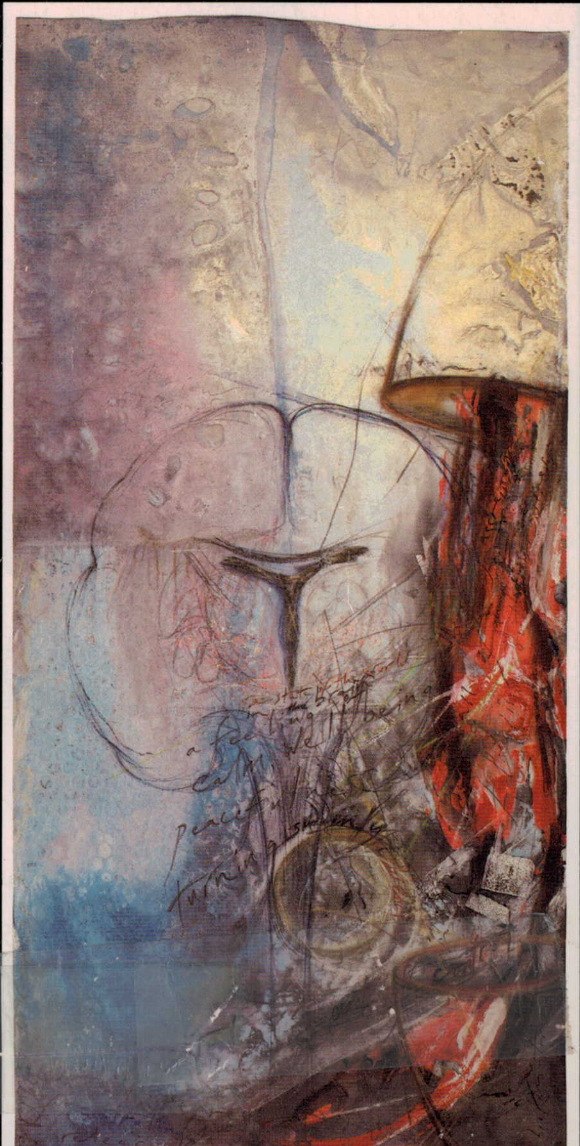
AN INVITATION TO COGNITIVE SCIENCE

THINKING

VOLUME 3



edited by
Daniel N. Osherson
and Edward E. Smith

Neisser, U. (1982). *Memory observed*. San Francisco: W. H. Freeman.

Nilsson, L.-G., and T. Archer, eds. (1985). *Perspectives on learning and memory*. Hillsdale, NJ: L. Erlbaum Associates.

Norman, D. A. (1976). *Memory and attention*. 2nd ed. New York: Wiley.

Osherson, D. N., and S. Weinstein (1984). Formal learning theory. In M. S. Gazzaniga, ed., *Handbook of cognitive neuroscience*. New York: Plenum Press.

Potter, M. C. (1975). Meaning in visual search. *Science* 187, 965–966.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory* 2, 509–522.

Potter, M. C. (1982). Very short-term memory: In one eye and out the other. Paper presented at the Annual Meeting of the Psychonomic Society.

Potter, M. C. (1983). Representational buffers: The eye-mind hypothesis in picture perception, reading, and visual search. In K. Rayner, ed., *Eye movements in reading: Perceptual and language processes*. New York: Academic Press.

Potter, M. C., and E. I. Levy (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology* 81, 10–15.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics* 2, 437–442.

Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory and Cognition* 2, 95–100.

Salomé, P., and A. Baddeley (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology* 41A, 107–122.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13, 501–518.

Schwartz, B. (1984). *Psychology of learning and behavior*. New York: Norton.

Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychological Review* 77, 406–418.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior* 6, 156–163.

Slamecka, N. J., and P. Graf (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory* 4, 592–604.

Smith, E. E. (1978). Theories of semantic memory. In W. K. Estes, ed., *Handbook of learning and cognitive processes*, vol. 6. Hillsdale, NJ: L. Erlbaum Associates.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs* 74 (11, Whole No. 498).

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Squire, L. R. (1989). On the course of forgetting in very long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15, 241–245.

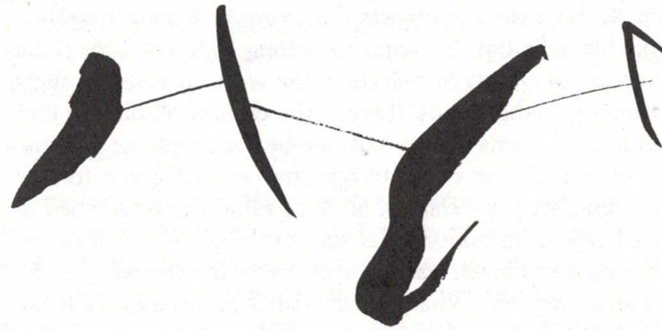
Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology* 25, 207–222.

Wiseman, S., and U. Neisser (1974). Perceptual organization as a determinant of visual recognition memory. *American Journal of Psychology* 87, 675–681.

Yates, F. A. (1966). *The art of memory*. Middlesex, England: Penguin.

Chapter 2
Categorization

Edward E. Smith



We are forever carving nature at its joints, dividing it into categories so that we can make sense of the world. If we see a particular child pet a particular dog at a particular time and a particular place, we code it as just another instance of "children like dogs." In doing this, we reduce a wealth of particulars to a simple relation between the categories "children" and "dogs" and free our mental capacities for other tasks.[1]

What exactly is a category? For now, let us take a *category* to be a class of objects that we believe belong together. (The word *believe* is critical here—we are dealing with the psychological sense of *category*, not the logical sense that is sometimes captured by linguistic theories.) Our major concern in this chapter is with the process by which people assign objects to categories, but this concern requires that we first consider the nature of categories. In section 2.1 we will analyze the nature of categories and consider three characteristics of a class of objects that make it into

Preparation of this chapter was supported by U.S. Public Health Service grant MH 37208. I thank Daniel Osherson for helpful comments on an earlier version.

