what implications they have for category structure.

CLASSIFICATION THEORIES AND CATEGORY STRUCTURE

The Classical View and Defining Features

One view of category structure is that natural concepts are characterized by simple sets of defining features that are singly necessary and jointly sufficient to determine category membership (Katz & Postal, 1964). A candidate exemplar either does or does not possess these defining features and thereby either is or is not a member of the category. The idea that categories are comprised of these defining features is considered in detail by Smith and Medin (1981), who dubbed this idea the "classical view."

The classical view has a lot to say about category structure. Defining features are what make a category a category and the difficulty of learning a category will depend on the difficulty in discovering these defining features and the extent to which the defining features of contrasting categories overlap with the category in question.

Despite its precise statements concerning category structure, the classical view will not be given further consideration. The major problem is that the classical view may be appropriate for only a small set of categories. The scholarly consensus (see Mervis & Rosch, 1981; Rosch & Lloyd, 1978; Smith & Medin, 1981, for reviews) has it that most natural concepts are not well-defined but rather are based on relationships that are only on the average true. Features are said to be characteristic rather than defining. Members of a category may vary in the number of characteristic features they possess and correspondingly vary in the degree to which they are judged to be good examples (typical) of a

[2]Larry Barsalou (personal communication, 1981) pointed out that this discussion proceeds as if a given exemplar can only belong to one category. In fact, however, items may belong to a variety of categories, and which category is employed may depend on an organism's plans and goals. This underlines the difficulty of using any single factor such as similarity as the sole determinant of category structure. At a minimum one would need a theory for how plans and goals modify similarity.

category (Mervis & Rosch, 1981; Rosch, 1973; Rosch & Mervis, 1975; Smith, Shoben, & Rips, 1974). For example, cows are rated to be better exemplars of the concept *mammal* than are whales (McCloskey & Glucksberg, 1978). In this view, instances are neither arbitrarily associated with categories nor strictly linked by defining features but rather reflect more nearly a "family resemblance" structure (Rosch & Mervis, 1975).

The Probabilistic View and Linear Separability

The Probabilistic View

If many natural categories do not have defining features, how do people acquire and use them? Posner and Keele (1968) proposed that people form an impression of the central tendency of a category as a result of experience with exemplars and that categorical judgments come to be based on this central tendency, or prototype. An eagle, for example, would be judged to be a bird and not a mammal because it is more similar to the bird prototype than to the mammal prototype.

What does prototype theory imply about the structure of categories? The main constraint is that categorizing on the basis of similarity to a prototype has to work in the sense that all members will be accepted and all nonmembers will be rejected. If, by some quirk of fate, sparrow were mammals rather than birds, a prototype process would not work, because sparrows have many characteristic features of birds and few characteristic features of mammals.

The idea that category representations are based on characteristic attributes and that classification decisions are based on how closely an examplar matches the summary representation of a category is known as the *probabilistic view*. Prototype theory is one instance of a class of models conforming to the probabilistic view. Because what is said about prototype theory applies as well to the other models, we use prototype theory as a shorthand way of referring to the probabilistic view.

One way of thinking about classifying stimuli on the basis of similarity to prototypes is that it involves a summing of evidence (e.g., characteristic features) against some criterion. The more typical the category member, the more quickly the summing of evidence should exceed the criterion. Therefore, it would not be surprising to find that people categorize robins as birds more rapidly than they categorize penguins as birds (see Mervis & Rosch, 1981, for a review). The key constraint is that this summing of evidence (or, alternatively, similarity to prototype) accepts members and excludes nonmembers of a category.

The formal term for the constraint just given is that categories be *linearly* separable or separable by a linear discriminant function (Sebestyen, 1962). Linearly separable categories are categories that can be partitioned on the basis of a weighted, additive combination of component information. If two categories are

linearly separable, then their members could be classified correctly on the basis of similarity to the respective prototypes, that is, every member of a category will be more similar to the prototype for its category than to the prototype for any contrasting categories.

Prototype, average distance, and versions of cue validity and frequency models fall under this domain (see Medin & Schaffer, 1978, for a more complete description of these models); all imply that categories should be linearly separable. Linear separability is also of interest because it has received important consideration in the closely allied area of pattern recognition and classification by machines (Nilsson, 1965).

Because linear separability is such an important constraint in formal models of classification, one might expect considerable interest in determining whether linear separability is a basic constraint in human information processing. Surprisingly little attention has been paid to this issue.

One way of evaluating the importance of linear separability to human categorization is to set up two categorization tasks similar in major respects, except that in one task the categories would be linearly separable and in the other categorization task they would not. A clear implication of independent cue models is that the task involving linearly separable categories should be easier to master than the task not conforming to this constraint; that is, the idea is to see if it is important that categories be separable by an additive combination of component information. Paula Schwanenflugel and I have recently completed a series of studies examining linear separability (Medin & Schwanenflugel, 1981), which I shall briefly summarize.

Is Linear Separability an Important Structural Principle?

