Psychological Review
1993, Vol. 100, No. 2, 254-278

0033-295X/93/$3.00

Respects for Similarity

Douglas L. Medin, Robert L. Goldstone, and Dedre Gentner

This article reviews the status of similarity as an explanatory construct with a focus on similarity judgments. For similarity to be a useful construct, one must be able to specify the ways or respects in which two things are similar. One solution to this problem is to restrict the notion of similarity to hard-wired perceptual processes. It is argued that this view is too narrow and limiting. Instead, it is proposed that an important source of constraints derives from the similarity comparison process itself. Both new experiments and other evidence are described that support the idea that respects are determined by processes internal to comparisons.

Similarity is one of the most central theoretical constructs in psychology. It pervades theories of cognition. Transfer of learning is said to hinge crucially on the similarity of the transfer situation to the original training context (Osgood, 1949; Thorndike, 1931). An important Gestalt principle of perceptual organization is that similar things will tend to be grouped together. The likelihood of successfully remembering depends on the similarity of the original encoding to those operations during retrieval (Roediger, 1990). Also, people's beliefs about whether, for example, ostriches have some property, given that robins have it, are assumed to vary as a function of how similar robins are to ostriches (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990). Lastly, most of the otherwise distinct theories of categorization share the assumption that the likelihood of assigning some example to a category depends on the similarity of the example to the category representation. David Premack (personal communication, October 12, 1990) nicely summarized the attitude of many researchers: "The human mind has a considerable investment in similarity."

What exactly has the human mind invested in? In this article we take a close look at similarity as a theoretical and empirical construct. Although the fundamental importance of similarity is both historically and intuitively compelling, the status of similarity as an explanatory construct has been called into question. This questioning has been particularly evident in domains in which researchers have suggested that people's conceptual systems are organized by naive theories (e.g., Keil, 1989). The basic claim is that similarity is just too flexible and underdetermined to ground cognition. Our aim will not be to resolve the tension between similarity and theory but to address the more general question of whether similarity has substance. To provide the context for this discussion, we first need to make clear the nature of the criticisms of similarity.

Citing the body of research associated with Tversky's featural theory of similarity (e.g., Gati & Tversky, 1984; Tversky, 1977), Murphy and Medin (1985) noted that "the relative weighting of a feature (as well as the relative importance of common and distinctive features) varies with the stimulus context and task, so that there is no unique answer to the question of how similar is one object to another" (p. 296). They argued that similarity is too flexible to define categories and that it is more like a dependent than an independent variable: "The explanatory work is on the level of determining which attributes will be selected, with similarity being at least as much a consequence as a cause of conceptual coherence" (Murphy & Medin, 1985, p. 296).

Philosopher Nelson Goodman (1972) was more blunt; he called similarity "invidious, insidious, a pretender, an imposter, a quack" (p. 437). Goodman claimed that the similarity of A to B is an ill-defined, meaningless notion unless one can say "in what respects" A is similar to B. He argued that similarity, like motion, requires a frame of reference. Just as one has to say what something is moving in relation to, one also must specify in what respects two things are similar. For example, if Mary says that John is similar to Bill, one may have no idea what she means until she adds the observation that they both are avid chess players. In Goodman's words, "we must search for the appropriate replacement in each case; and 'is similar to' functions as little more than a blank to be filled" (p. 445). In short, similarity seems to disappear when it is analyzed closely, because the meaning is conveyed by the specific respects, not the general notion of similarity.

If Nelson Goodman's argument is correct, then one should perhaps "pay last respects" to the concept of similarity and move beyond the emptiness of circular explanations. We believe that the crux of the matter is, in fact, *respects* and that Goodman's critique needs to be taken seriously. There are several distinct senses in which Goodman could be correct that vary dramatically in their implications for psychology. For example, one might assume that the perceptual system determines similarity in a fairly rigid manner. In that case, Good-

Douglas L. Medin and Dedre Gentner, Department of Psychology, Northwestern University; Robert L. Goldstone, Department of Psychology, Indiana University.