1. Quotation marks are used throughout the chapter to indicate categories.

a category. One characteristic is the similarity of the objects grouped together, and in section 2.2 we will discuss alternative means for measuring similarity. We will opt for a model in which the similarity of objects is measured in terms of their features. In section 2.3 we will apply this model to categorization tasks and see that it accounts for a variety of empirical findings. In section 2.4 we will briefly look at some other issues in research on categorization.

2.1 What Is a Category? Three Critical Characteristics

2.1.1 Classes and Categories

We take a category to be a class of objects that seem to belong together. The critical part of this definition is "seem to belong together," for there are an indefinite number of classes of objects in the world whose members do not seem to belong together. Thus, there is the class of all objects that weigh an even number of grams (or an odd number of grams, or a prime number of grams, etc.), the class of all things that are *not* green (or *not* round, or *not* democratic, etc.), the class of all things that can be scraped *or* worshiped (or tasted *or* mistrusted, or inflamed *or* envied, etc.), and so on. In all these cases the class of objects has some property in common yet the class is not treated as a category. What characteristics of a class give it the status of categoryhood? Three characteristics are discussed in the following sections.

2.1.2 Coding of Experience

Perhaps the most striking characteristic of a category is that we use it to code experience. We may perceive some complex object as a kind of "chair," remember it as a "chair," describe it to others as a *chair*, and reason about it in the same way. Coding by category is fundamental to mental life because it greatly reduces the demands on perceptual processes, storage space, and reasoning processes, all of which are known to be limited (see, for example, the discussion of short-term memories in chapter 1). This coding aspect of categories is presumably why human languages contain simple terms for categories, such as *tiger*, *chair*, and *mother*; that is, frequently used codes are associated with brief descriptions.

Categories vary in the extent to which they are used as codes. Categories are often structured into a *taxonomy*—a hierarchy in which successive levels refer to increasingly more specific objects—and categories at an intermediate level are more likely to be used to code experience than are categories at lower or higher levels (Rosch et al. 1976). Consider the taxonomy for fruits. The category "fruit" would be at a high or

superordinate level, "apple" would be at an intermediate or *basic* level, and "McIntosh apple" would be at a relatively low or *subordinate* level. (For objects, the basic level may be identified with the most abstract level that is associated with a specific shape; the superordinate and subordinate levels are simply the levels above and below the basic one.) Here, *apple*, which is at the basic level, would be the preferred code, as witnessed by the facts that (1) people overwhelmingly prefer to name a particular object *apple* rather than *fruit* or *McIntosh apple*, and (2) they can decide that a particular apple is an "apple" faster than they can decide that it is a "fruit" or a "McIntosh apple" (Rosch et al. 1976).

Note that this coding aspect of categories does not apply to classes that are not categories. Thus, generally we do not code things as "objects that weigh an even number of grams" or as "objects that can be scraped or worshiped." Nor are there simple terms in the language for these classes.

2.1.3 Inductive Inferences

Whenever we use existing beliefs to generate new ones, we have drawn an inference. An inference can be either "deductive," in which case it is *impossible* for the new belief to be false if the old ones are true, or "inductive," in which case it is *improbable* for the new belief to be false if the old ones are true (see Skyrms 1986). There is an intimate relation between inductive inferences and categories; namely, categorization of an object licenses inductive inferences about that object.

An experimental demonstration used by Gelman and Markman (1986) illustrates this relation. On each trial of the experiment subjects were presented three pictures, where the third picture looked like one of the first two but was from the same category as the other picture. For example, on one trial the pictures were of a flamingo, a bat, and a blackbird, where the blackbird resembled the bat. New information was given about the first two pictures, then a question was asked about the third one. For example: regarding the flamingo, subjects were told, "This bird's heart has a right aortic arch only"; regarding the bat, they were told, "This bat's heart has a left aortic arch only"; and regarding the blackbird, they were asked, "What does this bird's heart have?" Subjects responded with "right aortic arch only" almost 90 percent of the time, thus basing their decision on common category membership rather than physical similarity. More surprisingly, when 4-year-old children were tested in the same paradigm (though with simpler properties), they based their decision on category membership almost 70 percent of the time. Very early on, we know that members of the same category are likely to share many invisible properties even if they do not resemble one another.

Different kinds of categories differ in the extent to which they support inductive inferences. For one thing, basic and subordinate categories support more inferences than do superordinate categories (Rosch et al. 1976). For example, people will attribute far more properties to an object classified as an "apple" or a "McIntosh apple" than to an object classified as a "fruit." (There is little difference, though, between the number of inductive inferences supported by basic categories and the number supported by subordinate categories.)

Another distinction among categories that has implications for induction is that between *natural kinds* like "tiger" and "daisy," which deal with naturally occurring species of flora and fauna, and *artifact kinds* like "chair" and "shirt," which deal with person-made objects (see, for example, Schwartz 1979). Natural kind categories seem to support more inductive inferences about invisible properties than do artifact kinds. Having been told, for example, that some chair has a particular nonvisible property—say, that is has lignin all through it—we may be hesitant to conclude that another chair has this property, at least compared to the ease with which we generalize from a flamingo's having a right-aortic-arch heart to another bird's having such a heart (Gelman and O'Reilly 1988).

Categories in general support more inductive inferences than do classes that are not categories. We draw more inferences, say, about "fruit" or "furniture" than about "objects that weigh an even number of grams."

2.1.4 Similarity

Another characteristic of many categories is that their members tend to be physically similar to one another while being physically dissimilar from members of contrasting categories. Of course, there are limits to this, as in the earlier example where one bird was less similar to another bird than to a bat. Still, in general we divide the world so as to maximize within-category similarity while minimizing between-category similarity.

The extent to which this characteristic is manifest again depends on the taxonomic level of the categories. At the superordinate level members of a category need not resemble one another; instances of "fruit," for example, may share few physical properties (consider a raisin and a watermelon). At the subordinate level members of a category closely resemble one another, but they also resemble members of contrasting categories (two McIntosh apples look very much alike, but they also resemble a Delicious apple). It is primarily at the basic level that members of a category resemble one another *and* look different from members of contrasting categories (two apples look like each other yet differ from oranges or peaches).