The design of one experiment is shown in Fig. 7.2. In this example, stimuli are comprised of values on four component dimensions, described in terms of a binary notation. The number 1 represents the typical or characteristic value for members of category *A* and the value 0 is typical for category *B* members. No defining features exist to partition categories *A* and *B*, but the categories in the top half of the figure are linearly separable. Each category *A* member has three values typical of category *A* and no *B* member has more than two values typical of category *A*. Although the overall average similarity of the two categories in the bottom of Fig. 7.2 is the same as for the top of Fig. 7.2, the stimuli in the bottom half do not constitute linearly separable categories. Exemplar B_1 has more values typical of category *A* than either exemplar A_1 or A_2. If linear separability is important in classification, then the classification problem in the top half of Fig. 7.2 should be easier than that shown in the bottom half.

Not all categorization models share this prediction. According to the context model of Medin and Schaffer (1978), similarity of exemplars to each other is the major factor controlling classification difficulty. Specifically, high similarity of

LINEARLY SEPARABLE CATEGORIES

CATEGORY A					CATEGORY B				
	DIMENSION					DIMENSION			
EXEMPLAR	D_1	D_2	D_3	D_4	EXEMPLAR	D_1	D_2	D_3	D_4
A_1	1	1	1	0	B_1	1	0	1	0
A_2	1	0	1	1	B_2	0	1	1	0
A_3	1	1	0	1	B_3	0	0	0	1
A_4	0	1	1	1	B_4	1	1	0	0

CATEGORIES NOT LINEARLY SEPARABLE

CATEGORY A					CATEGORY B				
	DIMENSION					DIMENSION			
EXEMPLAR	D_1	D_2	D_3	D_4	EXEMPLAR	D_1	D_2	D_3	D_4
A_1	1	0	0	0	B_1	0	0	0	1
A_2	1	0	1	0	B_2	0	1	0	0
A_3	1	1	1	1	B_3	1	0	1	1
A_4	0	1	1	1	B_4	0	0	0	0

FIG. 7.2 Abstract representation of the alternative categorization tasks. Each task involved eight stimuli varying along four dimensions.

exemplars within categories facilitates categorization and high similarity of exemplars between categories impairs performance.

The context model is just one of a number of alternative classification models that do not imply that linear separability is an important constraint. These alternative models, which we refer to collectively as *relational coding models*, have in common the assumption that classification involves in one way or another combinations of attributes or features (Anderson, Kline, & Beasley, 1979, Hayes-Roth & Hayes-Roth, 1977; Neumann, 1974; Reitman & Bower, 1973). Because this use of combinations implies something other than an independent summation of component information, relational coding models do not require that categories be linearly separable. Although the models vary in their assumed underlying processes, they share the prediction of the context model that high similarity of exemplars within categories facilitates performance and high similarity of exemplars between categories impairs it.

If instances of high similarity are an important variable in classification, the task where the categories are not linearly separable (NLS) might be easier than the linearly separable (LS) task. According to this principle, instances of high similarity within a category should facilitate learning and high similarity across categories should impair it. We define the stimuli as highly similar if they differ in value along only one dimension. Inspection of Fig. 7.2 reveals no cases of high within-category similarity and six cases of high between-category similarity (A_1 and B_1, A_1 and B_2, A_1 and B_4, A_2 and B_1, A_3 and B_4, A_4 and B_2) for the LS task. In contrast, for the NLS task there are four cases of high similarity within

categories (A_1 and A_2, A_3 and A_4, B_1 and B_4, B_2 and B_4) and only three cases of high between-category similarity (A_1 and B_1, A_2 and B_3, A_3 and B_3). Therefore, on the basis of similarity relationships, the NLS categorization task is predicted to be easier to master than the LS. In other words, with overall similarity held constant, the experiment pits high similarity of exemplars against linear separability.

The basic procedure was straightforward. The stimuli were geometric shapes differing in color, form, size, and number. Stimuli were presented one at a time in a random order to participants. They assigned it to either category *A* or category *B* and then were told whether they were correct or incorrect. Subjects were told that their task was to learn to classify the stimuli correctly, that categorization could not be based on any one feature alone, and that the task was difficult but eventually they should be able to be correct all the time. Thirty-two subjects were given the linearly separable categorization task and an equal number were given the other task. Training continued until a participant met the learning criterion or until each exemplar had been presented 16 times.

Type of instruction was an additional variable. Subjects given *inference* instructions were told to think of the task as being analogous to learning two artistic styles. They were asked to focus on abstracting out general characteristics of each category (style) so that they could correctly classify new examples that would be presented later. They were told that the specific stimuli to be seen during learning were important only as aides to abstracting out general style.

Subjects given *facts* instructions were also told to think of the stimuli as being analogous to artistic styles, but they were also reminded that Picasso and other artists often change their style. They were further informed that the stimuli in their task might not be representative of the category and that they should perform the task by focusing on individual stimuli rather than be attempting to abstract out the general style.

Altogether the combinations of classification tasks and subjects created four distinct groups; the instructional manipulation was designed to see if the relative difficulty of the LS and NLS tasks would interact with the type of strategy used by subject. The *inferences* instructions were designed to maximize higher level abstraction and the *facts* instructions aimed to minimize abstraction.

