

Editor

Jacques Mehler

Laboratoire de Sciences Cognitives et Psycholinguistique, 54, Boulevard Raspail, 75006 Paris, France

Associate Editors

Thomas G. Bever

Department of Psychology, University of Rochester, Rochester, NY 14627, USA

John Morton

MRC Cognitive Development Unit, 4 Tavistock Street, London WC1H 0BT, UK

Steven Pinker

Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA

Editorial Board

- | | |
|----------------------------------|-------------------------------------|
| U. Frith London, UK | M. Piattelli-Palmarini Milan, Italy |
| A.M. Galaburda Boston, USA | M. Posner Eugene, USA |
| C.R. Gallistel Los Angeles, USA | M.C. Potter Cambridge, USA |
| G. Gigerenzer Chicago, USA | D. Preatoni Paris, France |
| L. Gleitman Philadelphia, USA | Z. Pylyshyn Piscataway, USA |
| G. Haslam Soka, Japan | L. Rips Evanston, USA |
| D.D. Hoffman Irvine, USA | L. Rizzi Geneva, Switzerland |
| F. Johnson-Laird Princeton, USA | A.G. Samuel Stony Brook, USA |
| M.I. Jordan Cambridge, USA | S. Sasayama Tokyo, Japan |
| P.W. Juszczyk Buffalo, USA | S. de Schonen Marseille, France |
| F. Kell Thaca, USA | H. Schacter Berlin, Germany |
| S.M. Kosslyn Cambridge, USA | N. Sebenius Barcelona, Spain |
| B. Landau Irvine, USA | M. Sedlitz Los Angeles, USA |
| S.G. Lapointe Davis, USA | E. Shafir Princeton, USA |
| A.M. Leslie Piscataway, USA | T. Shallice London, UK |
| W. Levelt Nijmegen, Netherlands | D.I. Shoben Berkeley, USA |
| S. McClelland Paris, France | D. Sperber Paris, France |
| C. Marcus Seattle, USA | M. Steedman Philadelphia, USA |
| J. Marshall Oxford, UK | S. Stich New Brunswick, USA |
| W. Marslen-Wilson London, UK | S. Strauss Ramat Aviv, Israel |
| D.L. Medin Evanston, USA | M. Taft Kensington, Australia |
| J.L. Miller Boston, USA | O. Tzeng Riverside, USA |
| J. Morais Brussels, Belgium | Y. Vallan New York, USA |
| D. Kemler Nelson Swarthmore, USA | D. Wilson London, UK |
| D. Osherson Marigny, Switzerland | A.W. Young Cambridge, UK |
| I. Peretz Montreal, Canada | |

Editorial Associate

Susana Franck

Laboratoire de Sciences Cognitives et Psycholinguistique, 54, Boulevard Raspail, 75006 Paris, France

Cognition (ISSN 0010-0277). For 1994, volumes 50-53 are scheduled for publication. Subscription prices are available upon request from the publisher. Single issues are available on a prepaid basis only and are sent by air to subscribers in the following countries where airfreight is necessary: Argentina, Brazil, Canada, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, Singapore, South Korea, Taiwan, Thailand, USA. For all other countries, mail rates are available upon request. Claims for missing issues should only be made if the issue has not yet appeared. Please address all your requests regarding orders and subscription queries to: *Cognition*, Journal 111, 1000 AE Amsterdam, Netherlands, tel. (+31-20)5803642, fax (+31-20)5803396.

For further information on this and other Elsevier journals please contact: Elsevier Science Inc., Journal Information Center, 655 Avenue of the Americas, New York, NY 10010, USA, tel. (212) 6333799, fax (212) 6333794, telex 420443 AEP UL.

Contact the publisher:

Advertising orders and enquiries may be sent to: Elsevier Science B.V., Advertising Department, P.O. Box 211, 1000 AE Amsterdam, Netherlands, tel. (+31-20) 6833901, in the UK: TG Scott & Son Ltd, atin Tim Blake/Vanessa Bird, Portland House, 21 Narborough Rd., Cobly, Leicestershire LE15 8AA, UK, tel. (+44-533)790522. In the USA and Canada: Western Media Associates, atin Daniel Lipner, P.O. Box 1110, Greens Farms, CT 06033, USA, tel. (203)2612500, fax (203)2610101.

Cognition (0010-0277) is published monthly by Elsevier Science B.V. (Moltenweg 1, Postbus 211, 1000 AE Amsterdam). Annual subscription price in the Netherlands is 110.00 (including postage and handling charges). Single issues are available on a prepaid basis only and are sent by air to subscribers in the following countries where airfreight is necessary: Argentina, Brazil, Canada, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, Singapore, South Korea, Taiwan, Thailand, USA. For all other countries, mail rates are available upon request. Claims for missing issues should only be made if the issue has not yet appeared. Please address all your requests regarding orders and subscription queries to: Elsevier Science B.V., Journal Information Center, 655 Avenue of the Americas, New York, NY 10010, USA, tel. (212) 6333799, fax (212) 6333794, telex 420443 AEP UL.

Cognition, 52 (1994) 125-157
0010-0277/94/\$07.00 © 1994 - Elsevier Science B.V. All rights reserved.

125

The role of similarity in categorization: providing a groundwork

Robert L. Goldstone*

Psychology Department, Indiana University, Bloomington, IN 47405, USA

Received May 28, 1993, final version accepted February 2, 1994

Abstract

The relation between similarity and categorization has recently come under scrutiny from several sectors. The issue provides an important inroad to questions about the contributions of high-level thought and lower-level perception in the development of people's concepts. Many psychological models base categorization on similarity, assuming that things belong in the same category because of their similarity. Empirical and in-principle arguments have recently raised objections to this connection, on the grounds that similarity is too unconstrained to provide an explanation of categorization, and similarity is not sufficiently sophisticated to ground most categories. Although these objections have merit, a reassessment of evidence indicates that similarity can be sufficiently constrained and sophisticated to provide at least a partial account of many categories. Principles are discussed for incorporating similarity into theories of category formation.

1. Introduction

The idea that concepts and categorization are grounded by similarity has been quite influential, providing the basis for many models in cognitive psychology.

*E-mail: rgoldsto@ucs.indiana.edu

The author would like to thank Douglas Medin and Dedre Gentner for many useful discussions, Larry Barsalou, Lloyd Komatsu, John Kruschke, Robert Melara, Jim Nairne, Paula Niedenthal, Robert Nosofsky, Steven Sloman, Jim Sherman, Richard Shiffrin, Ed Smith and Linda Smith also contributed helpful comments. Preparation of this article was supported in part by National Science Foundation Grant SBR-9409232.

SSDI 0010-0277(94)00621-0

According to some theories of categorization, an object is categorized as an A and not a B if it is more similar to A's best representation (its "prototype") than it is to B's (Posner & Keele, 1968; Reed, 1972; Rosch & Mervis, 1975). According to exemplar theories, an object is categorized as an A and not a B if it is more similar to the individual items that belong in A than it is to those that belong in B (Brooks, 1978, 1987; Medin & Schaffer, 1978; Nosofsky, 1986, 1992). Although there are important differences between prototype and exemplar models of categorization, they make the common assumption that categorization behavior depends on the similarity between the item to be categorized and the categories' representations.

Recently, a number of empirical and theoretical arguments have undermined the role of similarity in categorization. The arguments have taken several forms: similarity is too flexible and unconstrained to serve as a grounding explanation for categorization, similarity is too perceptually based to provide an account of abstract concepts, and many concepts exist that are organized around goals or theories rather than similarity. Together, these arguments impose a dilemma for researchers who would base categorization on similarity. Researchers can either allow abstract, theory-dependent commonalities to influence similarity, or they can restrict similarity to properties that are available directly to the senses. If abstract similarities are permitted, then similarity is too unconstrained to predict categorization results. If abstract similarities are disallowed, then too wide a gap exists between similarity and the requirements of categorization.

In the first section, these arguments and their associated empirical evidence are examined. In the second section, an assessment of the arguments is given. Finally, in the third section, a framework is proposed in which similarity plays a necessary and important role in grounding categorization.

2. Arguments for the insufficiency of similarity for grounding categorization

The following empirical and theoretical arguments undermine the role of similarity as a ground for categorization. Some of the arguments make the strong claim that similarity is a meaningless or empty notion, and thus cannot serve as a ground for any cognitive function. Others make the weaker claim that there is more to categorization than similarity; an understanding of similarity is not sufficient for an understanding of categorization. At a minimum, all of the arguments claim that similarity insufficiently constrains our categories.

2.1. Similarity is too flexible

Researchers have asserted that similarity is quite unconstrained, offering too mutable a foundation for categorization. If similarity is overly flexible and

context-dependent, then similarity would be in as much need of explanation as categorization.

A natural reply to make is, "But all *logically possible* properties are not psychologically important. Objects do not have an arbitrary number of *psychologically relevant* properties." However, this response begs the question; Goodman (1972) argues that determining whether a property is psychologically important is just as hard a task as determining similarity.

Similarity is a blank to be filled in

Goodman goes on to argue that "X is similar to Y" means nothing until it is completed by "X is similar to Y with respect to property Z." Psychologically important properties are determined by finding the property Z with respect to which X and Y are compared. Goodman argues that "when to the statement that two things are similar we add a specification of the property that they have in common . . . we render it [the similarity statement] superfluous" (p. 445). That is, all of the potential explanatory work is done by the "with respect to property Z" clause, and not by the similarity statement. Instead of saying "This object belongs to Category A because it is similar to A items with respect to the property 'red'," we can simplify matters by removing any notion of similarity with "This object belongs to Category A because it is red."