As usual, this characteristic of categories seems not to apply to classes that are not categories. On the average, there is little physical similarity among "objects that weigh an even number of grams," or among "objects that can be scraped or worshiped."[2]

2.1.5 Relations among the Three Characteristics

Two questions arise about the relations among the three characteristics. (1) Do they cohere; that is, do they pick out the same classes as categories? (2) Do they give the same kind of information about concepts?

With regard to coherence, there is substantial convergence among the three criteria, at least for basic categories. A basic category ("apple") is often used to code experience, affords numerous inductive inferences (particularly if it is a natural kind category), and tends to maximize within-category similarity while minimizing between-category similarity. For non-basic categories, there is less convergence. Although a subordinate category ("McIntosh apple") may be used to code experience in some contexts, the fact that it is rarely denoted by a single term suggests limits to its coding potential; further, although a subordinate category supports numerous inferences, it maximizes within-category similarity at the cost of substantial between-category similarity. In contrast, although a superordinate category ("fruit") also may be used to code experience in some contexts, it promotes few inductive inferences and clearly does not maximize within-category similarity.

With regard to the second question, the three characteristics seem to have different natures. Similarity represents a *guide* to categorization, whereas the other two characteristics generally reflect the *consequences* of categorization. To the extent that members of a category are similar to one another yet dissimilar from instances of other categories, we can decide whether or not a novel object belongs to the category by assessing its similarity to known category members (versus its dissimilarity from known nonmembers). Once this categorization is made, we can code the object in terms of the category (with a simple term) and infer hidden properties of the object.[3]

Because our primary interest lies in the process of categorization, and not in its products, we will focus on the similarity characteristic of categories. We will assume for the time being that assigning an object to a category rests on determining that the object is sufficiently similar either to known members of the category or to a summary of known members. Our

2. However, similarity considerations alone cannot explain why we have the categories that we do. For example, if the only criterion for categoryhood was to maximize within-class similarity, then all categories should have only one member (Medin 1983)!

3. I am oversimplifying here with regard to what is a guide versus what is a consequence of categorization. For example, knowing that two objects belong to the same category can make them seem more similar (Tversky 1977), in which case similarity is a consequence of categorization. Still, the basic distinction drawn in the text covers most cases.

next order of business is to find a means for measuring the similarity between a pair of objects or between an object and a summary of category members.[4]

2.2 Measurement of Similarity

There are two general approaches to the measurement of similarity: geometric and featural.

2.2.1 Geometric Approach

In the geometric approach, objects or items are represented as points in some multidimensional space such that the metric distance between two points corresponds to the *dissimilarity* between the two items. To illustrate, figure 2.1 represents 20 different fruits, as well as the category "fruit" itself, in a two-dimensional space. The shorter the metric distance between a pair of points, the more similar the corresponding fruits. For example, "apple" is more similar to "plum" than to "date," but more similar to "date" than to "coconut."

The space in figure 2.1 was constructed by a systematic procedure developed by Shepard (1962). First, a group of subjects rated the similarity between every possible pair of items ("apple"-"banana," "apple"-"plum," "apple"-"fruit," and so on—210 distinct pairs in all, for the items represented in figure 2.1). The similarity ratings were then input to a computer program that used an iterative procedure to position the items in a space (predetermined to have a certain dimensionality) so that the metric distance between items corresponded as closely as possible to the (inverse of) judged similarity between the items.

Crucial to the representation in figure 2.1 is the assumption that psychological distance is "metric" (just as ordinary physical space is). That is, it is assumed there is a function, d, that assigns to every pair of points a nonnegative number, their "distance," in accord with the following three axioms:

(1) *Minimality*
$d(a, b) \geqslant d(a, a) = d(b, b) = 0$

(2) *Symmetry*
$d(a, b) = d(b, a)$

4. There is more to categorization than similarity. For one thing, sometimes categorization involves determining whether or not an object satisfies a definition. Although natural kind and artifact kind categories lack true definitions (see, for example, Putnam 1975), "nominal kind" categories like "uncle," "felony," and "even number" seem to have them (Schwartz 1979). Nominal kind categories are tailor-made for some specialized system, such as kinship, law, or arithmetic. Deciding that something fits in such a category presumably involves determining that it meets the definition, though even here factors like similarity may play some role (Armstrong, Gleitman, and Gleitman 1983). Another matter is that categorization sometimes involves inductive reasoning; this matter is discussed in section 2.3.3.

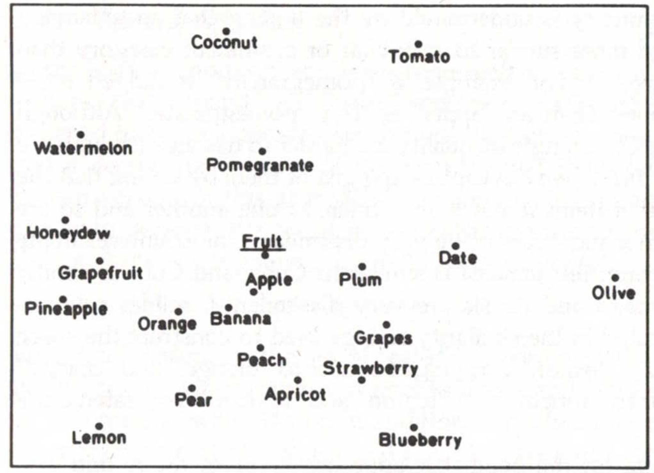


Figure 2.1
A two-dimensional space for representing the similarity relations among 20 instances of fruit and the category "fruit" itself. (From Tversky and Hutchinson 1986.)

(3) *Triangle inequality*
$d(a, b) + d(b, c) \geqslant d(a, c)$.

Minimality says that the distance between any item and itself is identical for all items, and is the mimimum possible. *Symmetry* says that the distance between two items is the same regardless of whether we start at one item or the other. And *triangle inequality* essentially says that the shortest distance between two points is a straight line. All three assumptions are evident in figure 2.1. For example, the distance between "peach" and "date" is (1) greater than that between "peach" and "peach," (2) equal to that between "date" and "peach," and (3) less than the sum of the distances between (a) "peach" and "apple" and (b) "apple" and "date."