Preparation of this article was supported by National Science Foundation Grant 89-18700. We thank Art Markman, Edward Wisniewski, Henry Gleitman, Ellen Markman, Ed Smith, Steve Sloman, Larry Barsalou, Javier Sainz, Evan Heit, our entire lab groups, and especially Keith Smith for valuable advice, comments, and suggestions. The research described in this article conforms with the American Psychological Association ethical standards for research involving human participants.

Correspondence concerning this article should be addressed to Douglas L. Medin, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston, Illinois 60208-2710.

man is technically correct in that the explanatory work is being done by the constraints embodied in the perceptual system. Still, psychologists would not find this too upsetting because they could say that "similarity" is just shorthand for similarity as constrained by the perceptual system. We shall argue, however, that similarity is often quite flexible and that, therefore, Goodman's critique is potentially very damaging. However, we shall also argue that respects are systematically fixed by the similarity comparison process and that a good part of the ongoing research on similarity can be viewed as providing respects for similarity. Our thesis is that quite a bit is known about respects, perhaps even enough that similarity can survive as an explanatory construct. For the moment, however, that conclusion requires support from a number of facts not now in evidence. Before turning to this body of evidence, we offer some further distinctions to define our scope of inquiry.

MEASURES OF SIMILARITY

On a broad level, one may distinguish at least three distinct types of similarity: similarity as measured indirectly, direct judgments of similarity, and similarity as a theoretical construct. As an example of addressing similarity indirectly, people may be asked to identify individual confusable stimuli, and the pattern of confusion errors may reveal underlying similarities. The idea is that the more similar two stimuli are, the more likely they are to be confused (e.g., J. E. K. Smith, 1980). Another measure might be false alarms in a new–old recognition memory test. The more similar a new item is to old items, the more likely that the new item should be judged as "old" (Gillund & Shiffrin, 1984). One might also index similarity by means of a same versus different judgment task. The more similar the stimuli, the longer it should take to respond that the stimuli are different and the more likely one should be to wrongly call the stimuli the same. Finally, as a fourth example, one may ask people to categorize novel examples on the idea that the more similar the new example is to instances of some category, the more likely it is that the new example will be assigned to that category.

Other tasks require a direct assessment of similarity. People may be asked to rate the similarity or dissimilarity of stimuli on some scale or to judge which of a set of alternatives is most similar to some standard stimulus.

Direct and indirect measures may be used as converging operations to get at similarity as a construct. The third and perhaps most central use of similarity is in models of cognition, which often assume that similarity is either directly or indirectly computed. For example, similarity-based models of categorization assume that people evaluate the similarity of a new item to representations associated with alternative categories (e.g., Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1989; Reed, 1972). The Osherson et al. (1990) category-based induction model posits that the similarity of conclusion members to premise members is explicitly computed to determine inductive judgments. To the extent that these models are successful, they indirectly support the theory of similarity they embody. Other models (e.g., Estes, 1972; Lee & Estes, 1977) contain processing mechanisms that lead to performance varying as a function of similarity. In these models, one should be more likely to confuse two stimuli that differ in only one way than stimuli that differ in two ways. Again, a successful model provides support for both the processing assumptions and the associated similarity analysis.

In the present article, our focus will be on similarity judgments and their associated processes. Although we will make no strong claims about the status of indirect measures of similarity, later we take up the question of whether similarity is a unitary construct. In the next section, we further explicate criticisms of similarity and then turn to a detailed consideration of their implications.

CASE AGAINST SIMILARITY

Why have people like Murphy and Medin suggested that similarity is too unconstrained to ground other cognitive processes such as categorization? The argument is as follows: Similarity is assumed to be based on matching and mismatching properties or predicates. Two things are similar to the extent that they share predicates and dissimilar to the extent that predicates apply to one entity but not the other. However, any two things share an arbitrary number of predicates and differ from each other in an arbitrary number of ways (see Goodman, 1972; Watanabe, 1969). The only way to make similarity nonarbitrary is to constrain the predicates that apply or enter into the computation of similarity. It is these constraints and not some abstract principle of similarity that should enter one's accounts of induction, categorization, and problem solving. To gloss over the need to identify these constraints by appealing to similarity is to ignore the central issue.