Empirical research on similarity supports the contention that similarity can change markedly depending on the properties that are implicated as relevant. For example, *raccoon* and *snake* were judged to be less similar when no explicit context was provided than when a context was established by placing the words above the comparison (Barsalou, 1982). Barsalou (1983) went on to show that items that are seemingly quite different (e.g. *children* and *jewelry*) can be rated as highly similar if they are placed in a context ("things to retrieve from a burning house") that highlights a respect on which they are similar. Goldstone, Medin, and Gentner (1991) have shown that shared abstract relations increased similarity more than shared superficial attributes when the overall relational similarity between compared items was high; the opposite trend was found when the overall superficial similarity was high. Thus, whether a particular relation or attribute serves as the primary basis for fixing Z in the "with respect to property Z" clause depends on the other shared properties.

More generally, the properties that are relevant for a similarity comparison vary widely with age (Gentner, 1988), expertise (Sjoberg, 1972), environment (Harnad, 1987), method of presentation (Gati & Tversky, 1984), and even cerebral hemisphere of processing (Umlilt, Bagnara, & Sinton, 1978). In order to have a complete theory of similarity, it would seem that we first must have theories about how various factors influence the importance of different properties. If Goodman is correct, then these additional theories are entirely responsible for all of the explanatory power of similarity.

Similarity is context-dependent

A number of researchers have explicitly manipulated the context of a similarity comparison, and have found wide variation in the resulting similarity assessments. The sum of research indicates that assigning a single unitary estimate of similarity to pairs of items is insufficient.

Suzuki, Ohnishi, and Shigemasa (1992) have shown that similarity judgments depend on level of expertise and goals. Expert and novice subjects were asked to solve the Tower of Hanoi puzzle, and judge the similarity between the goal and various states. Experts' similarity ratings were based on the number of moves required to transform one position to the other. Less expert subjects tended to base their judgments on the number of shared superficial features. Similarly, Hardiman, Dufresne, and Mestre (1989) found that expert and novice physicists evaluate the similarity of physics problems differently, with experts basing similarity judgments more on general principles of physics than on superficial features (see Sjöberg, 1972 for other expert/novice differences in similarity ratings).

Whorf (1941) argues for similarities that depend on cultural context. For Shawnee Native Americans, the sentences "I pull the branch aside" and "I have an extra toe on my foot" are highly similar. Roughly speaking, the first sentence would be represented as "I pull it (something like the branch of a tree) more open or apart where it forks" and the second sentence is represented as "I have an extra toe forking out like a branch from a normal toe." Similarity evidently depends on factors other than the objective perceptual features of the compared objects. Kelly and Keil (1987) show that the similarity between items from two domains (periodicals and food) is altered by presenting subjects with metaphors between the domains. For example, subjects who receive metaphors such as "The New Yorker is the quiche of newspapers and magazines" give greater similarity assessments to cross-domain pairs that have similar values on a *tastefulness* dimension than do subjects who do not receive such metaphors.

Similarity also depends on the surrounding presentation context of an item. Roth and Shoben (1983) presented subjects with a general category term instantiated in two different sentential contexts. For example, subjects saw *beverage* in one of two sentences: "During the midmorning break the two secretaries gossiped as they drank the beverage" or "Before starting his day, the truck driver had the beverage and a donut at the truckstop." In the first context, subjects rate *tea* to be more similar than *milk* to *coffee*. In the second context, there is a greater tendency for subjects to rate *milk* as being more similar than *tea* to *coffee*. Medin and Shoben (1988) extend these results to similarity relations between adjectives that are influenced by their accompanying nouns. They find, for example, that when *hair* is modified by a color adjective, *white* is selected more often than *black* as being similar to *gray*; however, the opposite trend is

found when *cloud* is modified. Likewise, Half, Ortony, and Anderson's (1976) subjects responded that the redness of flags was more similar to the redness of lights than it was to the redness of sunburns, wine, hair, or blood. These studies indicate that nouns and adjectives do not have context-free similarities. In many cases, to know how similar two words are, a context is needed.

In addition to the general context effects that occur when rating scales are used (Parducci, 1965), Krumhansl (1978, also see Nosofsky, 1991; Sjöberg, 1972) argued for specific context effects for similarity ratings. In particular, similarity decreased between objects when they were surrounded by many close neighbors. Sjöberg (1972) and Tversky (1977) found that the same two items were rated as more similar when the set of items covered a broader range. For example, the similarity of *falcon* to *chicken* increased when the entire set of items to be compared included *wasp* rather than *sparrow*.

While the above studies indicate that the context defined by the stimulus set influences similarity, the within-comparison context also influences similarity. According to the diagnosticity effect (Tversky, 1977), how much a feature affects similarity depends on how diagnostic it is for categorization purposes. When choosing the most similar country to Austria from the set {Sweden, Poland, Hungary}, subjects chose Sweden more often than Hungary. In this case, the dimension "form of government" is important because it highlights a difference between one of the choices (Sweden) and the other two choices. When choosing the most similar country to Austria from the set {Sweden, Norway, Hungary}, subjects chose Hungary more often than Sweden, because now the feature "Scandinavian" singles out Hungary from the other two candidates. Medin, Goldstone, and Gentner (1993) obtained another within-comparison context effect whereby antonymically related words markedly altered their similarity. When judged by separate paired similarity ratings, *sunrise* was more similar to *sunbeam* than it was to *sunset*. However, when both *sunbeam* and *sunset* were presented simultaneously, subjects tended to choose *sunset*, rather than *sunbeam*, as most similar to *sunrise*.

Perhaps the most radical suggestion has been that the features that enter into similarity assessments are, themselves, subject to context effects (Asch, 1952). Consistent with this suggestion, Medin et al. (1993) found that ambiguous objects produce mutually inconsistent feature interpretations depending on their comparison. One object was interpreted as possessing three prongs when it was compared to an object that clearly possessed three prongs, and was interpreted as possessing four prongs when compared to a four-prong object.

Similarity is not an unitary phenomenon

There is evidence that various measures of similarity do not converge on a single construct (Goldstone & Medin, in press-a; Medin et al., 1993). Similarity

can be measured by ratings (e.g. on a scale from 1 to 10). Similarity can be assessed by measuring the average time required for subjects to respond that two things are different (Podgorny & Garner, 1979). Hypothetically, the more similar two things are, the longer it will take to say that they are different. Similarity can also be assessed via confusions in an identification task. The more similar two things are, the more often subjects will mistakenly respond that have seen one of things are, the more often subjects will mistakenly respond that have seen one of the things when they have actually seen the other (Getty, Swets, & Green, 1979; Gilmore, Hersh, Caramazza, & Griffin, 1979; Townsend, 1971).

Similarity as measured by ratings is not equivalent to similarity as measured by perceptual discriminability or frequency of perceptual confusion. Although these measures correlate highly, systematic differences are found (Keren & Baggen, 1981; Podgorny & Garner, 1979; Sergeant & Takane, 1987). In general, abstract or conceptual properties are more influential for similarity ratings than they are for the other measures (Torgerson, 1958). For example, Beck (1966) finds that an upright T is rated as more similar to a tilted T than an upright L, but that it is also more likely to be perceptually grouped with the upright Ls.

Given the systematic discrepancies between different measures that hypothetically tap into similarity, a possible conclusion is that similarity is not a coherent notion at all. The term *similarity*, like the terms *bug* or *family values*, may not pick out a consolidated or principled set of things. Similarity, then, is flexible not only because it varies with context, intentions, and characteristics of the comparison-maker, but also because it is calculated by divergent processes in diverse tasks.

Summary

That similarity is often flexible seems unquestionable. The conclusion to draw from this is more controversial. Goodman (1972) argues that the flexibility of similarity is pervasive enough to doom it: "Similarity tends under analysis either to vanish entirely or to require for its explanation just what it purports to explain" (p. 446). The literature cited above clearly indicates that similarity assessments are not based solely on perceptual input. The expertise, intentions, and goals of the comparison-maker also influence judged similarity. Thus, the dangers of an overly flexible similarity construct indicated by Goodman (and also Rips, 1989) must be taken seriously.

Similarity cannot explain categorization if it is dependent on categorization for definition. Goodman argues for such a dependency. In one example, he argues that there is no way to explain why different letter As all belong to the category *Letter A* by similarity unless we claim that "a's are alike in being a's, which presumes exactly the categorization that similarity was supposed to explain" (p. 439). Indurkha (1992) provides several examples of similarities that emerge only

after items are grouped together, arguing at one point that the similarities between cats and fog may only be apparent when they are paired together, as they are in Carl Sandburg's poem (also see Shannon, 1988). More empirically based, Goldstone (in press-b) finds that giving subjects prolonged training on different categorization rules alters the perceptual similarity of items as measured by a task that requires subjects to respond as to whether two displayed items are physically identical.

In sum, similarity seems to be a rather flexible ruler, dependent on several contexts - contexts that are defined by: the individual comparison-maker's goals and knowledge, the items currently being compared, the background set of items, and the method of measuring similarity. Several researchers have argued that similarity is too flexible to provide a solid basis for categorization that does not beg questions. By these claims, explaining categorization by similarity is question-begging if similarity requires sophisticated, knowledge-dependent, and flexible analyses of the items being compared.

2.2. Categorization depends on factors other than similarity

While some researchers have argued that similarity is too affected by non-perceptual factors to provide a productive explanation of categorization, others maintain that similarity is too perceptually constrained to explain our categorization. A good deal of evidence has found dissociations between categorization and similarity assessments, with similarity assessments grounded more in perception, and categorization depending more on a categorizer's theories, goals, culture, and other high-level factors.

Categorization as theory-dependent

People's categorizations seem to depend on the theories they have about the world (for a recent review of the experimental and theoretical support of this claim, see Komatsu, 1992). Theories, although often not clearly defined, involve organized systems of knowledge. In making an argument for the use of theories in categorization, Murphy and Medin (1985) provide the example of a man jumping into a swimming pool fully clothed. This man may be categorized as drunk because we have a theory of behavior and inebriation that explains the man's action. Murphy and Medin argue that the categorization of the man's behavior does not depend on matching the man's features to the category *drunk's* features. It is highly unlikely that the category *drunk* would have such a specific feature as "jumps into pools fully clothed." It is not the similarity between the instance and the category that determines the instance's classification; it is the fact that our category provides a theory that explains the behavior.