The results were clear-cut. For both types of instructions the NLS task proved to be easier than the LS task, contrary to prototype theory but consistent with theories that emphasize the similarity of exemplars to other exemplars. The *facts* instructions were associated with fewer errors than the *inferences* instructions.

One can argue that the classification task in the previous experiment is not representative of categorization in general. There are three ways in which the geometric stimuli may not have been optimal for studying classification performance. First it might be argued that the stimuli used in the previous study were not as complex as those found in natural categories, that is, the geometric forms that were used differed solely on four dimensions, whereas instances of natural

categories differ on many dimensions. The fact that instances of natural categories differ on many dimensions means that a person is required to ignore certain dimensions of an instance and attend to others when making a categorization decision.

Second, it might be argued that natural categories usually consist of many instances, whereas in the previous study only a few instances of each category were used. The small stimulus set used may have permitted the subjects to use the equivalent of a paired-associate learning strategy.

Finally, the attributes of the geometric stimuli in the study just described were binary valued and a given value (e.g., red) was identical wherever it appeared. For natural categories, attributes such as *feathered* are themselves an abstraction and the exact realization of an attribute varies from instance to instance (e.g., feathers of a peacock versus feathers of a crow).

The stimuli for the next experiment were selected to address all three of these possibilities. The categorization stimuli consisted of a potentially infinite set of photographs of faces. These faces varied systematically along four dimensions (hair color, hair length, smile type, and shirt color), while varying freely on all other dimensions. Although attributes of the faces were encoded as binary values (e.g., light hair versus dark hair), the realization of some value (light hair) could vary considerably from face to face. If the factors of stimulus set size, presence of irrelevant features, and variability within a given attribute value, either alone or in conjunction, elicit or constrain the process of summing evidence from component dimensions, then linear separability should now become important. The actual stimuli used, taken from old college yearbooks, were faces of women that varied in hair color, shirt color, smile type, and hair length, as well as in other numerous, irrelevant, attributes. The structure of the linearly separable and nonlinearly separable categories is shown in Fig. 7.3. There are several aspects of note for these structures: First, both the linearly separable and nonlinearly separable categories show the same overall similarity (that is, the frequency of the values on the component dimensions is the same for the LS as for the NLS categories). Second, the two category types differ with respect to between-category similarity. Although *average* between-category similarity is equated, the LS categories show more instances of high between-category similarity of types (four) than the nonlinearly separable categories (no instances of high between-category similarity). Therefore, relational coding models predict that the linearly separable categories will be more difficult to learn than the nonlinearly separable categories.

Although both the LS and the NLS classification task proved to be fairly difficult, performance improved steadily with practice, as seen in Fig. 7.4. Figure 7.4 also shows an advantage for the NLS condition, which developed quickly and did not change with practice. The error data also do not support the idea that linear separability is important in classification. Subjects in the LS task averaged 34.0 errors, whereas subjects in the NLS task averaged 28.8 errors, a

LINEARLY SEPARABLE CATEGORIES

CATEGORY A TYPE	DIMENSION D_1	D_2	D_3	D_4	CATEGORY B TYPE	DIMENSION D_1	D_2	D_3	D_4
A_1	0	1	1	1	B_1	1	0	0	0
A_2	1	1	1	0	B_2	0	0	0	1
A_3	1	0	0	1	B_3	0	1	1	0

CATEGORIES NOT LINEARLY SEPARABLE

CATEGORY A TYPE	DIMENSION D_1	D_2	D_3	D_4	CATEGORY B TYPE	DIMENSION D_1	D_2	D_3	D_4
A_1	1	1	0	0	B_1	0	0	0	0
A_2	0	0	1	1	B_2	0	1	0	1
A_3	1	1	1	1	B_3	1	0	1	0

FIG. 7.3 Abstract representation of the alternative categorization tasks. Note that individual stimulus *types* rather than individual stimuli are represented. Each task involved six stimulus types and a potentially infinite set of individual stimuli representing these types.

difference in the direction predicted by relational coding models but short of statistical reliability ($t[62] = 1.35$, $p < .10$).

In a final attempt to find evidence for the importance of linear separability, the category strucutres were simplified. Specifically, the number of relevant dimensions was reduced from four to three. Independent cue models, such as prototype theory, assume that classification is equivalent to a summing of evidence and our

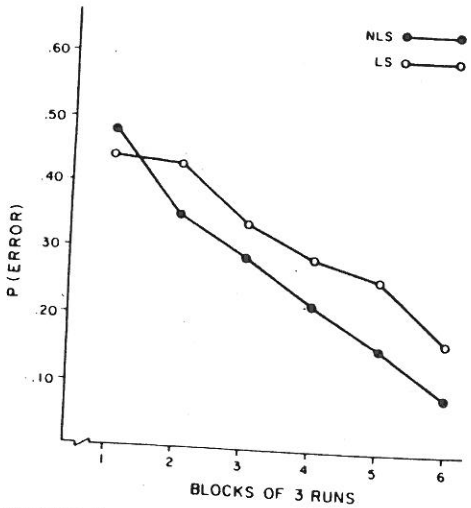

FIG. 7.4 Proportion of errors as a function of practice for the linearly separable (LS) and not linearly separable (NLS) tasks involving an unlimited set of stimuli.

intent was to examine the simplest case of summation. There were only three relevant dimensions, and in the linearly separable task subjects could correctly classify all stimuli by looking at the three attributes and using a two out of three rule: that is, if two of the three values were typical of category *A*, then that stimulus belonged to category A.