Defenders of similarity might point out that similarity seems to work quite well despite its detractors. For example, multidimensional scaling algorithms produce stable and informative solutions (e.g., Nosofsky, 1988; E. E. Smith, Shoben, & Rips, 1974), and these representations of similarity are instrumental in determining property verification times and categorizations. In addition, there is fairly good across-subject agreement in similarity ratings (although perhaps less agreement than one might expect; see Hutchinson & Lockhead, 1977). This agreement belies the argument that similarity is hopelessly ambiguous. Similarity seems to work well enough to support inductive inferences, categorization, and generalizations concerning learning, memory, and transfer.

However, similarity critics are unlikely to be persuaded by the empirical success of similarity. Their response might be twofold. One ploy would be to argue that natural constraints on features for research participants tend also to be those for experimenters and that this "hidden contract" gives similarity apparent stability. If this is so, then one needs to learn what these natural constraints are rather than attributing them to some objective state of affairs in the world. For example, Melara, Marks, and Lesko (1992) recently reported that whether a city block or Euclidean metric best described relations among multidimensional stimuli could vary as a function of instructions. Previously, these metrics had been primarily linked to stimuli rather than optional processes. A related counterargument is that researchers furthermore restrict their attention to only those experimental contexts in which their conjectures about the relevant aspects of similarity are likely to be sup-

ported. If important factors are not allowed to vary, then these factors will not influence observations.

One of the most influential similarity theorists, Amos Tversky, explicitly acknowledged the problem of constraints.

> When faced with a particular task (e.g. identification or similarity assessment) we extract and compile from our data base a limited set of relevant features on the basis of which we perform the required task. Thus, the representation of an object as a collection of features is viewed as a product of a prior process of extraction and compilation. (Tversky, 1977, pp. 329–330)

It is this prior process that similarity detractors would argue should be the focus of attention. (Despite the preceding declaration, as we will show, Tversky's work bears at least partially on this issue.)

IMPLICATIONS

At the very least, it seems that Nelson Goodman's critique cannot be dismissed out of hand. Just how serious the respects problem is may depend on both the goals of researchers and the domain in question. Consider some possibilities along with their implications for the stability of similarity.

Similarity Is Hard Wired

Earlier we mentioned this perspective, along with the idea that the term *similarity* could be thought of as shorthand for fixed, perceptual similarity. There are several limitations with this approach. One problem is that similarity relations may vary with processing time. In this case, the goal should be to specify a mechanism that could account for these changes. The explanatory power would then derive partly from the theory of similarity structure and partly from the processes that operate on it.

Even if similarity appears to be fixed in particular domains, the basis for similarity may vary across domains. A model for generating past tense may be successful if it bases its generalization on phonological similarities (e.g., Rumelhart & McClelland, 1986), but it is likely doomed to failure if it formulates generalizations in terms of semantic similarity. Semantic similarity, however, will be a necessary component in lexical priming models (Meyer & Schvaneveldt, 1971). Therefore, using the general term *similarity* risks ignoring the important problem of how the right kind of similarity becomes linked with a given generalization.

In our opinion, to claim that similarity is hard wired and perceptual is to draw an ill-advised sharp distinction between cognitive and perceptual processes. First of all, measures of perceptual similarity do not completely converge on a single construct. Palmer (1978) found that common structural relations influence similarity, as measured by a perceptual same–different task. However, the influence of relational similarity is greater when figures are presented simultaneously than when one figure appears 1 s before the other figure, so similarity seems to be tied up with process. Furthermore, we have found that subjects judging the similarity of purely visual displays show sensitivity to abstract relations such as "in both displays, the left shape is larger and darker than the right shape" and "this display has elements of increasing darkness, just as this display has elements of increasing size" (Goldstone, Gentner, & Medin, 1989; Medin, Goldstone, & Gentner, 1990). These observations belie clear boundaries between perception and conception. The reason to restrict similarity to purely perceptual aspects is to firmly ground it, but the cost of this restriction is a drastic reduction in similarity's dominion and, consequently, its explanatory power. Similarity cannot be completely inflexible for the work it has to do. Later we shall argue that it is not necessary for similarity to be hard wired and perceptual because cognitive constraints may also be powerful.