Other researchers have empirically supported the dissociation between theory-derived categorization and similarity. In one experiment, Carey (1985) observes that children choose a toy monkey over a worm as being more similar to a human, but that when they are told that humans have spleens, are more likely to infer that the worm has a spleen than that the toy monkey does. Thus, the categorization of objects into "spleen" and "no spleen" groups does not appear to depend on the same knowledge that guides similarity judgements. Carey argues that even young children have a theory of living things. Part of this theory is the notion that living things have self-propelled motion and rich internal organizations. Susan Gelman and her colleagues (Gelman, 1988; Gelman & Markman, 1986) have shown that children as young as three years of age makes inferences about an animal's properties on the basis of its category label even when the label opposes superficial visual similarity.

Using different empirical techniques, Keil (1989) has come to a similar conclusion. In one experiment, children are told a story in which scientists discover that an animal that looks exactly like a racoon actually contains the internal organs of a skunk and has skunk parents and skunk children. With increasing age, children increasingly claim that the animal is a skunk. That is, there is a developmental trend for children to categorize on the basis of theories of heredity and biology rather than visual appearance. In a similar experiment, Rips (1989) shows an explicit dissociation between categorization judgments and similarity judgments in adults. An animal that is transformed (by toxic waste) from a bird into something that looks like an insect is judged by subjects to be more similar to an insect, but is also judged to be a bird still. Again, the category judgment seems to depend on biological, genetic, and historical knowledge, while the similarity judgments seems to depend more on gross visual appearance.

In another experiment, Rips asks subjects to imagine a three-inch-round object, and then asks whether the object is more similar to a quarter or a pizza, and whether the object is more likely to be a pizza or a quarter. There is a tendency for the object to be judged as more similar to the quarter, but as more likely to be a pizza. Our knowledge about the relative variability of pizzas (some variability) and quarters (very little variability) seems to play a larger role in the categorization decision than in the similarity judgment.

Other researchers have tested groups of subjects that have specialized theoretical knowledge, and have detected an influence of these theories on categorization behavior. Chi, Feltovich, and Glaser (1981) gave expert and novice physicists a group of physics problems to sort into categories. The novices grouped the problems on the basis of surface properties such as whether the problems involved springs or inclined planes. The experts grouped the problems on the basis of the deep law of physics required for solution, such as Newton's second law or conservation of energy. Experiments such as this, along with work on cross-cultural cognition (e.g. Lakoff, 1986; Shweder, 1977), argues that the theories

that one develops over a lifetime, and not simply the immediate statistical associations in the world, affect people's categorizations (also see Chapman & Chapman, 1969; Johnson, Mervis, & Boster, 1992).

Other researchers (Ahn, 1991; Ahn, Brewer, & Mooney, 1992; Medin, 1989; Medin, Wattenmaker, & Hampson, 1987; Wattenmaker, Dewey, Murphy, & Medin, 1986; Wisniewski & Medin, in press) manipulate subjects' theoretical knowledge by explicitly giving some subjects an abstract description or background knowledge that accounts for a categorization. The typical empirical result is that categorizations are highly sensitive to the background theories provided. Wisniewski and Medin (in press) give different groups of subjects the same sets of children's drawings with different category labels. For example, some subjects would receive the category labels "drawn by city children"/"drawn by rural children" while others would receive "drawn by gifted children"/"drawn by normal children." The category rules that subjects devise to distinguish the sets of drawings depends on the category labels. For example, subjects are much more likely to mention "unusual and creative perspective" when asked to describe the pictures supposedly drawn by gifted children. In general, categorizations vary widely depending on whether theories are experimentally provided to subjects or not.

Categorization as goal-dependent

While theories are seen as fairly permanent knowledge structures, researchers have also argued that categorization depends on transient goals and perspectives. Barsalou and his colleagues (Barsalou, 1982, 1983, 1987, 1991; Barsalou & Medin, 1986) have argued that categories can be created as needed in order to fulfill a particular goal. While some aspects of a concept may be permanently linked to the concept (e.g. people seem to automatically activate the property "smells" when the category "skunk" is mentioned), other aspects are only activated when they are appropriate for a particular goal/context (Barsalou, 1982). For example, the fact that basketballs float is only considered (primed) in a context such as a shipwreck at sea with a cargo of basketballs. Barsalou (1983) describes "ad hoc" categories that collect together apparently dissimilar members into a single temporary category to meet a goal.

Temporary shifts in perspectives influence category structure. Barsalou (1987) reports evidence where subjects judged the typicality of instances from one of several perspectives (Average American, African, Chinese, businessman, housewife, hippie, etc.). The adopted perspective had large effects on typicality. For instance, swans were rated as more typical birds from a Chinese perspective than they were from an American perspective. Schank, Collins, and Hunter (1986) also argue that categories are determined on the basis of contextual, pragmatic, and goal-related information.

Categorization as dependent on non-local information

Categorization decisions can depend on information that is available at a category-wide level but not at the level of individual instances. For example, Fried and Holyoak (1984; also see Homa & Vosburgh, 1976) show that subjects are sensitive to the variability within categories. Two prototype pictures were constructed for two artists; one artist's pictures deviated substantially from his prototype, while the other artist's pictures showed less variability. Subjects categorized with respect to category variability, such that some pictures that were more similar on average to the low-variability artist's pictures were categorized as being created by the high-variability artist.

Rips (1989) provides another example of the distribution of information within a category influencing categorization. Subjects were given temperature values falling along a bimodal distribution with peaks at 30 and 80 degrees. Temperatures around 55 degrees were judged to be similar to the observed temperatures, but were not judged to be likely members of the presented set (see also Bourne, 1982). Rips again concludes that similarity and categorization judgments are dissociated, with similarity judgments being more sensitive to the central tendency of a category, and categorization judgments being more sensitive to the distribution of instances within the category.

As a final example, Medin et al. (1993) report the results from an experiment in which one group of subjects judges that doberman pinschers are more similar to raccoons than sharks. Other subjects judge doberman pinschers to more likely belong to the category {boar, lion, shark} than the category {boar, lion, raccoon}. Doberman pinschers seem to be placed in the first category because the first category's members are all ferocious, as are doberman pinschers. The feature "ferocious" emerges as an important category-level generalization that guides categorizations, despite the greater overall similarity of doberman pinschers to the other category's items (for a similar experimental result, see Elio & Anderson, 1981). In this example, as with the preceding examples, categorizations are influenced by properties that do not exist at the local level of individual instances, but do exist for the category as a whole. In these cases, categorizations cannot be completely accounted for by the similarity of items to category instances taken individually.

Summary

The sum of the evidence in this section indicates that categorization and similarity are not based on exactly the same information. Categorization appears to be more theory-dependent, more goal-driven, and involves properties that are not obtainable from individual item similarities. The reviewed evidence shows that there are empirically observed dissociations between similarity judgments and categorization decisions. The burden of proof at this point seems to lie with

those who would base categorization on some function of the similarities between the item to be categorized and category members.

3. A re-evaluation of the connection between similarity and categorization

The above arguments can be summed up by: similarity is too flexible to ground categorization ("flexible similarity"); and categorization is too flexible to be grounded by "mere" similarity ("flexible categorization").

The first thing to note is that, to some extent, these arguments weaken each other. Much of the evidence that supports "flexible categorization" is countered by evidence used to support "flexible similarity." Empirical support for "flexible categorization" comes from dissociations between similarity and categorization, with similarity being more perceptually based and categorization being based more on theories, knowledge, goals, and context. However, studies used to support "flexible similarity" have shown this split to be too simple: similarity, like categorization, depends on context, goals, theories, and culture. Conversely, evidence in support of "flexible categorization" argues against Goodman's claim [in his argument for "flexible similarity"] that similarity requires exactly the sort of categorization that it purports to explain. Evidence in support of "flexible categorization" shows that categorization decisions often make use of more sophisticated knowledge than is used to compute similarity. Thus, evidence for "flexible similarity" argues for a "fancy," sophisticated notion of similarity, and evidence for "flexible categorization" argues for a more sensory-based notion of similarity.

The burden of this section is to argue for a compromise between "flexible similarity" and "flexible categorization" that allows similarity to play an important role in explaining/grounding categorization. Arguments against extreme versions of "flexible similarity" and "flexible categorization" will be put forth. It will be argued that similarity is not limited to superficial sensory properties on the one hand, but also does not typically require theories or knowledge as sophisticated as those that are required for the categorization to be explained.

3.1. Constraints on similarity

The conclusion that similarity is too flexible to explain categorization only follows if constraints on similarity are not forthcoming or are no different from the constraints on categorization. Although the reviewed evidence does indicate that similarity depends on culture, goals, and context, there are still grounds for believing in a variety of similarity that is relatively constrained and principled.

The strongest and most obvious constraint on similarity comes from perceptual

processes. It may be true that any two things can share any number of properties in common if we allow strange properties such as "weighs less than 50001 kilograms," but it does not follow that we always need *theories* to determine what properties will be salient. Our perceptual systems also aid in this determination (Murphy & Medin, 1985). Excellent work is currently underway to identify exactly what perceptual aspects are likely to be used for recognition and categorization (Biederman, 1987; Yuille & Ullman, 1990). There is good reason to think that many perceptual similarities are hard-wired (Shepard, 1987). It is difficult *not* to notice the similarity between a 400 Hz tone and a 402 Hz tone, or two shades of red. Jones and Smith (1993; also see Smith & Heise, 1992) argue that previous experiments have underestimated the role that perception plays in inferring novel properties of objects because only perceptually sparse representations have been used in many cases. If richer properties are included, then categorical inductions are related to perceptual similarity in most cases.