The geometric approach has a history of success in representing perceptual objects (for a partial review, see Shepard 1974). Given a two-dimensional representation of color, for example, one can use the distances between the colors to accurately predict the likelihood that a subject in a memory experiment will confuse one color with another. However, the geometric approach works less well in representing conceptual items, such as categories and their instances. Indeed, for conceptual items, Tversky (1977) has produced evidence against each one of the metric axioms.

Minimality is compromised by the fact that the more we know about an item, the more similar it is judged to itself. The president of the United States, for example, seems more similar to himself than does some obscure member of the House or Senate. A familiar category like "apple" seems more similar to itself than does an unfamiliar category like "pomegranate."

The axiom of symmetry is undermined by the finding that an unfamiliar category is judged more similar to a familiar or prominent category than the other way around. For example, a "pomegranate" is judged more similar to an "apple" than an "apple" is to a "pomegranate." Although exact violations of the triangle inequality are harder to describe (though see Tversky and Gati 1982), we can capture the gist of them by noting that the axiom implies that if items a and b are similar to one another and so are items b and c, then a and c cannot be very dissimilar. One counterexample to this involves countries: Jamaica is similar to Cuba, and Cuba is similar to Russia, but Jamaica and Russia are very dissimilar. A milder counterexample is manifested in the similarity ratings used to construct the space of fruits in figure 2.1: "lemon" was judged similar to "orange," and "orange" was judged similar to "apricot," but "lemon" and "apricot" were rated quite dissimilar.

Another problem for the geometric approach involves the notion of a "nearest neighbor" (Tversky and Hutchinson 1986). If we obtain similarity ratings for pairs of items, as in the fruit example, then for each item we can refer to the item rated most similar to it as its "nearest neighbor." We can now characterize an item by how many other items it is the nearest neighbor to. When this was done with the similarity ratings for fruits, the category "fruit" turned out to be the nearest neighbor for 18 of the 20 other terms. This finding is problematic because it is impossible for one item in a metric space to be a nearest neighbor to so many other items as long as the space is of relatively low dimensionality. In fact, in a two-dimensional space the maximum number of items to which another item can serve as nearest neighbor is five (look at figure 2.1). At a minimum, a nine-dimensional space is needed to accommodate "fruit" being the nearest neighbor to 18 items. And once "fruit" is positioned in such a space so that it is the nearest neighbor to the appropriate 18 items, there is no guarantee that the distances between the 18 items themselves will adequately capture the similarity ratings for the relevant pairs of items. The general problem is that a category serves as the nearest neighbor to many of its instances, so many as to call into question the appropriateness of low-dimensional metric representations.

The above challenges to the geometric approach are not without their critics. Defenders of the geometric approach have argued, for example, that violations of symmetry and the triangle inequality arise more often when similarity is judged directly ("Rate the similarity of a to b") than when it is judged indirectly (say, by the frequency with which a and b are confused with one another). This suggests that direct judgments require complex decision processes that are the source of the asymmetries (Krumhansl 1978; Nosofsky 1986). Still, at this moment the weight of the evidence points away from geometric representations of categories.

2.2.2 Featural Approach

In the featural approach, an item is represented as a set of discrete features, such as "red", "round," and "hard," and the similarity between two items is assumed to be an increasing function of the features they have in common and a decreasing function of the features that they differ on. The best-known version of this approach is Tversky's (1977) contrast model. The similarity between the set of features characterizing item i (labeled I) and the set characterizing item j (labeled J) is given by (4):

(4) $\mathrm{Sim}\,(I, J) = af(I \cap J) - bf(I - J) - cf(J - I).$

Here $I \cap J$ designates the set of features common to the two items, $I - J$ designates the set of features distinct to item i, and $J - I$ designates the set of features distinct to item j. In addition, f is a function that measures the salience of each set of features, and a, b, and c are parameters that determine the relative contribution of the three feature sets.

Table 2.1 illustrates the contrast model with examples drawn from the domain of fruits. Each panel of the table deals with a phenomenon that surfaced in our discussion of the geometric approach. Panel 1 is concerned with minimality. It contains possible feature sets for the categories "apple" and "pomegranate." There are more features for "apple" than for "pomegranate," reflecting the fact that "apple" is the more familiar item. This difference will result in "apple" being rated more similar to itself than is "pomegranate," because the more features an item has, the more common features there are when the item is compared to itself. This idea is detailed in the calculations given below each pair, where the contrast model has been used to calculate the similarity between the members of the pair. For purposes of simplicity, here and elsewhere, we will assume that the function f simply assigns a value of 1 to each feature in a set of common or distinctive features.[5]

Panel 2 is concerned with symmetry. It compares the similarity of "pomegranate" to "apple" versus that of "apple" to "pomegranate." As the calculations show, the contrast model is compatible with the fact that "pomegranate" is more similar to "apple" than vice versa as long as parameter b exceeds parameter c.

Panel 3 demonstrates that the contrast model is compatible with violations of the triangle inequality: "lemon" is similar to "orange," and "orange" is similar to "apricot," but "lemon" is not similar to "apricot." As the calculations show, the violation will be pronounced whenever the weight given to common features, a, exceeds that given to either set of distinctive features, b or c, because then similarity will be relatively large for the first two pairs but not for the third.

5. These feature sets are derived from the work of Smith et al. (1988), who had 30 subjects list features of 15 different instances of "fruit" (including "apple," "pomegranate," "orange," "lemon," and "apricot").

Table 2.1
Some illustrations of the contrast model.