To accomplish this simplification, the second dimension from the structures used in the last experiment was eliminated (held constant). Because the LS task remains separable and the NLS task remains nonseparable, independent cue models predict that the LS task should be easier.

Although performance consistently improved with practice, less than half the subjects in either group met the learning criterion. Mean errors during learning were virtually identical—NLS subjects averaged 38 errors and LS subjects averaged 39.5 errors. The LS task was no easier to learn than the NLS task, contrary to the prediction of independent cue models.

Despite our varied attempts, we have been unable to find any evidence that linear separability is a factor in classification learning. To my knowledge, these studies represent the first direct comparison of learning of LS and NLS categories. Therefore, cautions concerning the generality of these results are even more pertinent than is usually the case.

One reaction to these experiments is that they are not a fair test of the importance of linear separability because the difficulty of the tasks (there were many nonlearners in the last two experiments) is clear evidence that the categories were unnatural. Weird categories produce weird results supporting weird theories. Although this claim has merit, it is not truly applicable to these experiments. The category structure for the LS task was dictated by theories that imply linear separability is important, and from the perspective of these theories it is hard to imagine an easier task than using a two out of three rule. If such a simple rule can be associated with such a difficult classification task, then perhaps linear separability is not the key to category structure.

The Exemplar View

The exemplar view accepts the idea that many concepts may not have defining features but differs from the probabilistic view in that it assumes that category judgments may be based on the retrieval of specific-item information rather than category-level information (e.g., by reference to a prototype).

Do exemplar-based theories provide meaningful constraints on category structure? The most optimistic answer is a hedge. To begin with, this class of models is somewhat diffuse. Smith and Medin's (1981) review lumped into this class all models that either do not rely on an abstract summary representation or are not based on an additive summation of component information. For example, the context model is neither tied to the assumption that a distinct representation is set up for each exemplar nor does it assume an additive summation rule. An average

distance model does assume that every exemplar is represented and does use an additive combination rule (and thus implies that linear separability is an important constraint). Both are considered exemplar models.

Exemplar models seem to imply that something belongs in a category if you have learned that it does. A lamprey is a fish because you have been told that lampreys are fish. But this is totally unconstrained and provides no guidelines at all concerning category structure.

At this point one may despair of any likelihood that theories will have anything useful to say concerning what makes a category cohesive. In the remainder of this chapter we pursue an alternative approach to these questions. Although many models do not say anything direct about category structure, they do imply that some categorization tasks should be easier than others. If these predictions are generally supported, then we may obtain some hints as to the structural principles that these theories are exploiting, by looking at the basis of these predictions.

CONTEXT THEORY AND CORRELATED ATTRIBUTES

Context Model

Although the context model often has been presented as an exemplar view model and contrasted with probabilistic view models, it can be argued that the key aspect of the model is not that it assumes exemplar storage but rather that it assumes components are treated in a nonindependent manner.

The context model assumes that when some stimulus or cue is presented in some experimental context and some event occurs (e.g., the classification assignment), information concerning the cue, the context, and the event are stored together in memory and that both cue and context must be activated simultaneously in order for information about the event to be retrieved. A change in either the cue or the context can impair the accessibility of information associated with both. It is further assumed that a particular stimulus component serves a cue function *and* acts as context for other cues. This means that components will combine in an interactive rather than independent manner.

This formulation is closely related to the assumptions of the Estes hierarchical association model (Estes, 1972, 1973, 1975). Estes employs the notion of a "control element" and assumes (1) that neither cue nor context is directly associated with an event or outcome; and (2) that inputs from both the cue and context are needed to activate the node and provide access to the representation of an event.

As applied to classification, the context model proposes that when an item is presented to be classified, that item acts as a retrieval cue to access information associated with similar stored exemplars. The various cue dimensions comprising stimuli in some context are assumed to be combined in an interactive, specifically multiplicative, manner to determine the similarity of two stimuli.

The multiplicative rule has the implication that an exemplar may be classified more efficiently if it is highly similar to one instance and dissimilar to a second than if it has medium similarity to two instances of a category. Hence the context model predicts that categorization performance will vary with the number of stored exemplars that are highly similar to the test item. Independent-cue models are insensitive to such density effects. In a series of four experiments, Medin and Schaffer (1978) obtained clear support for the context model. Data from original learning, transfer, and speeded classification were in each case more in line with the context model than with a generalized independent-cue model. In addition, a mathematical version of the context model gave an excellent quantitative account of classification performance on transfer tests involving new and old instances. A follow-up study varied the strategies subjects employed and although strategy variations produced large differences in performance, certain relationships were invariant over strategy, relationships that were captured by the context model and not by independent-cue models (Medin & Smith, 1981).

The fact that high similarity to one instance and low similarity to another instance will outweigh medium similarity to two instances in some sense makes the context theory a configural model. Correlated attributes represent one type of configural information. If people are sensitive to correlated attributes, then it would be very convenient if natural categories were organized around correlated attributes. In the next section, this rationale is slightly expanded and then the application of the context model to some recent experiments on correlated attributes are described.