Other constraints on similarity come from the comparison process itself. Even when the relevance of properties for a similarity judgment cannot be determined by considering the objects separately, the properties often become relatively fixed when the objects enter into a comparison. For example, similarity comparisons seem to be constrained such that a shared aspect between objects increases similarity more if there are many other shared aspects of a similar sort (Goldstone et al., 1991). An object with ambiguous properties can become disambiguated when it is placed in a comparison (Medin et al., 1993). Other research (Goldstone, in press-a; Goldstone & Medin, in press-a; Markman & Gentner, 1993) provides evidence that the importance of a shared property depends on whether it belongs to aligned parts – parts that are likely to be placed in correspondence with each other. The alignment of one part will depend on the other alignments that are simultaneously being created. Thus, even when perceptual constraints are not sufficient to completely specify the properties that will be considered for an individual object, it may be possible to provide constraints that arise from the interaction between pairs of objects.

Finally, task and stimulus factors may further constrain what properties are considered, and how much they are considered. It is widely agreed that similarity increases as the number of common features between objects increases, and as the number of features possessed by only one object (distinctive features) decreases. Common features, as compared to distinctive features, are given relatively more importance in similarity judgments for verbal as opposed to pictorial stimuli (Gati & Tversky, 1984), for cohesive as opposed to non-cohesive stimuli (Ritov, Gati, & Tversky, 1990), for similarity as opposed to difference judgments (Tversky, 1977), and for entities with a large number of distinctive as opposed to common features (Gati & Tversky, 1984). To take another example, abstract structural features, as opposed to superficial features, have a larger influence on similarity when subjects are given more time to respond (Goldstone

& Medin, in press-a, in press-b; Goldstone et al., 1991), when sparse rather than rich objects are used (Gentner & Rattermann, 1991), and with increasing practice in making judgments (Goldstone et al., 1991; also see Gentner, 1988 for developmental evidence). In short, there seem to be systematic constraints on the importance of properties that come from task and stimulus factors.

From Goodman's philosophical perspective, any two things could be judged to have any degree of similarity. However, if empirically warranted constraints are taken into account, similarity is substantially less haphazard. Furthermore, there are sources of constraints other than perception and theories. Interactions between compared objects, and task and stimulus variables, also serve to resolve uncertainties about the basis of similarity. These latter constraints are important because they illustrate that "*constrained*" is not the same thing as "*fixed and static*." If we know the two things that are being compared and we know about various task variables, then we can say quite a bit about what properties will be important for the similarity comparison. However, much less can be said about the importance of properties if we only know what one of the objects looks like. The context-sensitive nature of property selection is consistent with similarity being governed by systematic constraints (cf. Jones & Smith, 1993).

3.2. *Similarity as the integration of multiple sources of information*

Although the above constraints serve to establish a systematic and fairly stable notion of similarity, Goodman (1972) would still complain that the perceptual constraints, and not similarity proper, are performing the explanatory work. This argument is less compelling when the constraints come from the particular process or task of comparing objects for similarity. In these cases, the operation of determining similarity itself establishes the properties that will be considered and their importance.

However, similarity is explanatory even in situations where the perceptual system determines the importance of properties. A notion of similarity is still needed because comparisons along several properties must be integrated into a single estimate of similarity. The specification of how this integration takes place constrains similarity in ways that are not explained by the perception of individual properties.

Goodman (1972) ignores the integrative nature of similarity by asserting that "to say that two things are similar in having a specific property in common is to say nothing more than they have the property in common" (p. 445). In fact, cognitive psychologists investigating similarity have dedicated the bulk of their efforts to formulating how multiple properties are combined to form an impression of similarity (for a review, see Melara, 1992). According to feature matching models of similarity, the similarity of A to B is a function of three components:

$(A \cap B)$, the features shared by A and B; $(A - B)$, the features possessed by A but not B; and $(B - A)$, the features possessed by B but not A. According to Tversky's (1977) Contrast model, the similarity of A to B is expressed by

$$S(A, B) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

where f is a monotonically increasing function, and θ , α , and β are weighting terms. Thus, similarity increases as A and B share more features, and decreases as they possess more distinctive features. In multidimensional scaling approaches (Ashby, 1992; Carroll & Wish, 1974; Nosofsky, 1992), similarity is conceived as inversely related to $D_{i,j}$, the distance between objects i and j in a geometric space. Distance is defined as

$$D_{i,j} = \left[\sum_{k=1}^n |X_{ik} - X_{jk}|^r \right]^{1/r}$$

where n is the number of dimensions, X_{ik} is the value of item i on dimension k , and r is a parameter that allows different spatial metrics to be used (if $r = 1$ then the distance between items is equal to the sum of their dimensional differences, if $r = 2$ then the distance is the length of shortest line that connects the items). The value of r that best fits human similarity assessments depends on the stimuli (Garner, 1974) and subjects' strategies (Melara, Marks, & Lesko, 1992). Stimuli that are composed of dimensions that are psychologically fused together (such as the saturation and brightness of a color) or have very small value differences (Nosofsky, 1987), are often best modeled by setting r equal to 2. Stimuli that are composed of separable dimensions (such as the size and brightness of an object) are often best modeled by setting r equal to 1.

The issue of determining an appropriate value for r raises the more general point - that the form that the integration function takes is non-arbitrary and has important consequences. As Goodman notes, we may always flesh out the statement "X is similar to Y" with the clause "with respect to Z," but "Z" may include many properties, and each of the properties may be quite broad. If we say "The large red square is similar to the large red pentagon with respect to visual appearance," several properties (color, size, and shape) must all be considered and integrated. If any property changes substantially, the statement's credentials diminish. Counter to Goodman, it is not the "with respect to visual appearance" clause that is doing all of the explanatory work; it is also the specific manner in which the components of visual appearance are integrated.

Evidence from children's perception of similarity suggests that children are particularly prone to combine multiple sources of information when determining similarity. Even dimensions that are perceptually separable are treated as fused in similarity judgments (Smith, 1983; Smith & Kemler, 1978). Children under five years of age tend to classify on the basis of overall similarity and not on the basis

of a single criterial attribute (Keil, 1989; Smith, 1989a). Children often have great difficulty in identifying the dimension along which two objects vary even though they can easily identify that the objects are different in some way (Kemler, 1983). Smith (1989b) argues that it is relatively difficult for young children to say whether two objects are identical on a particular property, but relatively easy for them to say whether they are similar across many dimensions. Thus, we have an early bias not to treat the "with respect to Z" term as severely restricting the properties considered.

There is evidence that adults also often integrate many respects into their similarity judgments. Ward (1983) finds that subjects that tend to sort objects quickly into piles based on similarity tend to sort objects like children, by considering overall similarity across all dimensions instead of maximal similarity on one dimension¹. Likewise, Smith and Kemler Nelson (1984) find that subjects who are given a distracting task produce more overall similarity judgments than subjects who are not. Several of the cited researchers have argued for a *primitive* similarity computation that is used when cognitive resources are limited due to age, level of intelligence (Ward, Stagner, Scott, & Marcus-Mendoza, 1989), distraction, or speed. The essential characteristic of this primitive similarity computation is that it considers a broad number of properties simultaneously. In opposition to Goodman, our most basic similarity computation appears *not* to be one of determining identity on a particular dimension. It appears to be one of determining proximity across many dimensions.

The mandatory perception of similarity

The previously reviewed evidence indicates that similarity depends on goals, knowledge, perspective, and culture. This evidence is consistent with the automatic or mandatory perception of a more-or-less "generic" similarity. Although similarity displays flexibility, it also may possess a stable core that is relatively context-independent. Weak evidence for this comes from the high correlations between measures of similarity that involve quite different procedures (e.g. similarity ratings versus confusion errors in identification).

Stronger evidence for the obligatory use of context-independent similarity comes from situations where subjects are influenced by similarity despite the subjects' intentions and the task's demand characteristics. Allen and Brooks (1991) provide such evidence. Subjects were given an easy rule for categorizing cartoon animals into two groups, such as "if the animal has at least two of the features {long legs, angular body, spots} then it is a builder; otherwise it is a

¹ There is some controversy, however, about whether object classification tasks that have been assumed to measure overall similarity responses (Kemler, 1984; Kemler Nelson, 1989) could also be explained by single-dimension, analytic processes that are obscured by averaging over individual subjects (Ward, 1989; Ward & Scott, 1987).

digger." Subjects were trained repeatedly on eight animals. Then subjects were transferred to new animals. Some of the animals looked very similar to one of the eight training stimuli, but belonged in a different category. These animals were categorized more slowly and less accurately than animals that were equally similar to an old animal and belonged in the same category as the old animal. Subjects seem not to have been able to ignore similarities between old and new animals, even though they knew a fairly straightforward and perfectly accurate categorization rule.

Same/different experiments provide further evidence that similarities cannot be ignored when irrelevant for a task. Eggeth (1966) showed that the speed with which subjects decided that two stimuli are the same with respect to a particular property is greater if they are also the same with respect to irrelevant properties. This generalization holds even when the stimulus is composed of perceptually separable properties such as size and color. Thus, at least early in processing, overall similarity seems to be necessarily processed.

Gentner and Toupin (1986), and Ross (1987) find converging evidence that people use superficial similarities when solving tasks in which these similarities are irrelevant. For example, in solving a word problem, subjects are highly influenced by the previous solution of a problem if it involves the same superficial "cover story" (e.g. both problems involve golf). Subjects are frequently misled into using inappropriate formulas for solving problems because they are reminded of a problem that is only superficially similar. Sadler and Shoben (1993) have recently shown that people's similarity ratings are influenced by features that are not relevant for a particular task-defined context but are relevant for a generic context. When subjects rate the similarity of occupations in a generic context (without additional instructions), two of the most important dimensions that determine similarity concern whether the occupation is outdoor/indoor and whether it is mental/manual. When subjects are asked to rate the similarity of occupations from the perspective of an IRS auditor who is trying to determine whether a given occupation is likely to be involved in tax fraud, the two most important dimensions are "likelihood of committing tax fraud" and, once again, mental/manual skill. Even though subjects are told to base their similarity judgments on a particular dimension, this particular dimension does not completely account for the subjects' similarity ratings. The generic dimension of mental/manual skill intrudes on similarity judgments that should be evaluated solely from the IRS auditor's perspective.