Apple	Apple	Pomegranate	Pomegranate
red	red	red	red
round	round	round	round
hard	hard		
sweet	sweet		
trees	trees		
$\mathrm{Sim}(A, A) = a(5) - b(0) - c(0)$		$\mathrm{Sim}(P, P) = a(2) - b(0) - c(0)$	

Pomegranate	Apple	Apple	Pomegranate
red	red	red	red
round	round	round	round
	hard	hard	
	sweet	sweet	
	trees	trees	
$\mathrm{Sim}(P, A) = a(2) - b(0) - c(3)$		$\mathrm{Sim}(A, P) = a(2) - b(3) - c(0)$	

Lemon	Orange	Orange	Apricot	Lemon	Apricot
yellow	orange	orange	red	yellow	red
oval	round	round	round	oval	round
sour	sweet	sweet	sweet	sour	sweet
tree	tree	tree	tree	tree	tree
citrus	citrus	citrus		citrus	
ade	ade	ade		ade	
$\mathrm{Sim}(L, O) = a(3) - b(3) - c(3)$		$\mathrm{Sim}(O, A) = a(3) - b(3) - c(1)$		$\mathrm{Sim}(L, A) = a(1) - b(5) - c(3)$	

Apple	Plum	Apple	Fruit
red	red	red	red
round	round	round	round
hard	soft	hard	hard
sweet	sweet	sweet	sweet
trees	trees	trees	
$\mathrm{Sim}(A, P) = a(4) - b(1) - c(1)$		$\mathrm{Sim}(A, F) = a(4) - b(1) - c(0)$	

Finally, panel 4 establishes that the contrast model is compatible with the fact that a category can serve as a nearest neighbor to numerous instances. Among "fruit" instances, "plum" is often rated most similar to "apple." But as the calculations show, "fruit" is an even closer neighbor to "apple." The reason is that "fruit" is more abstract than "plum" and hence includes fewer distinctive features.

In sum, the contrast model offers a satisfactory account of the phenomena that plagued the geometric approach, and we will use the model in what follows.

However, we should note that the contrast model does have some limitations. First, it does not tell us what an item's features are. For each domain of inquiry, like that of plant categories, researchers need independent procedures for determining the features of the various items (asking people to list features of items is one such procedure, albeit a rough one).

Second, the contrast model does not offer any theory of the function f that measures the salience of each set of features. Such a theory would have to address issues about the intensity of individual features (for instance, a more saturated color might be assigned a greater salience than a less saturated one), as well as issues about the diagnosticity of features (for instance, a feature that discriminates among relevant objects might be assigned a higher salience than one that does not). The theory would also have to specify how and why people differ in the salience they assign to the same feature in the same context.

Third, although the contrast model tells us *what* is computed—measures of sets of common and distinctive features—it says little about the algorithms used to effect the computation. Thus, the model does not tell us whether the features of two items are compared simultaneously or sequentially, and if the latter, in what order.

As we will see, in applying the contrast model we will have to add auxiliary assumptions to deal with these three limitations.[6]

2.3 Similarity and Categorization

Now that we have some insight into the measurement of similarity, we are in a position to appreciate that similarity underlies some important phenomena in categorization.

6. A more specific (and more remediable) limitation of the contrast model concerns additivity. Most applications of the model assume that the salience assigned to a set of features is an additive function of the individual saliences of the features that constitute the set. In fact, though, Tversky (1977) derived the contrast model from a set of qualitative axioms, and his derivation does not yield additivity of feature saliences. More recently Osherson (1987) has derived the contrast model from a different set of qualitative axioms, and his derivation does guarantee the additivity of feature saliences.

2.3.1 Typicality Effects

People can reliably order the instances of any category with respect to how "typical" or "prototypical" or "representative" they are of the category. Table 2.2 presents typicality ratings for the categories "fruit" and "bird." These ratings were obtained by instructing subjects to rate typicality on a 7-point scale, with 7 corresponding to the highest typicality and 1 to the lowest (Malt and Smith 1984). "Apple" and "peach" are considered typical fruits, "raisin" and "fig" less typical, and "pumpkin" and "olive" atypical. Similar variations are found among the instances of "bird." Ratings like these have been obtained for numerous categories and have been shown to be relatively uncorrelated with the frequency or familiarity of the instances (Mervis, Catlin, and Rosch 1976).

What is most important about these ratings is that they predict how efficiently people can categorize various instances. Consider an experimental task that is frequently used to study categorization. On each trial a subject is given the name of a target category, such as "bird," followed by a test item. The subject must decide as quickly as possible whether the test item names an instance of the target category, such as "robin," or a noninstance, such as "trout." The main data of interest are the decision times for correct categorizations. When the test item in fact names a member of the target category, categorization times decrease with the

Table 2.2
Typicality ratings for 15 instances of "fruit" and "bird" (from Malt and Smith 1984).

Fruit	Rating	Bird	Rating
apple	6.25*	robin	6.89
peach	5.81	bluebird	6.42
pear	5.25	seagull	6.26
grape	5.13	swallow	6.16
strawberry	5.00	falcon	5.74
lemon	4.86	mockingbird	5.47
blueberry	4.56	starling	5.16
watermelon	4.06	owl	5.00
raisin	3.75	vulture	4.84
fig	3.38	sandpiper	4.47
coconut	3.06	chicken	3.95
pomegranate	2.50	flamingo	3.37
avocado	2.38	albatross	3.32
pumpkin	2.31	penguin	2.63
olive	2.25	bat	1.53

*Ratings were made on a 7-point scale, with 7 corresponding to the highest typicality.

typicality of the test item. With "bird" as the target, for example, test items corresponding to "robin" and "swallow" are categorized more quickly (by somewhere between 50 and 100 milliseconds) than those corresponding to "owl" and "vulture," which in turn are categorized more quickly (again by between 50 and 100 milliseconds) than test items corresponding to "flamingo" and "penguin" (see, for example, Smith, Shoben, and Rips 1974).

These results in no way rest on the verbal nature of the paradigm. If the task is modified so that the test items are pictures of particular objects (for instance, a pictured robin or vulture or trout), the results are virtually unchanged. Furthermore, to the extent that there is variation in the accuracy of these categorizations (in either the verbal or the pictorial task), error rates also decrease with the typicality of the test items. These effects are extremely reliable: they have been documented in more than 50 experiments that have used many different variants of the verbal and pictorial categorization tasks (for a partial review, see Smith and Medin 1981).

There is also evidence that categorization depends on typicality in more naturalistic settings. A child developing language acquires the names of typical category members before those of atypical ones. And if children are asked to sort pictured objects into categories, their sortings resemble those of adults more if the objects are typical than if they are atypical (Mervis 1980; Rosch 1978).