Correlated Attributes

Rosch and her associates have recently argued persuasively that real-world categories are formed to take advantage of correlated attribute clusters (Mervis & Rosch, 1981; Rosch, 1975, 1978; Rosch & Mervis, 1975). The perceived world of objects is not a total set (in Garner's, 1974, sense) but rather is a subset. In other words, certain attributes tend to co-occur. For example, animals with feathers are very likely to have wings and beaks, whereas animals with fur are very unlikely to have wings and beaks. There are two very important advantages to being sensitive to correlated attribute clusters. First, this allows an organism to predict attributes of an object from knowledge of other attributes (as noted on page 207). Second, those categories that best follow the natural correlation of attributes are likely to be maximally differentiated from each other. When combined, these two ideas remind one of the principle of maximizing within category similarity and minimizing between category similarity. The analogy, however, is not quite correct. It is incorrect in the sense that there may be correlated attributes *within* a category that provide further internal structure. Correlated attributes

within a category could provide clusters of exemplars that might act as subcategories at a more refined level of analysis. There is evidence that natural categories do have correlated attributes within categories (Malt & Smith, 1981; Smith & Medin, 1981).

People also seem to use correlated attributes as a basis of classification when categories are ill-defined. In one set of studies designed to evaluate correlated attributes, we (Medin, Altom, Freko, & Edelson, 1982) used a simulated medical diagnosis task. Subjects learned about a fictitious disease, burlosis, from hypothetical case studies of patients having the disease. The case studies included descriptions of symptoms that tend to be characteristic of burlosis. Some symptoms were correlated with each other, whereas others were independent. After subjects studied the descriptions, they were presented with pairs of new cases and asked to judge which was more likely to have the disease, based on what they had learned from the earlier case studies.

During an initial training phase, subjects were presented with nine different case studies of the fictitious disease burlosis. The basic design of the learning cases is depicted in Fig. 7.5. For each subject, two symptoms were completely correlated in that a given patient either had both symptoms or neither symptom. For example, in Fig. 7.5, eye condition and weight condition were perfectly correlated symptoms. Each description involved five symptom dimensions: blood pressure (high or low), skin condition (splotches or rash), muscle condition (stiffness or loss of control), eye condition (puffy or sunken), and weight condition (loss or gain). One symptom of each pair was selected to be typical (notation 1 in Fig. 7.5) and the other was selected to be atypical (notation 0).

The transfer tests of primary interest are shown in Fig. 7.6. For one type of test, both new case descriptions broke the correlation between the symptom dimensions, eye condition and weight condition. On these tests virtually all

SYMPTOMS OF BURLOSIS

CASE STUDY	BLOOD PRESSURE	SKIN CONDITION	MUSCLE CONDITION	CONDITION OF EYES	WEIGHT CONDITION
1 RL	0	1	0	1	1
2 LF	1	1	0	1	1
3 JJ	0	0	1	1	1
4 RM	1	0	1	1	1
5 AM	1	1	1	1	1
6 JS	1	1	1	1	1
7 ST	1	0	0	0	0
8 SE	0	1	1	0	0
9 EM	1	1	1	0	0

FIG. 7.5 Case studies presented during learning according to the binary notation. The fourth- and fifth-symptom dimensions are perfectly correlated.

TEST TYPES

UNCORRELATED vs UNCORRELATED	UNCORRELATED vs CORRELATED
01110 vs 11101	11100 vs 11101
11001 vs 11110	01111 vs 11101
01010 vs 11010	01011 vs 11110
10001 vs 10101	10000 vs 10010

FIG. 7.6 Transfer tests evaluating the effects of number of typical symptoms and the presence or absence of correlated symptoms in the burlosis experiment.

classification models would expect people to choose the description having the greater number of typical symptoms to be more likely to have burlosis. The other type of test pitted number of typical symptoms against whether or not the symptom correlation was broken. For example, the pattern 10010 has two typical symptoms, whereas the pattern 10000 has only one, but the latter pattern preserves the correlated attribute structure. The context model predicts that people will select descriptions involving correlated attributes as more likely to have burlosis. Independent-cue models predict that number of typical symptoms will continue to govern choices.

The results were clear. For the tests where pattern of symptom correlations was broken in both descriptions, the description having the greater number of typical symptoms was judged more likely to be associated with burlosis. In the tests where the descriptions with fewer number of typical symptoms preserved the pattern of symptom correlations and the description with more typical symptoms did not, the pattern of choices was reserved; that is, the description with fewer typical symptoms but preserving the correlations was judged to be more likely to have burlosis, as predicted by configural-cue models such as the context model.

The task used in the previous experiment may seem strange in that only a single category was used. Although no subject complained about the oddness of the task, we ran an additional experiment using two contrasting categories.