Barsalou (1982) has also obtained evidence for context-independent similarities. As reviewed earlier, subjects give different similarity ratings to some pairs of items (e.g. *slaves* and *jewelry*), depending on whether they are accompanied by an ad hoc category label (e.g. *plunder taken by conquerors*). However, items from common, familiar categories do not change their similarity ratings when their category label is presented. For example, the similarity between *robin* and

eagle is not altered by the presence of the word *bird*. Similarly, Barsalou and Ross (1986) found that subjects clustered items into the same familiar category even when the items were experimentally distributed across diverse ad hoc categories. Subjects' frequency estimates showed that they registered, for example, that *robin* belongs to the same group as *sparrow* even though the materials and instructions led subjects to code *robin* as an object that was red.

At a more perceptual level, several researchers have argued that some featural similarities cannot be ignored despite their hindrance. In the classic demonstration of Stroop interference (Stroop, 1935), subjects are slower to name the color of a word's ink if the word is the name of a conflicting color than if it is a neutral word. In Garner interference (Garner, 1978; Pommerantz, 1986), variation on an irrelevant dimension slows responses to a relevant dimension. In both cases, featural similarities that subjects know to be irrelevant still influence task performance. Likewise, in categorization tasks, features that are, by themselves, nondiagnostic about the category that an item belongs to, still exert an influence on categorization judgments (Brooks, 1978; Goldstone, 1991).

Cross-cultural evidence indicates a strong non-normative use of similarity. According to the principle of homeopathy, causes and effects tend to be similar (Frazer, 1959; also see Wattenmaker, Nakamura, & Medin, 1988). For example, the Azande culture uses the burnt skull of the red bush monkey to cure epilepsy, apparently because the monkeys exhibit seizure-like stretches. Lest we believe ourselves to be immune to such biases, Shweder (1977) argues that Americans perceive a relationship between leadership and self-esteem because of their conceptual similarity, despite the empirically non-existent correlation between the two variables. More generally, Kahneman and Tversky's (1982) argue that people assess "the probability of an uncertain event, or a sample, by the degree to which it is . . . similar in essential properties to its parent population" (p. 33). These examples do not necessarily show that similarity is perceived in a mandatory fashion, but they do show that similarity is used in inappropriate situations.

Summary

The principle conclusion from the last three sections is that similarity is more constrained than is expected by a strong version of the "Similarity is too flexible" argument. Constraints on similarity come from perception, task characteristics, and context. Further constraints come from the manner in which multiple sources of information are integrated into a single similarity estimate. There also appears to be a relatively context-independent similarity that sometimes intrudes on tasks automatically and inappropriately. This similarity is context-independent in the sense that even though a task would be best accomplished with a specialized similarity computation, a more general, overall similarity assessment is used instead. This is not to say that similarity does not vary at all as a function of task,

a position cast into doubt by research cited earlier. The claim is that tasks are partially influenced by general purpose and untailored similarities.

3.3. *Sophisticated properties of similarity*

The second argument against similarity's use in categorization was that categorization is too rich, flexible, and sophisticated to be grounded in similarity. One line of evidence against the strong version of this thesis comes from the previous section: there are occasions (Allen & Brooks, 1991; Smith & Sloman, submitted) when categorization judgments use overall similarity, even when the correct categorization rule is known. Thus, categorization judgments may not always be very sophisticated and flexible. Quine (1977) argues that as science develops (and people mature), categories become increasingly dissociated from "primitive" similarity, but the empirical evidence indicates that adults have not completely discarded similarity as a categorization principle, at least at this point in human evolution.

The other line of evidence that serves to narrow the gap between similarity and categorization concerns the sophistication of perceptual similarity. People seem to be influenced by abstract similarities even when given perceptual tasks. For example, work by Melara and Marks (Melara, 1989; Melara & Marks, 1989; Marks, 1987) shows that people perceive the correspondences between color and pitch, size and loudness, and pitch and position automatically. Stroop interference exists between these dimensions. Thus, subjects who are supposed to respond that a color is "white" are slower if there is a simultaneous low pitch than if there is a high pitch. The following correspondences have been found: white = high, black = low, big = loud, small = soft, high pitch = high spatial position, and low pitch = low spatial position. The interfering effect of incongruent dimension values suggests that subjects automatically perceive dimensional correspondences. They cannot help but to perceive the correspondence even when it impairs their performance. Smith and Sera (1992) have recently found that even subjects as young as 2 years of age perceive a natural correspondence between large sizes and loud sounds. Thus, some cross-dimensional similarities seem to be primitive in that they appear early in development and are difficult for adults to ignore.

Similarity also seems to depend on relational properties, and not simply isolated stimulus attributes. Researchers have shown that subjects will respond to such similarities as: "in both scenes, the left object is larger than the right object," "in both scenes, one thing is providing nutrients to another," and "both scenes have one object surrounded by two identical objects" (Goldstone, Gentner, & Medin, 1989; Goldstone et al., 1991; Markman & Gentner, 1993; Medin, Goldstone, & Gentner, 1990). Sensitivity to these abstract relations are found even for speeded similarity judgments and perceptual same/different

judgments. Similarity and more abstract analogical reasoning seem to have important commonalities, and are in fact hard to distinguish at times (Gentner, 1983, 1989).

Other researchers have stressed the importance of structural descriptions, as opposed to simple feature lists, for similarity. Palmer (1978; also see Hock, Tromley, & Polmann, 1988; Palmer, 1977) argues against the hypothesis that similarity is based on lines/points treated as independent structural units. Line figures with similar high-level structures are found to be more similar than figures with different high-level structures, holding constant the line/point similarity. Similarity is measured by similarity ratings, discrimination errors, and discrimination response times. Some of the particular high-level structures implicated are closedness (whether a closed figure is present in the figure) and connectedness (whether all line segments of a figure are connected to each other).

A reassessment of experimental dissociations between categorization and similarity

The above evidence argues that similarity is not limited to simple attributes of the sort that might be available from feature detectors. Such a limitation seems far too strong, not even explaining Beck's (1966) results that a T is rated as more similar to a tilted T than to an L even though the T is harder to distinguish from the L. More sophisticated perceptual and conceptual aspects clearly influence similarity. However, previously cited evidence does suggest that similarity and categorization are dissociated; similarity and categorization are unequally influenced by various factors. The force of this empirical evidence cannot be dismissed completely, but the conclusions can be somewhat tempered.

For example, in the experiments by Rips (1989) and Keil (1989), subjects judged an animal to be more similar to one species, but more likely to belong to another species. However, one might argue that these experiments have not necessarily tapped into subjects' assessments of similarity. One might consider alternative ways of probing subjects' appraisal of similarity in a Rips-like experiment:

- (1) "Which species (insect or bird) does this animal look more like?"
- (2) "Which species is this animal more to similar to?"
- (3) "Which species is this animal *really* more like?"
- (4) "Which species is this animal more like, taking into consideration *all* of the information that you have available?"
- (5) "Which species is this animal more likely to belong to?"

These questions are ordered roughly on a continuum between perceptually-driven and conceptually-based similarity. Certainly, by Question 5, most people

would prefer to call the judgment an inductive inference or categorization and not a similarity judgment. Rips and Keil essentially use variations of Question 2. However, there is no *a priori* reason to think that Question 2 reveals "true similarity." The fact that Question 2 contains the word "similar" does not guarantee that it provides evidence about what psychologists refer to as "similarity." In Rips' and Keil's experiments, there may very well be a strong task demand to interpret "similar" as "visually similar," but it is doubtful that the subjects, in their everyday life, only adopt a similarity measure tapped by Question 2. In an informal classroom experiment conducted using Rips' materials but probing similarity via Question 4, I find that similarity and categorization judgements largely correlate with each other. That is, if subjects are probed for their similarity assessments in a manner that stresses *overall* similarity as opposed to simple visual similarity, then the observed dissociation between categorization and similarity that Rips observed is no longer found.

Similar considerations can be raised with respect to apparent dissociations between similarity and categorical induction (Carey, 1985; Gelman & Markman, 1986). For example, Carey finds that even adult subjects judge mechanical monkeys to be more similar to people than are fish, worms, or bugs, yet adults and young children are much more likely to make inferences from people to fish/worms/bugs than from people to mechanical monkeys across a wide range of biological properties (e.g. sleeps, eats, has babies, and has bones). However, it might well be argued that adults, in some important sense, do *not* think that mechanical monkeys are more similar to people than are fish. Once again, the instructions to "rank the similarity of pairs of objects" may not be the best way to measure conceptual or abstract similarity. In an earlier section, evidence was reviewed that indicated that similarity is not completely a unitary construct. Different measures of similarity do not converge on exactly the same notion. Even though Rips and Carey use very reasonable methods for obtaining similarity data, if similarity is not a unitary notion, then even the best measure of similarity is in danger of disregarding some aspects of similarity.

Thus, it may be premature to argue from the results of one measure of similarity to a general dissociation between similarity and categorization. The current argument, taken together with Jones and Smith's (1993) observation that the salience of perceptual similarities may be underestimated in several studies because overly sparse materials are used, suggests that the dissociation between similarity and categorization may have been exaggerated.

Implicit sensitivity to category-level information from similarity-based processes

A final point to raise is that similarity may provide a sufficient basis for grounding categorization even when categorization depends on information that appears not to be available from "local" pairwise comparisons. For example,

Fried and Holyoak (1984) find that subjects will tend to place an item into the category with members of lower average similarity to the item if the category has sufficiently greater variability than the alternative category. Although this type of result would appear to require the postulation of a category-level property such as category variability, similarity-based models can predict this result (Medin, 1986). These models simply measure the similarity of the item to be categorized to each of the members of the possible categories, but they give particular weight to members that are highly similar to the item in question (also see Nosofsky, 1986). Even though the *average* similarity of an item to a category is low, if the category has highly variable instances, then there is a good chance that one of the instances will be close to the item. In the extreme, we could imagine a categorization procedure that placed an item into whatever category contains the instance that is most similar to the item. With this similarity-based model, *average* similarity would not influence categorization at all, and there would be a strong tendency to place items into the category with greater variability.