2.3.2 Typicality as Similarity

A general interpretation of the above findings is that the typicality of an instance is a measure of its similarity to its category, and categorization amounts to determining that an item is sufficiently similar to the target category. In what follows we will flesh out this interpretation.

If typicality is really similarity, then the contrast model should be able to predict typicality ratings. To test this, we (1) select a domain of instances, (2) estimate the features of the instances and the category (remember, the contrast model does not supply these), (3) apply the contrast model to each instance-category pair, and (4) see whether this estimate of instance-category similarity correlates with the rated typicality of the instance in the category.

The instances we will use as well as their features are taken from a study by Malt and Smith (1984), in which subjects had 90 seconds to list all the features they could think of for each instance. Table 2.3 contains a small subset of the features obtained. In the experiment 30 subjects were each presented 15 instances of "bird"; they collectively produced more than 50 features, each feature being produced by more than one subject. Table 2.3 considers only nine of the instances and six of the features: flies, sings, lays

Table 2.3
Illustrations of how to use listed properties to calculate an instance's similarity to prototype.

Features	Robin	Bluebird	Swallow	Starling	Vulture	Sandpiper	Chicken	Flamingo	Penguin	Bird
Flies	+	+	+	+	+	+	−	−	−	+
Sings	+	+	+	+	−	+	−	−	−	+
Lays eggs	+	+	+	−	−	+	+	−	+	+
Is small	+	+	+	+	−	+	−	−	−	+
Nests in trees	+	+	+	+	+	−	−	−	−	+
Eats insects	+	+	+	+	−	+	−	−	−	+
Similarity to bird	6−0−0=6	6−0−0=6	6−0−0=6	5−.5−0=4.5	2−2−0=0	5−.5−0=4.5	1−2.5−0=−1.5	0−3−0=−3	1−2.5−0=−1.5	

eggs, is small, nests in trees, and eats insects. The rows of the table list the six features; the columns give the instances in order of decreasing typicality, with the last column representing the category "bird." Each entry in the resulting matrix is a + or a −, where + indicates that at least two subjects listed the feature for that instance and a − indicates that either one or no subjects did. To determine the entries for "bird," a feature was assigned a + only if a majority of the instances had a + for that feature. The category "bird" thus contained the frequent features of the instances.

The contrast model was used to determine the similarity of each instance in table 2.3 to "bird." In making the calculations (given at the bottom of the table), it was assumed that (1) all features are equally salient (that is, f assigns a value of 1 to each feature, which means that the salience of a set of common or distinctive features is simply the number of features it contains), and (2) common features count more than distinctive ones, with features distinct to the category counting more than those distinct to the instance (specifically, $a = 1$, $b = \frac{1}{2}$, $c = \frac{1}{4}$). The contrast model correctly segregates the instances in table 2.3 into three levels of typicality (3 high, 3 medium, and 3 low), though it makes few distinctions among the instances within each level. Finer distinctions can readily be made by assuming that features differ in their salience or by considering more features.

You can verify that had the average similarity of an instance to all other instances been computed (rather than its similarity to "bird"), virtually the identical similarity scores would have been obtained. Hence, the success of the contrast model in predicting typicality does not depend on whether a category is taken to be an abstraction or a set of instances.

Let us now look briefly at how the above account could be extended into a model of categorization that could explain some of the experimental results mentioned earlier. The general ideas are (1) an item will be categorized as an instance of a category if and only if it exceeds some criterial

level of similarity to the category, and (2) the time needed to determine that an item exceeds this criterial level of similarity is less the more similar the item in fact is to the category. When these two assumptions are joined with the claim that an item's typicality reflects its similarity to its category, it follows that more typical items will be categorized faster.

Fleshing out this model requires making specific assumptions about the algorithms used to implement the contrast model. One possibility is to assume that all features of the instance and category are compared in parallel—with common features incrementing a similarity counter and distinctive features decrementing it—and the outcomes of these feature comparisons become available at different points in time. If an instance is only moderately similar to its category, the process may have to wait for late-arriving feature matches (common features) to reach threshold. In contrast, if an instance is highly similar to its category, the early-arriving feature matches may suffice to pass threshold.

A related approach is expressed in terms of *spreading activation* (a mechanism that figures centrally in memory; see chapter 1). When an item and category are presented, activation from these two sources begins to spread to the features associated with them, with the activation from each source being subdivided among its features. If the two sources of activation intersect at some features (common features), further processing is undertaken to determine that an instance-category relation holds. Because the number of intersecting or common features generally increases with the typicality of an instance to its category, there are more opportunities for an intersection with typical than atypical instances, and hence more opportunities for an early termination of the process (Collins and Loftus 1975). (In this model, features distinct to the category or item slow the process by thinning the activation from each source.)

The above models may suffice for the case where only one category is relevant (as in the experiments described earlier), but often people have to decide which of *n* relevant categories is the correct one (Is this plant a mushroom or a toadstool? Is that car a Chevy or a Ford?). In such cases a categorization model has to consider the relation of an item to the categories that contrast with the correct one. Thus, a categorization decision may consider something like the ratio between the similarity of the instance to the target category versus the similarity of the instance to all contrasting categories (Nosofsky 1986).[7]

2.3.3 Beyond Similarity

The approach we have taken accurately describes categorization in many cases, but some recent experiments demonstrate situations where categorization is based on something other than similarity.

Two studies by Rips (1989) suffice to make the point. In the first study, on each trial a subject was presented a description of an object that mentioned only a value on a single dimension (say, an object's diameter). Then the subject decided which of two categories the object belonged to, where prior work had established that the object was between the subject's average values for the two categories. For example, one item was "an object three inches in diameter," and the associated categories were "pizzas" and "quarters." Although the object was if anything closer to the average diameter of a quarter (indeed, another group of subjects had judged it more similar to a quarter), subjects judged it more likely to be a pizza than a quarter, presumably because there is an official constraint on the size of quarters but not on the size of pizzas. This kind of situation obtained on all trials, as one category always allowed more variability on the relevant dimension than did the other, and subjects consistently chose the high-variability category. These results indicate that categorization decisions consider variability as well as similarity.