The design of the follow-up experiment is shown in Fig. 7.7. Case studies involved one of two diseases, terrigitis or midosis, and as in the earlier studies the values along the last two symptom dimensions were perfectly correlated. For the first two symptom dimensions the value 1 was typical of terrigitis and the value 0 typical of midosis. Following original training, subjects were presented with new case studies and asked to classify them as having terrigitis or midosis. All 16 possible combinations of symptom patterns were presented on these transfer tests. Of these new patterns, 8 exactly matched one of the original case studies, but the other 8 did not. The context model predicts that classification is based on similarity to case studies but does not assume that only exact matches influence performance. The multiplicative similarity function also implies a sensitivity to correlated symptoms for both exact matches and nonexact matches that

TERRIGITIS	DIMENSIONS			
CASE	D_1	D_2	D_3	D_4
E.M.	1	1	1	1
S.T.	0	1	1	1
R.M.	1	1	0	0
L.F.	1	0	0	0

MIDOSIS	DIMENSIONS			
CASE	D_1	D_2	D_3	D_4
R.L.	1	0	1	0
A.M.	0	0	1	0
S.E.	0	1	0	1
J.J.	0	0	0	1

FIG. 7.7 Abstract representation of the contrasting categories for the terrigitis/midosis experiment. The third- and fourth-symptom dimensions are perfectly correlated.

preserve symptom correlations. The success of the context model's predictions can be assessed by how well it fits the transfer data quantitatively.

The main results are shown in Table 7.2. Subjects performed quite accurately on symptom patterns that exactly matched a case study from the training phase. There was also a systematic effect of the first two symptom dimensions—subjects were 12% to 16% more accurate on matching patterns if the first and second symptoms were typical of the disease than if one of the two symptoms was atypical.

TABLE 7.2
Transfer Tests and Response Proportions Associated with Fig. 7.7

Test Description		*Observed (Predicted) Proportion of Terrigitis Responses*
Exact match to old terrigitis patterns	1111	.88 (.85)
	0111	.73 (.76)
	1100	.89 (.86)
	1000	.77 (.73)
Exact match to old midosis patterns	1010	.25 (.24)
	0010	.12 (.15)
	0101	.33 (.27)
	0001	.17 (.14)
New patterns	0000	.53 (.56)
	0100	.75 (.73)
	0110	.36 (.34)
	1110	.45 (.46)
	0011	.53 (.54)
	0100	.67 (.66)
	1001	.28 (.27)
	1101	.38 (.44)

Predicted values derived from the context model are shown in parentheses.

Exactly the same pattern held for new transfer patterns. In general, new patterns were placed in the category associated with the symptom correlation that the new pattern manifested, but there were also clear effects of the first two symptom values. Thus it appears that, for both exact matches to old case studies and for new combinations of symptoms, subjects used correlated symptoms and typical symptoms in making their classifications. This is the pattern of performance predicted by the context model.

Quantitative Fits to Context Model. According to the context model, when new items are presented on a transfer test, they will act as retrieval cues to access stored information associated with similar case studies. Similarity along a symptom dimension is assumed to vary between 0 and 1, with 1 representing identity or maximum similarity. Overall similarity of a transfer item is assumed to be a multiplicative function of individual symptom similarities. If we represent the similarity on the four symptom dimensions in Fig. 7.7 by a, b, c, and d respectively, then transfer description 0110 would have an overall similarity to case study R.L. (1010) of $a \cdot b \cdot 1 \cdot 1$, or $a \cdot b$; that is, matching symptoms have a similarity of 1 whereas mismatches are represented by a corresponding similarity parameter. By the same system, the new item would have similarity b to case study A.M., $c \cdot d$ to case study S.E., and so on. Small values for similarity along a symptom dimension imply that a mismatch on that dimension would be salient.

The similarity parameters were estimated by a grid search designed to minimize the difference between predicted and observed transfer performance, according to a least-squares criterion. The probability of classifying a transfer description as terrigitis was assumed to be equal to the sum of similarities of the description to the terrigitis case studies divided by that sum plus the sum of the similarities of the description to the four midosis case studies. The resulting estimates for a, b, c, and d were .52, .40, .32, and .06, respectively, and the predictions associated with these parameters are shown in parenthesis in Table 7.2. The parameter estimates suggest that the correlated dimensions, c and d, tended to be more salient than the first two dimensions. The fit to the data shown in Table 7.2 is excellent—the average absolute deviation of predicted and observed proportions is less than .03, the root mean squared deviation is .03, and 98% of the variance is accounted for by the model.[3]

Overall, the idea that correlated attributes have a critical role in categorization appears to have considerable promise. Sensitivity to correlations, first, is consistent with the principle of being able to infer the presence of other attributes from known attributes. And sensitivity, moreover, may be an indirect property of memory retrieval processes (at least according to the context model).

[3]More recent research in our laboratory by William Wattenmaker shows that the correlation between symptoms need not be perfect for people to use them on transfer tests.