Thus, models that only base categorization on item-to-item similarity can still show sensitivity to category-level information such as category variability. Sensitivity to other category-level information is obtained by the inclusion of selective attention to particular dimensions, a characteristic of many similarity-based models of categorization (Medin & Schaffer, 1978; Nosofsky, 1986, 1992; Kruschke, 1992). Because of the ability of processes that use only similarity to mimic sensitivity to category-level information (e.g. category variability, correlations between features within a category, feature diagnosticity, category rule), we cannot indiscriminately use evidence of sensitivity to this sort of information to exclude models that only use item-to-item similarity (cf. Nosofsky, Clark, & Shin, 1989; Smith & Medin, 1984; Wattenmaker, 1993).

Summary

This section has reviewed arguments in favor of the position that similarity is often quite sophisticated, and consequently, that similarity may well have sufficient power to ground many categorizations. Similarity is sophisticated in the following senses: it is sensitive to the relational structure of the compared items; under natural circumstances, perceptual properties can provide a rich source of information for categorization; and item-by-item similarities, when integrated properly, can mimic sensitivity to at least some category-level information. Similarity, conceived as raw perceptual overlap, may not be a promising candidate for grounding categorization. However, if similarity is expanded to incorporate perceptual relations between object parts, selective attention, rich multisensory perceptual inputs, and conceptual as well as perceptual qualities, then many of the earlier objections for the insufficiency of similarity as a ground for categorization lose their force.

4. Developing a role for similarity in categorization

In the last section, arguments were reviewed that suggest that similarity is both constrained and sophisticated enough to provide a potential ground/explanation for many categorizations. The current section pursues some specific proposals for developing a role for similarity in explaining categorization.

4.1. Building categories from lower-level similarities

If similarity is to provide a ground for categorization, it must successfully navigate between the Scylla of a purely perceptual basis and the Charybdis of an unconstrained set of postulated aspects. Part of the solution is that similarity may depend on sophisticated aspects, but that these aspects might still not be as sophisticated as the ones that will eventually come to characterize the category. There exists a continuum from low-level perceptual feature detectors to highly abstract theories. Explanatory progress occurs when concepts at more abstract levels are explained, in part, by concepts at lower levels.

As an example, consider the concept *dog*. Whatever the features are that allow us to view two dogs as similar, they seem to be less sophisticated than the elaborate “theory” that we have about dogs. Our *dog* theory includes notions involving genes, cellular organization, dog psychology (e.g., “dogs often refuse to bring a fetched stick *all* the way back to the thrower”), and stories about heroic dogs. On the other hand, what determines a poodle’s similarity to other dogs is often less elaborate, involving features like tail length, fur color, size, and the spatial organization of its limbs. This information may be sufficient to group dogs together in a common category. Once this category has been created, further abstract commonalities can be discovered.² Scientists investigated the genetic similarity of dogs *because* of their more superficial similarities. Even if poodles and paramecia were genetically quite similar, it would take scientists a fairly long time to discover this fact, because their apparent dissimilarities impede even considering the comparison.

Other examples of low-level similarities providing a catalyst for developing more elaborate theories for categories come from scientific development. For example, the concept of a log-log linear law of learning (Newell & Rosenbloom, 1981) was developed to explain the relation between practice and response time. The motivation for the development of this law came from apparent similarities between learning curves in many domains. Before the category *log-log linear*

learning curve could be constructed, it was necessary for researchers to see various manifestations of the concept as similar. Again, this initial similarity could be purely visual, or it could be theory-based. But, if the similarity is theory-based, then it is not based on the theory $RT(T+1) = A + B(RT(T)^C)$. After the category is invented, curves may appear even more similar because they instantiate this equation, but the original noticing of similarities between curves was prior to this mathematical law. Scientific theories provide excellent cases of abstract concepts that are also coherent. Even for scientific theories, a strong case can be made that the original grounding/motivation for the category is based on perceptual (or at least lower-level) similarity.

The current argument is that new concepts are suggested by previously developed concepts. Previously developed concepts provide properties that serve as a basis for similarity. Similarity, in turn, provides a heuristic for developing new concepts. Once a new concept is developed, more sophisticated commonalities between the concept’s members are likely to be discovered.

Evidence that categories are sometimes developed before people know the theoretical basis for the category supports this view. Brooks (1978, 1987) showed that people categorize according to abstract grammars before they learn the theory behind the grammar. They can do this by comparing the overall similarity of test items to category exemplars. Other researchers have argued that categorization can occur without a full theory ever developing. According to Medin and Ortony’s (1989) notion of “psychological essentialism,” people act as if there are necessary and sufficient features that define categories even though people may not know these criterial features. They argue that people assume that objects that are superficially similar have deeper “essences” in common as well, and that these essences are responsible for the superficial appearances of the objects. Even when people are unable to define what the underlying essence of a category is, they assume that it has one. Similar to the current claim, Medin and Ortony argue that “surface similarity... [serves]... as a good heuristic for where to look for deeper properties” (p. 182).

Category bootstrapping by similarity

Several mechanisms for creating abstract categories out of simple similarities have received some empirical confirmation. Markman and Gentner (1993) show that the very act of making a similarity comparison promotes a deeper analysis of the compared entities. Rescorla and Furrow (1977) show that associations between events are easier to acquire if the events are similar to each other. If abstractly similar word problems also have similar superficial cover stories, their abstract similarity is more likely to be noticed (Gentner, Rattemann, & Forbus, 1993; Ross, 1987, 1989). Several researchers have found that analogous problems are more likely to be accessed when trying to solve a problem if the analogous

² This is not to deny that the complementary mechanism, categorizations creating similarities, plays an equally important role. Once objects are placed in a category, because of a theory or instruction, similarities between the objects may be changed (Barsalou, 1991; Goldstone, in press-b).

problems also have superficial resemblances to the unsolved problems (Gentner, 1989; Gentner & Toupin, 1986; Holyoak & Koh, 1987; Ross, 1984).

Once superficial similarity prompts two things to be compared, abstract information from one thing can be “carried over” or applied to the other thing (Gentner, 1989). Furthermore, recent evidence indicates that similarity-based reminders can promote category generalizations by highlighting common aspects (Ross, Perkins, & Tenpenny, 1990). When subjects are spontaneously reminded of a previous category member when shown a new member, they tend to create a generalization for the category that fits both members. Both “carry over” and reminding-based generalization mechanisms generate novel abstract generalizations that are triggered by similarity.

Relatively simple perceptual properties also can serve to bootstrap the development of more sophisticated properties. For example, Spelke (1990) argues that among the first principles that an infant uses to break a scene into objects are: “assume objects move as wholes” and “assume objects move independently of one another.” Interestingly, other developmental evidence suggests that the perception of movement is also instrumental in acquiring the distinction between living things and man-made objects (Gelman, 1990), a highly theory-bound distinction (Keil, 1989). However, Spelke’s evidence suggests that the original inspiration for the distinction probably has a much more perceptually grounded basis.

In sum, there is much evidence that people form categories before they have developed full theories for the categories. In fact, it is the act of grouping items together in a category, on the basis of lower-level similarities, that promotes the later discovery of higher-level theories.

4.2. Similarity and category-based induction

Just as similarity is not a unitary concept, there is good reason to think that the term “category” covers disparate notions. Similarity does not provide an explanatory ground for some types of categories, but that it does ground others. Furthermore, the categories that are grounded by similarity represent an important subclass, because of their primary role in inference-making.

Types of categories

Categories can be arranged roughly in order of their grounding by similarity: natural kinds (*dog* and *oak tree*), man-made artifacts (*hammer*, *airplane*, and *chair*), ad hoc categories (*things to take out of a burning house*), and abstract schemas or metaphors (e.g., *events in which a kind action is repaid with cruelty*, *metaphorical prisons*, and *problems that are solved by breaking a large force into*

units that converge on a target). For the latter categories, explanations by similarity are mostly vacuous. An unrewarding job and a relationship that cannot be ended may both be metaphorical prisons, but this categorization is not established by overall similarity. The situations may seem similar in that both conjure up a feeling of being trapped, but this feature is highly specific, and is almost as abstract as the category to be explained.

On the other hand, overall similarity is a useful ground for many natural things and several artifacts. In a series of studies, Rosch (Rosch, 1975; Rosch & Mervis, 1975) has shown that the members of such “basic level” categories as *chair*, *trout*, *bus*, *apple*, *saw*, and *guitar* are characterized by high within-category overall similarity. Subjects listed features for these categories, and for broader superordinate (i.e. furniture) and narrower subordinate (i.e. kitchen chair) categories. An index of within-category similarity was obtained by tallying the number of features common to items in the same category. Items within a basic-level category tend to have several features in common, in contrast to the members of categories such as *metaphorical prisons*.

Rosch (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) argues that categories are defined by family resemblance; category members need not all share a definitional feature, but they tend to have several features in common. Furthermore, Rosch argues that people’s basic level categories preserve the intrinsic correlational structure of the world. All feature combinations are not equally likely. For example, in the animal kingdom, flying is correlated with laying eggs and possessing a beak. There are “clumps” of features that tend to occur together. Some categories do not conform to these clumps (e.g. ad hoc categories), but many of our most natural-seeming categories do.

As Rosch et al. (1976) note, this view does not entail that natural categories are objectively present in the world. Determination of the features to be correlated depends on our perceptual/cognitive apparatus. At the same time, Rosch et al.’s experiments indicate a large observer-independent structuring component for the categories they tested.³ In fact, their Experiment 3 deserves special note as being one of the few experiments in the history of human cognitive psychology that involves no human subjects at all. Silhouette outlines were created from randomly selected and canonically positioned photographs of objects from different categories. Category members belonging to the same basic level category had significantly greater objective similarity (defined by amount of overlapping physical area) than members belonging only to the same superordinate category. Rosch et al. also show by several converging experiments that basic level categories are psychologically “privileged” in that they are first accessed, first learned, most quickly confirmed, and most efficiently represented.