A second study by Rips (1989) provides further evidence against similarity-based categorization. Subjects were told about an animal that started out with typical bird properties but suffered an accident that caused many of its properties to resemble those of an insect. Subjects were further told that eventually this animal mated with a normal female of its species, who produced normal young. Subjects rated this creature as more likely to be a

7. It is worth pointing out that tasks other than categorization are affected by typicality, including memory and reasoning tasks. As one example, when asked to generate from memory all instances of a category, subjects retrieve instances in order of decreasing typicality (Rosch 1978). Examples of typicality effects on reasoning will be discussed in chapter 3. The general point is that a similarity-to-category computation may be a general component of mental life.

"bird" than an "insect," but more similar to an "insect" than a "bird." Here we have a situation where categorization and similarity go in different directions.

In these studies subjects seem to be reasoning more than categorizing. Rather than restricting themselves to the features of the test item and target categories—for instance, the features of the accident-prone bird and the categories "bird" and "insect"—subjects seem to be bringing to bear other beliefs and knowledge—for instance, "Animals produce offspring of the same kind as themselves." And rather than just comparing features, subjects seem to be constructing arguments—for instance, "This animal accidentally acquired insect properties but it produced normal bird offspring, so probably it's still a bird." In short, subjects seem to be reasoning inductively. Although it is not yet known which situations lead to reasoning-based categorization and which to similarity-based categorization, it seems plausible that similarity is involved in rapid, automatic decisions, whereas reasoning comes into play in slower, more deliberative decisions.[8]

2.4 Summary and Other Issues

The essential story goes as follows. Categories, at least basic ones, seem to be readily distinguishable from other classes in that they have far greater coding potential and induction potential. Further, categories, at least basic ones, tend to maximize within-category similarity while minimizing between-category similarity. The latter property allows categorization to occur by determining a test item's similarity to known exemplars, or to a summary of the category.

Detailing the categorization process requires specifying a precise means for computing similarity between a pair of items. Both geometric and featural approaches to similarity offer such means, but a number of empirical phenomena (such as asymmetries in similarity judgments) indicate that the featural approach, particularly Tversky's (1977) contrast model, is best for measuring the similarity among categories and their instances. Studies have shown that the similarity of an instance to a category as determined by the contrast model is part of what lies behind the instance's typicality to its category. And an instance's typicality to its category predicts numerous aspects of categorization decisions such as their speed and accuracy. Although the extent of these claims is somewhat compromised by demon-

8. The distinction of interest is phrased here in terms of "similarity versus reasoning" for purposes of exposition. In a more extensive treatment of the issue, similarity itself would be a kind of inductive reasoning, and the distinction of interest would be between similarity and quasi-deductive forms of reasoning (Osherson, Smith, and Shafir 1986).

strations of reasoning-based categorization, our essential story still covers a lot of ground.

In order to keep our focus on similarity, we have had to deemphasize other issues in categorization research. Three such issues deserve at least brief mention.

First, it was noted at a couple of points that a category may be thought of either as an abstract summary or as a set of exemplars. (The similarity proposal was phrased so that it was noncommittal on this issue.) The category "bird," for example, could be mentally represented either by its own set of features or by a set of specific exemplars ("robin," "bluejay," and so on), each with its own set of features. Though at first blush it seems more natural to think of a category as an abstraction, it is apparent that an exemplar representation coupled with the right similarity algorithms can account for much of the data in categorization.

Further, studies on this issue have often found that the ease of learning an instance-category relation is better predicted by the similarity of the instance to the other category exemplars than by the similarity of the instance to a summary representation (see, for example, Estes 1986; Medin and Schaffer 1978). However, other arguments favor an abstract summary. For example, frequently we learn facts about a general class rather than about specific exemplars, such as "All birds lay eggs," and it seems likely that we store such facts as summary information. Given the mixed evidence on this issue, some sort of hybrid position (abstraction-plus-exemplar) may be called for.

Second, we have dealt in this chapter mainly with "simple" categories—roughly those denoted by single words like "apple" and "bird"—and have ignored conjunctive categories—roughly those denoted by more than one word like "dry apple" and "very large bird." Because many conjunctive categories are novel combinations and hence cannot be learned from experience, there must be some procedures for composing conjunctive categories out of simple ones. (This is similar to the composition-of-meaning issue addressed in Larson 1990.)

A number of composition processes have been proposed, particularly for the case where a single modifier is applied to a simple category as in "dry apple." One of these proposals is an extension of the similarity model advanced earlier (Smith et al. 1988). Roughly, the modifier selectively changes those features of the simple category that are mentioned in the modifier (for example, *dry* changes the taste feature of *apple* but not its size feature); then a decision about whether or not an item is an instance of the conjunctive category can be made in exactly the same way as before (by employing the contrast model). Some other models of composition involve reasoning-based categorization (see, for instance, Cohen and Murphy 1984); still other proposals involve applications of *fuzzy set theory*, a generalization of traditional set theory that provides functions for relating membership in conjunctive sets (categories) to membership in simpler ones (see, for instance, Zadeh 1982).

Third, virtually everything in this chapter assumes that natural kind and artifact kind categories do not have definitions and consequently that categorization with such categories involves something other than instantiating a definition. This position has been widely accepted in psychology (see, for example, Smith and Medin 1981). Though some of the best-known arguments for the position come from work in philosophy of mind (for example, Kripke 1972; Putnam 1975; and see chapter 7), there is a gap between the psychological and philosophical work on this problem (see Rey 1983).

Suggestions for Further Reading

For further discussion of the distinguishing characteristics of categories and their dependence on taxonomic level, see Rosch et al. 1976 and Rosch 1978. For a look at other distinctions between categories, particularly those that have to do with kinds, see Schwartz 1979.

On the matter of measuring similarity between instances and categories, perhaps the single most important paper is Tversky 1977. For a discussion of the geometric approach in general, see Shepard 1974.

A psychological perspective on typicality effects and categorization is provided in Smith and Medin 1981. For a philosophical perspective on these same issues, see Rey 1983. Murphy and Medin 1985 offers a summary of the reasoning-based approach to categorization. Finally, for a more advanced treatment of many of the problems considered in this chapter, with a particular emphasis on the similarity-reasoning distinction, see Smith 1989.