CONTEXT THEORY, STRATEGIES, AND DEVELOPMENT

Ill-Defined Concepts and Hypothesis Testing

Much of this chapter has been concerned with the structure of ill-defined categories, that is, categories that do not have defining features. Earlier work on concepts used well-defined categories and focused on such issues as the relative difficulty of acquiring different rules, strategies for formulating and testing alternative hypotheses, and the transfer of behavior to new stimulus sets (Levine, 1975; Trabasso & Bower, 1968). It may be worth pointing out that current hypothesis-testing models would have difficulty in predicting that ill-defined concepts would be learned at all, because there may be no "correct" hypothesis to describe category membership of fuzzy concepts like *furniture*. A child learning about birds might entertain the hypothesis that birds fly, only to discard this conjecture when encountering either a nonflying bird (e.g., an ostrich) or a flying nonbird (e.g., a bat). A hypothesis-testing process that discarded hypotheses whenever contradictory information was encountered would throw out important information about concepts (i.e., their characteristic features).

The further evidence that young children are less likely than older children to use strategies suggests a moral here, but it takes the form of a speculation. Instead of viewing the tendency of young children to be nonstrategic as a shortcoming, it may be more accurate to view nonstrategic behavior as precisely the means by which children are able to learn ill-defined concepts. Strategies may get in the way.

There is some evidence to support this speculation. Kossan (1981) tested second- and fifth-grade children on different types of category structures and encouraged children either to attack the problem conceptually or to be nonstrategic (i.e., to use a paired-associate learning strategy). Although fifth-grade children performed better than second-grade children overall, for the ill-defined category structures second-grade children given paired-associate instructions outperformed fifth-grade children given conceptual instructions (see also Farah & Kosslyn, 1981, for a more extensive discussion of children's learning of ill-defined concepts).

One reason why strategic behavior may not always be efficient is that the set of possible hypotheses is very large. Consider correlated attributes. Suppose that in some domain the entities could be characterized in terms of eight binary-valued attributes (this is, of course, an oversimplification). Then there would be 28 possible pairwise correlations to be examined. A child might take a long time to hit on the correct hypothesis. In contrast, consider what nonstrategic behavior might accomplish. The context model is based on storing exemplar information and, as we have seen, is sensitive to correlated attributes. If children store information in the nonindependent manner implied by the context model, then learning a few examples will allow the child's classification to reflect correlated attributes in the absence of any higher level abstraction.

Selective Attention and Context

So far the context model has been presented as if it had no strategic component. Actually, in both its original application to discrimination-learning phenomena (Medin, 1975) and more recent applications to classification performance (Medin & Schaffer, 1978), selective attention has been an integral part of the model. In some attention theories (Zeaman & House, 1963) it is assumed that learning is confined to the dimension to which a subject attends. The context model makes the less stringent assumption that the similarity parameter of two cues along a dimension is less when that dimension is attended than when it is not attended; that is, differences on a dimension are more salient when attention is directed toward that dimension. Only as a special case, however, would the similarity parameter be zero for an attended dimension and one for a nonattended dimension (for related evidence, see Kellogg, 1980).

In the original presentation of the context model (Medin, 1975), I offered the not entirely original suggestion that developmental shifts in discrimination performance might involve an increased ability to overcome the effects of particularized contexts; that is, older children may be more efficient at attending to relevant information and ignoring irrelevant information. Several years earlier, Tighe and Tighe (1972) similarly had argued that whereas all subjects can and usually do learn something about instance–reward relations and the category–reward relations of simple concept tasks, the contribution of each type of solution varies with age. Younger subjects are relatively more likely to learn and remember instance properties (Tighe, Tighe, & Schechter, 1975).

Recently there has been accumulating evidence that there is a fundamental developmental shift in childrens' processing of multidimensional stimuli. Specifically, young children are more likely to process stimulus dimensions as integral (and less likely to process them as separable) than older children (see chapters by Kemler and by Shepp, this volume). The operational definitions of integral and separable dimensions suggest that this change represents an increased ability to attend selectively. This is consistent with the Tighes' theory.

In light of the preceding discussion and analysis of ill-defined concepts, however, it may be erroneous to view the younger child's attentional behavior as a limitation. The same process that is associated with highly specific learning and limited transfer may also be associated with learning correlations among attributes that give structure to many categories. Abstraction is an efficient process when one knows what information should be discarded. In many discrimination-learning tasks it may be clear which information is irrelevant but relevancy

would be difficult to judge in the broader set of classification tasks where the potential contrast categories are not always present. For example, in learning to tell the difference between cats and dogs, size is a good cue whereas domesticity is not. But ignoring domesticity and learning only about size could prove costly later when one might want to distinguish dogs and wolves or dogs and coyotes.

SUMMARY AND A RESIDUAL PROBLEM

Probably very few readers will come away with the conviction that they know just what makes a category a category and what makes one task easy and another hard. I have tried to undermine both the viability of "atheoretical" guidelines for category structure and any residual confidence that categorization theories themselves place psychologically valid constraints on what may be a category. In their stead, the idea that correlated attributes may provide an organizing principle in categorization was briefly touted, but its main attraction may turn out to be that it has not yet been subject to stringent tests.

It remains to be seen just how category representations that exploit correlated attributes relate to the maxims with which we began this chapter. For example, both sensitivity to correlated attributes and cue validity can be thought of as facilitating inferences from attributes and category membership to allow organisms to "go beyond the information given." These inferences will also tend to be more accurate to the extent that within-category similarity is maximized and between-category similarity is minimized. But the main point is not that these maxims have no value; rather they seem more to be *consequences* rather than the *cause* of what makes a category good or sensible. To use a somewhat overdrawn analogy, the maxims may be like stating that good football teams score a lot of touchdowns and give up few touchdowns. Presumably, points scored or given up are by-products of more basic principles of skill. The maxims for category structure may best be thought of as byproducts or as acting in the service of more basic principles. To some extent the same criticisms can be leveled at the principle of correlated attributes.