³ More accurately, their experiment indicates that some categories are structured by physical properties that are probably perceivable by almost all people.

These results should not be taken to imply that our basic level categories are defined completely by the world. The history of domestication is as relevant to our concept of *dog* as is general shape. However, the results do show that relatively atheoretic, objectively determined similarities can provide excellent cues to category membership for at least some categories. Experiments by Tanaka and Taylor (1991) argue against the stronger claim that what level is privileged is established completely by objective, observer-independent criteria. Expert bird-watchers and dog handlers were asked to make speeded categorizations of dog and bird photographs at subordinate, basic, and superordinate levels. Subordinate level categorizations were made as quickly as basic level categorizations for experts in a field, whereas the typical basic level advantage was found for novices. These results, although they illustrate an influence of observer characteristics on categorization, are consistent with the view that many natural basic level categories including *bird* and *dog* have similarities that are perceived by virtually all adults (cf. Boster, 1986; Boster & Johnson, 1989).⁴ Consistent with this latter view, Tanaka and Taylor did not find that expertise results in reliably faster categorization at the subordinate than the basic level (the basic level is still among the most privileged levels), and their results, with a few exceptions, showed that the effect of expertise was generally to speed subordinate responding rather than slow basic level responding.

The inductive potential of different categories

Categories that are not structured by similarity allow few inductive generalizations to be made. For example, if we know that an object belongs in Barsalou's ad hoc category of *things to take from a burning house*, we do not know much more about the object. We may suspect that the object is valuable and portable, but not much else can be inferred. Metaphorical concepts and abstract schemes also permit relatively few inductions. Barsalou (1993) argues that ad hoc categories show our ability to organize the world in unusual ways in order to satisfy temporary and context-specific goals. To borrow an example from Barsalou (1991), we create the category *things that can be stood on to change a lightbulb* only when it is needed. The fact that *chair* is a member of this category does not occur to us until we have the goal of changing a lightbulb.

On the other hand, there are concepts that permit many inductive inferences. If we know something belongs to the category *dog*, then we know that it probably has four legs and two eyes, cats dog food, is somebody's pet, pants, barks, is bigger than a breadbox, and so on. Generally, natural kind objects, particular-

⁴ It is interesting to note that several computational object recognition theories (e.g. Biederman, 1987) explicitly attempt to model *only* basic level categorization. The researchers argue that basic level categorization can be accomplished using objective physical features, and does not undergo significant experiential change.

ly those at Rosch's basic level, permit many inferences. Basic level categories allow many inductions because their members share similarities across *many dimensions/features*.

Thus, there is a sense in which some important categories are relatively context-free. Barsalou's (1987) warning that the structure and extension of all concepts may depend on goals and perspective must be taken seriously. Still, some categories are fairly stable with respect to changes in context precisely because there are a large number of converging features that indicate the same categorization. We may be able to select contexts that alter categorization (e.g. Coho salmon and Atlantic salmon are strikingly similar, but only the Atlantic salmon belongs in the group *things found on the coast of Northern Europe*), but most contexts preserve basic level categorizations.

5. Conclusion

It has been argued that overall similarity can provide a useful ground for an important subset of categories. Similarity is neither too unconstrained to provide a firm base for categories, nor too simple-minded to account for rich natural categories.

Similarity is constrained by our perceptual system and by the process for integrating multiple sources of information. Evidence that similarity assessments are often influenced by properties that are irrelevant or even counterproductive to a task indicates that similarity is not completely context-sensitive or task-specific. Similarity is not too simple-minded if rich perceptual stimuli are used and if sophisticated perceptual features are admitted into the calculation of similarity. Furthermore, low-level similarities can bootstrap categories that will evolve deep commonalities.

We have also seen grounds for caution in propounding similarity's role in categorization. Neither *similarity* nor *category* is a unitary construct – there are variations of each that are importantly different. Similarity cannot ground all category types. Still, the class of categories for which overall similarity provides a partial account are an important class because of their wide inductive potential (Smith, Shafir, & Osherson, 1993). The fact that similarity integrates multiple sources of information is an important part of many natural categories' ability to provide useful inferences across many contexts.

The conclusions drawn here are compatible with many of the researchers who conclude that similarity is not sufficient to ground all categorization. However, given the multitude of articles suggesting fundamental problems for similarity's role in categorization, it is easy to incorrectly conclude that similarity provides no role, or only a secondary role. The aim of this review has been to suggest that similarity, despite the real and perceived objections to its use, does play an

important role in establishing many of our categories. Similarity may not necessarily be sufficient for categorization, but similarity is sufficiently necessary to categorization to merit a reassessment of its role.

References

- Ahn, W. (1991). Effects of background knowledge on family resemblance sorting: Part II. *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 203–208). Hillsdale, NJ: Erlbaum.
- Ahn, W., Brewer, W., & Mooney, R. (1992). Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 391–412.
- Allen, S.W., & Brooks, L.R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*, 3–19.
- Asch, S.E. (1952). *Social psychology*. New York: Prentice-Hall.
- Ashby, F.G. (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.
- Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*, 82–93.
- Barsalou, L.W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.
- Barsalou, L.W. (1987). The instability of grade structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development* (pp. 101–140). New York: Cambridge University Press.
- Barsalou, L.W. (1991). Deriving categories to achieve goals. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, (Vol. 27). New York: Academic Press.
- Barsalou, L.W. (1993). Structure, flexibility, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A.C. Collins, S.E. Gathercole, & M.A. Conway (Eds.), *Theories of memory* (pp. 29–101). London: Erlbaum.
- Barsalou, L.W., & Medin, D.M. (1986). Concepts: Static definitions or context-dependent representations? *Cahiers de Psychologie Cognitive*, *6*, 187–202.
- Barsalou, L.W., & Ross, B.H. (1986). The roles of automatic and strategic processing in sensitivity to superordinate and property frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *1*, 116–134.
- Beck, J. (1966). Effect of orientation and of shape similarity on perceptual grouping. *Perception and Psychophysics*, *1*, 300–302.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.
- Boster, J.S. (1986). Agreement between biological classification systems is not dependent on cultural transmission. *American Anthropologist*, *89*, 914–920.
- Boster, J.S., & Johnson, J.C. (1989). Form or function: A comparison of expert and novice judgments of similarities among fish. *American Anthropologist*, *91*, 866–889.
- Bourne, L.E. (1982). Typically effects in logically defined categories. *Memory & Cognition*, *10*, 3–9.
- Brooks, L.R. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.
- Brooks, L.R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: The ecological and intellectual factors in categorization* (pp. 141–174). Cambridge: Cambridge University Press.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.
- Carroll, J.D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D.H. Krantz, R.C. Atkinson, R.D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 57–105). San Francisco: Freeman.
- Chapman, L.J., & Chapman, J.P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, *74*, 272–280.
- Chi, M.T.H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.
- Egeth, H.E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, *1*, 245–252.
- Elio, R., & Anderson, J.R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning & Memory*, *7*, 397–417.
- Frazier, J.G. (1959). *The new golden bough*. New York: Criterion Books.
- Fried, L.S., & Holyoak, K.J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 234–257.
- Garner, W.R. (1974). *The processing of information and structure*. New York: Wiley.
- Garner, W.R. (1978). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General*, *107*, 287–308.
- Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology*, *16*, 341–370.
- Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate–inanimate distinction as examples. *Cognitive Science*, *14*, 79–106.
- Gelman, S.A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, *20*, 65–95.
- Gelman, S.A., & Markman, E.M. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.
- Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, *59*, 47–59.
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity, analogy, and thought* (pp. 199–241). New York: Cambridge University Press.
- Gentner, D., & Rattermann, M.J. (1991). Language and the career of similarity. In S.A. Gelman & J.P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). London: Cambridge University Press.
- Gentner, D., Rattermann, M.J., & Fortou, K.D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, *25*, 524–575.
- Gentner, D., & Toupin, C. (1986). Symmetrality and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), 277–300.
- Gety, D.J., Swets, J.A., Swets, J.B., & Green, D.M. (1979). On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics*, *26*, 1–19.
- Gilmore, G.C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics*, *25*, 425–431.
- Goldstone, R.L. (1991). Feature diagnosticity as a tool for investigating positively and negatively defined concepts. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 263–268). Hillsdale, Erlbaum.
- Goldstone, R.L. (in press-a). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Goldstone, R.L. (in press-b). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*.
- Goldstone, R.L., Gentner, D., & Medin, D.L. (1989). Relations relating relations. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 131–138). Hillsdale, NJ: Erlbaum.
- Goldstone, R.L., & Medin, D.L. (in press-a). The time course of similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Goldstone, R.L., & Medin, D.L. (in press-b). Interactive activation, similarity, and mapping. In K. Holyoak & J. Barnden (Eds.), *Advances in connectionist and neural computation theory*, Vol. 2: *Connectionist approaches to analogy, metaphor, and case-based reasoning*. New Jersey: Ablex.
- Goldstone, R.L., Medin D.L., & Gentner, D. (1991). Relations attributes and the non-independence of features in similarity judgments. *Cognitive Psychology*, *23*, 222–262.

- Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects*. New York: Bobbs-Merrill.
- Half, H.M., Ortony, A., & Anderson, R.C. (1976). A context-sensitive representation of word meanings. *Memory & Cognition*, 4, 378-383.
- Hardman, P.T., Dufresne, R., & Mestre, J.P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition*, 17, 627-638.
- Harnad, S. (1987). *Categorical perception*. Cambridge: Cambridge University Press.
- Hock, H.S., Tromley, C., & Polmann, L. (1988). Perceptual units in the acquisition of visual categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 75-84.
- Holvoet, K.J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, 15, 332-340.
- Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 322-330.
- Indurkha, B. (1992). *Metaphor and cognition*. Dordrecht: Kluwer.
- Johnson, K.E., Mervis, C.B., & Boster, J.S. (1992). Developmental changes within the structure of the mammal domain. *Developmental Psychology*, 28, 74-83.
- Jones, S.S., & Smith, L.B. (1993). The place of perception in children's concepts. *Cognitive Development*, 8, 113-140.
- Kahneman, D., & Tversky, A. (1982). Subjective probability: A judgment of representativeness. In Kahneman, D., Slovic, P., & Tversky, A. (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 32-47). New York: Cambridge University Press.
- Keil, F.C. (1989). *Concepts, kinds and development*. Cambridge, MA: Bradford Books/MIT Press.
- Kelly, M.H., & Keil, F.C. (1987). Metaphor comprehension and knowledge of semantic domains. *Metaphor and Symbolic Activity*, 2, 33-51.
- Kemler, D.G. (1983). Holistic and analytic modes in perceptual and cognitive development. In T.J. Tighe & B.E. Shepp (Eds.), *Perception, cognition, and development: Interactional analyses*. (pp. 77-101). Hillsdale, NJ: Erlbaum.
- Kemler, D.G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior*, 23, 734-759.
- Kemler Nelson, D.G. (1989). The nature and occurrence of holistic processing. In B. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process*. Hillsdale, NJ: Erlbaum.
- Keren, G., & Baggett, S. (1981). Recognition models of alphanumeric characters. *Perception and Psychophysics*, 29, 234-246.
- Komatsu, L.K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, 112, 500-526.
- Krumhansl, C.L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 85, 450-463.
- Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.
- Lakoff, G. (1986). *Women, fire and dangerous things: What categories tell us about the nature of thought*. Chicago: University of Chicago Press.
- Markman, A.B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25, 431-467.
- Marks, L.E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 384-394.
- Medin, D.L. (1986). Comment on "Memory storage and retrieval processes in category learning." *Journal of Experimental Psychology: General*, 115, 373-381.
- Medin, D.L. (1989). Concepts and conceptual structure. *American Psychologist*, 44, 1469-1481.
- Medin, D.L., Goldstone, R.L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science*, 1, 64-69.
- Medin, D.L., Goldstone, R.L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254-278.
- Medin, D.L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge: Cambridge University Press.
- Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- Medin, D.L., & Shoben, E.J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, 20, 158-190.
- Medin, D.L., Watermaker, W.D., & Hampson, S.E. (1987). Family resemblance, concept cohesiveness, and category construction. *Cognitive Psychology*, 19, 242-279.
- Melara, R.D. (1989). Similarity relations among synthetic stimuli and their attributes. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 212-231.
- Melara, R.D. (1992). The concept of perceptual similarity: From psychophysics to cognitive psychology. In D. Algom (Ed.), *Psychophysical approaches to cognition* (pp. 303-388). Amsterdam: North-Holland.
- Melara, R.D., & Marks, L.E. (1989). Similarity relations among synthetic stimuli and their attributes. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 212-231.
- Melara, R.D., Marks, L.E., & Lesko, K. (1992). Optional processes in similarity judgments. *Perception & Psychophysics*, 51, 123-133.
- Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.
- Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.
- Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.
- Nosofsky, R.M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 87-108.
- Nosofsky, R.M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23, 94-140.
- Nosofsky, R.M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F.G. Ashby (Ed.), *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.
- Nosofsky, R.M., Clark, S.E., & Shin, H.J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 282-304.
- Palmer, S.E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9, 441-474.
- Palmer, S.E. (1978). Structural aspects of visual similarity. *Memory & Cognition*, 6, 91-97.
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, 72, 407-418.
- Podgorny, P., & Garner, W.R. (1979). Reaction time as a measure of inter-intraobject visual similarity: Letters of the alphabet. *Perception & Psychophysics*, 26, 37-52.
- Pomerantz, J.R. (1986). Visual form perception: An overview. In *Pattern recognition by humans and machines: Visual perception* (Vol. 2). New York: Academic Press.
- Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract idea. *Journal of Experimental Psychology*, 77, 353-363.
- Quine, W.V.O. (1977). Natural kinds. In S.P. Schwartz (Ed.), *Naming, necessity, and natural kinds*. Ithaca, NY: Cornell University Press.
- Reed, S.K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382-407.
- Rescorla, R.A., & Furrrow, D.R. (1977). Stimulus similarity as a determinant of Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 3, 203-215.
- Rips, L.J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity, analogy, and thought* (pp. 21-59). Cambridge: Cambridge University Press.
- Ritov, I., Gati, I., & Tversky, A. (1990). Differential weighting of common and distinctive components. *Journal of Experimental Psychology: General*, 119, 30-41.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 303-322.
- Rosch, E., & Mervis, C.B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Rosch, E., Mervis, C.B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 7, 573-605.

- Ross, B.H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology*, 16, 371-416.
- Ross, B.H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 629-639.
- Ross, B.H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 456-468.
- Ross, B.H., Perkins, S.J., & Tempenny, P.L. (1990). Reminding-based category learning. *Cognitive Psychology*, 22, 460-492.
- Roth, E.M., & Shoben, E.J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, 15, 346-378.
- Sadler, D.D., & Shoben, E.J. (1993). Context effects on semantic domains as seen in analogy solution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 128-147.
- Schank, R.C., Collins, G.C., & Hunter, L.E. (1986). Transcending inductive category formation in learning. *Behavioral and Brain Sciences*, 9, 639-686.
- Sergent, J., & Takane, Y. (1987). Structures in two-choice reaction-time data. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 300-315.
- Shanon, B. (1988). On similarity of features. *New Ideas in Psychology*, 6, 307-321.
- Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317-1323.
- Shweder, R.A. (1977). Likeness and likelihood in everyday thought: Magical thinking in judgments about personality. *Current Anthropology*, 18, 4.
- Sjöberg, L. (1972). A cognitive theory of similarity. *Göteborg Psychological Reports*, 2(10).
- Smith, E.E., & Medin, D.L. (1984). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Smith, E.E., Shafir, E., & Osherson, D. (1993). Similarity, plausibility, and judgments of probability. *Cognition*, 49, 67-96.
- Smith, E.E., & Sloman, S.A. (submitted). Similarity versus rule-based categorization.
- Smith, L.B. (1983). Development of classification: The use of similarity and dimensional relations. *Journal of Experimental Child Psychology*, 36, 150-178.
- Smith, L.B. (1989a). From global similarity to kinds of similarity: The construction of dimensions in development. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 146-178). Cambridge: Cambridge University Press.
- Smith, L.B. (1989b). A model of perceptual classification in children and adults. *Psychological Review*, 96, 125-144.
- Smith, L.B., & Heise, D. (1992). Perceptual similarity and conceptual structure. In B. Burns (Ed.), *Percepts, concepts, and categories: The representation and processing of information* (pp. 233-272). Amsterdam: North-Holland.
- Smith, L.B., & Kemler, D.G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology*, 10, 502-532.
- Smith, J.D., & Kemler Nelson, D.G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General*, 113, 137-159.
- Smith, L.B., & Sera, M.D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, 24, 99-142.
- Spelke, E.S. (1990). Principles of object perception. *Cognitive Science*, 14, 29-56.
- Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Suzuki, H., Ohnishi, H., & Shigemasa, K. (1992). Goal-directed processes in similarity judgment. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 343-348). Hillsdale, NJ: Erlbaum.
- Tanaka J.W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, 23, 457-482.
- Torgerson, W.S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Townsend, J.T. (1971). Alphabetic confusion: A test for individuals. *Perception & Psychophysics*, 9, 449-454.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, 89, 123-154.
- Umiltà, C., Bagnara, S., & Simion, F. (1978). Laterality effects for simple and complex geometrical figures. *Neuropsychologica*, 16, 43-49.
- Ward, T.B. (1983). Response tempo and separable-integral responding: Evidence for an integral-to-separable processing sequence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 103-112.
- Ward, T.B. (1989). Analytic and holistic modes of processing in category learning. In B. Shepp & S. Ballsteros (Eds.), *Object perception: Structure and process*. Hillsdale, NJ: Erlbaum.
- Ward, T.B., & Scott, J.G. (1987). Analytic and holistic modes of learning family-resemblance concepts. *Memory & Cognition*, 15, 42-54.
- Ward, T.B., Stagner, B.H., Scott, J.G., & Marcus-Mendoza, S.T. (1989). Classification behavior and measures of intelligence: Dimensional identity versus overall similarity. *Perception & Psychophysics*, 45, 71-76.
- Wattenmaker, W.D. (1993). Incidental concept learning, feature frequency, and correlated properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 203-222.
- Wattenmaker, W.D., Dewey, G.I., Murphy, T.D., & Medin, D.L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology*, 18, 158-194.
- Wattenmaker, W.D., Nakamura, G.N., & Medin, D.L. (1988). Relationships between similarity-based and explanation-based categorization. In D. Hilton (Ed.), *Science and natural explanation: Common sense conceptions of causality* (pp. 204-240). New York: New York University Press.
- Whorf, B.L. (1941). *Languages and logic*. In J.B. Carroll (Ed.), *Language, thought, and reality: Selected papers of Benjamin Lee Whorf* (pp. 233-245). Cambridge, MA: MIT Press.
- Wisniewski, E.J., & Medin, D.L. (in press). Harpoons and long sticks: The interaction of theory and similarity in rule induction. In D. Fisher & M. Pazzani (Eds.), *Computational approaches to concept formation*. San Mateo, CA: Morgan Kaufman.
- Yuille, A.L., & Ullman, S. (1990). Computational theories of low-level vision. In D. Osherson, S. Koslyn, & J. Hollerbach (Eds.), *An invitation to cognitive science*. (Vol. 2). Cambridge, MA: MIT Press.