Categorization is intimately connected to concepts; indeed, psychologists often assume that a concept is a mental representation of a category. The psychological study of concept development is discussed in chapter 6. An analysis of concepts is also essential for an understanding of language, particularly its development, as words may be construed as names of concepts. Language development is the subject of Pinker 1990. In addition to psychological studies, there is of course a rich tradition of analyses of concepts in philosophy of mind; here, chapter 7 and Schwartz 1977 provide a useful entry point into the recent literature.

Categorization is also related to inductive reasoning. As noted earlier, categorization sometimes involves a kind of inductive reasoning. In other cases, however, inductive reasoning reduces to something like similarity-based categorization, and these cases are discussed in chapter 3.

Research on how we categorize objects has also proven useful for understanding how we categorize people. For recent reviews of work on person categorization, see Cantor and Kihlstrom 1986 and Markus and Zajonc 1985.

Questions

2.1 Use the contrast model to determine the ordering by typicality of the five vegetables given below (the features are given under the name of each instance). Assume that all features are weighted equally and that $a = 1$, $b = \frac{1}{2}$, and $c = \frac{1}{2}$.

Stringbean	Carrot	Cauliflower	Seaweed	Broccoli	Vegetable
green	orange	white	green	green	green
long	long	round	long	long	long
hard	hard	hard	stringy	bushy	hard

2.2 In question 2.1, what changes are there in the ordering of typicality if color is weighted three times more than the other features?

2.3 It was noted in the text that the feature similarity model can be extended to explain categorization with a conjunctive category like "dry apple." Can such a feature model be extended to a conjunction like "fake apple"?

References

Armstrong, S. L., L. R. Gleitman, and H. Gleitman (1983). What some concepts might not be. *Cognition* 13, 263–308.

Cantor, N., and J. Kihlstrom (1986). *Personality and social intelligence*. Englewood Cliffs, NJ: Prentice-Hall.

Cohen, B., and G. L. Murphy (1984). Models of concepts. *Cognitive Science* 8, 27–60.

Collins, A. M., and E. F. Loftus (1975). A spreading-activation theory of semantic processing. *Psychological Review* 82, 407–428.

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

Gelman, S. A., and E. Markman (1986). Categories and induction in young children. *Cognition* 23, 183–209.

Gelman, S. A., and A. W. O'Reilly (1988). Children's inductive inferences with superordinate categories: The role of language and category structure. *Child Development* 59, 876–887.

Kripke, S. (1972). Naming and necessity. In D. Davidson and G. Harman, eds., *Semantics of natural language*. Dordrecht, Holland: Reidel.

Krumhansl, C. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review* 85, 445–463.

Larson, R. K. (1990). Semantics. In D. N. Osherson and H. Lasnik, eds., *Language: An invitation to cognitive science, volume 1*. Cambridge, MA: MIT Press.

Malt, B. C., and E. E. Smith (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior* 23, 250–269.

Markus, H., and R. B. Zajonc (1985). The cognitive perspective in social psychology. In G. Lindzey and E. Aronson, eds., *The handbook of social psychology*. New York: Random House.

Medin, D. L. (1983). Structural principles in categorization. In T. Tighe and B. Shepp, eds., *Perception, cognition and development: Interactional analyses*. Hillsdale, NJ: L. Erlbaum Associates.

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

Mervis, C. B. (1980). Category structure and the development of categorization. In R. Spiro, B. C. Bruce, and W. F. Brewer, eds., *Theoretical issues in reading comprehension*. Hillsdale, NJ: L. Erlbaum Associates.

Mervis, C. B., J. Catlin, and E. Rosch (1976). Relationships among goodness-of-example, category norms and word frequency. *Bulletin of the Psychonomic Society* 7, 268–284.

Murphy, G. L., and D. L. Medin (1985). The role of theories in conceptual coherence. *Psychological Review* 92, 289–316.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115, 39–57.

Osherson, D. N. (1987). New axioms for the contrast model of similarity. *Journal of Mathematical Psychology* 31, 93–103.

Osherson, D. N., E. E. Smith, and E. B. Shafir (1986). Some origins of belief. *Cognition* 24, 197–224.

Pinker, S. (1990). Language acquisition. In D. N. Osherson and H. Lasnik, eds., *Language: An invitation to cognitive science, volume 1*. Cambridge, MA: MIT Press.

Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson, ed., *Language, mind, and knowledge*. Minneapolis, MN: University of Minnesota Press.

Rey, G. (1983). Concepts and stereotypes. *Cognition* 15, 237–262.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Voisniadou and A. Ortony, eds., *Similarity, analogy, and thought*. New York: Cambridge University Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch and B. B. Lloyd, eds., *Cognition and categorization*. Hillsdale, NJ: L. Erlbaum Associates.

Rosch, E., C. Mervis, D. Gray, D. Johnson, and P. Boyes-Braehm (1976). Basic objects in natural categories. *Cognitive Psychology* 3, 382–439.

Schwartz, S. P. (1977). *Naming, necessity and natural kinds*. Ithaca, NY: Cornell University Press.

Schwartz, S. P. (1979). Natural kind terms. *Cognition* 7, 301–315.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* 27, 125–140.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika* 39, 373–421.

Skyrms, B. (1986). *Choice and chance: An introduction to inductive logic*. 3rd ed. Belmont, MA: Wadsworth.

Smith, E. E. (1989). Concepts and induction. In M. I. Posner, ed., *Foundations of cognitive science*. Cambridge, MA: MIT Press.

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

Smith, E. E., D. N. Osherson, L. J. Rips, and M. Keane (1988). Combining prototypes: A modification model. *Cognitive Science* 12, 485–527.

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

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

Tversky, A., and I. Gati (1982). Similarity, separability and the triangular inequality. *Psychological Review* 89, 123–154.

Tversky, A., and J. W. Hutchinson (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review* 93, 3–22.

Zadeh, L. (1982). A note on prototype theory and fuzzy sets. *Cognition* 12, 291–297.