Some different approaches may be in order, and I mention only one. Within the category bird, there is probably a correlation between the type of feather and whether or not the feet are webbed. But this is not a raw correlation that just happens to emerge. For example, adjusting to an aquatic environment may bring about a number of adaptations (e.g., webbed feet, water-repellent feathers) that would manifest themselves as correlated attributes. Likewise, correlated symptoms in medical diangosis may arise from a common underlying source. I find it intriguing that many of the subjects in our burlosis experiments not only noticed that certain symptoms were correlated but they also offered numerous explanations as to why. It may be that "relatedness" acts as a conceptual glue holding correlated attributes together directly and categories together indirectly.

ACKNOWLEDGMENTS

This research was supported by U.S. Public Health Service Grant MH32370. Ed Shoben, Elissa Newport, Gerald Dewey, Jerry Busemeyer, Carolyn Mervis, and Larry Barsalou provided helpful comments on earlier drafts of this chapter.

REFERENCES

Anderson, J. R., Kline, P. J., & Beasley, C. M. A general learning theory and its application to schema abstraction. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 13). New York: Academic Press, 1979.

Estes, W. K. An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston, 1972.

Estes, W. K. *Memory and conditioning*. In F. J. McGuigan & D. B. Lumsden (Eds.), *Contemporary approaches to conditioning and learning*. New York: Wiley, 1973.

Estes, W. K. Structural aspects of associative models for memory. In C. N. Cofer (Ed.), *The structure of human memory*. New York: Freeman, 1976.

Farah, M. J., & Kosslyn, S. M. Learning concepts. In H. Reese & L. P. Lipsett (Eds.), *Advances in child development and behavior* (Vol. 16). New York: Academic Press, 1981.

Fried, L. S. *Perceptual learning and classification with ill-defined categories*. Michigan Mathematical Psychology Program technical report, 1979.

Garner, W. R. *The processing of information and structure*. New York: Wiley, 1974.

Hayes-Roth, B., & Hayes-Roth, F. Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 321–338.

Homa, D., Rhoads, D., & Chambliss, P. The evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, *5*, 11–23.

Katz, J. J., & Postal, P. M. *An integrated theory of linguistic descriptions*. Caudbridge, Mass.: MIT Press, 1964.

Kellogg, R. T. Is conscious attention necessary for long-term storage? *Journal of Experimental Psychology: Human Learning and Memory*, 1980, *6*, 379–390.

Kosan, N. E. Developmental differences in concept acquisition strategies. *Child Development*, 1981, *52*, 290–298.

Levine, M. *A cognitive theory of learning: Research on hypothesis testing*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Malt, B. C., & Smith, E. E. *Correlational structure in semantic categories*. Unpublished manuscript. Stanford University, 1981.

McCloskey, M. E., & Glucksberg, S. Natural categories: Well-defined or fuzzy sets? *Memory & Cognition*, 1978, *6*, 562–572.

Medin, D. L. A theory of context in discrimination learning. In G. H. Bower (Ed.), *The Psychology of learning and motivation* (Vol. 9). New York: Academic Press, 1975.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1982.

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

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

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

Mervis, C. B., & Rosch, E. Categorization of natural objects. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology*, 1981, *32*, 89–115.

Murphy, G. L. Cue validity and basic levels in categorization. *Psychological Bulletin*, 1982.

Neumann, P. G. An attribute frequency model for the abstraction of prototypes. *Memory & Cognition*, 1974, *2*, 241–248.

Nilsson, N. J. *Learning machines*. New York: McGraw-Hill, 1965.

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

Reitman, J. S., & Bower, G. H. Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, 1973, *4*, 194–206.

Rosch, E. On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.

Rosch, E. Universals and cultural specifics in human categorization. In R. Breslin, S. Bochner, & W. Lonner (Eds.), *Cross-cultural perspectives on learning*. New York: Halsted Press, 1975.

Rosch, E. Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Rosch, E., & Lloyd, B. B. *Cognition and categorization*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. Basic objects in natural categories. *Cognitive Psychology*, 1976, *8*, 382–439.

Sebestyn, G. S. *Decision-making processes in pattern recognition*. New York: Macmillan, 1962.

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

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and processes in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214–241.

Tighe, T. J., & Tighe, L. S. Stimulus control in children's learning. In A. D. Pick (Ed.), *Minnesota symposia on child psychology* (Vol. 6). Minneapolis: University of Minnesota Press, 1972.

Tighe, T. J., Tighe, L. S., & Schechter, J. Memory for instances and categories in children and adults. *Journal of Experimental Child Psychology*, 1975, *20*, 22–37.

Trabasso, T., & Bower, G. H. *Attention in learnings theory and research*. New York: Wiley, 1968.

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

Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency: Psychological theory and research*. New York: McGraw-Hill, 1963.