Similarity Changes With Experience

Selective Learning

In the domain of discrimination learning, it has long been assumed that organisms can learn to attend to relevant dimensions and to ignore irrelevant or uninformative dimensions of stimuli. Theories of selective attention have been quite successful in accounting for a variety of learning and transfer phenomena (e.g., Fisher & Zeaman, 1973; Sutherland & Mackintosh, 1971; Zeaman & House, 1963). Therefore, it seems natural to formulate models in which similarity can change with experience. Indeed, L. B. Smith and Heise (1992) have argued that the role of perceptual similarity in conceptual development has been substantially underestimated because of a tendency to view perceptual similarity as fixed.

Developmental Changes

There has been a variety of proposals suggesting that relative to adults, children process stimuli in a more holistic manner (e.g., Keil & Batterman, 1984; Kemler-Nelson, 1989; Shepp, 1983; L. B. Smith, 1989a; L. B. Smith & Kemler, 1977). The general idea is that children are less likely to analyze a stimulus into its components and instead respond in terms of overall similarity. Linda Smith (1989b) has offered a quantitative model of perceptual classification based on the conjecture that there is a developmental increase in the tendency to differentially weight dimensions and specifically to give special weighting to identity of values along a dimension. These proposals and associated observations suggest that adults may be more analytic and more flexible about similarity than are children.

Another observed trend is that, as children mature, their similarity judgments become increasingly based on more abstract, more relational, less superficial properties (Gentner, 1988; Gentner & Toupin, 1986). Gentner (1988) cited evidence that children give attributional interpretations to comparisons that adults interpret relationally. Given "a cloud is like a sponge," a 5-year old typically explains that "They both are round and fluffy," whereas the adult typically responds, "They both hold water and give it back later." Although these developmental changes could stem from general processing differences, Gentner and Rattermann (1991) reviewed work in development of similarity and concluded that the relational shift can largely be accounted for in terms of changes in the content and structure of knowledge (see also Carey, 1984). Consistent with this conclusion are demonstrations that young children are capable of responding to abstract relations when they possess

appropriate knowledge (Brown, 1989, 1990; Gentner, 1977, 1989).

Knowledge and Expertise

Chi, Feltovich, and Glaser (1981) noted that the basis for similarity appears to change with expertise. They examined how novices and experts classified physics problems. Novices tended to classify on the basis of superficial or surface features, whereas experts classified on the basis of deeper underlying principles. This difference between novices and experts appears to be quite general.

Knowledge effects extend even to infants. Kolstad and Baillargeon (1991) have studied how 10–12-month-olds generalize the concept of containing object on the basis of experience with particular cups. In the absence of specific knowledge, generalization was based on overall similarity. When infants were given experience with the functional property that bottoms help contain substances, generalization was based on the functional property rather than overall similarity.

Implications

Each of the sets of observations just discussed seems to suggest that similarity is flexible rather than fixed. Goodman's criticism of similarity appears to be supported. However, similarity is not vacuously flexible as long as systematic changes in the process of determining similarity can be established. For example, the claim that similarity assessment shifts from a holistic to an analytic process asserts that specific processing constraints are observable at particular ages. Similarity is constrained to follow systematic changes in how it is determined. If similarity is conceived of as a process that acts on representations, then it is partially fixed by the systematic processing changes suggested here.

Similarity Changes in Context-Specific Ways

Contextual Cues

People's judgments of the typicality or goodness of example of instances of a concept have been shown to vary with the context provided (e.g., Roth & Shoben, 1983). Barsalou (1982) has demonstrated that similarity judgments also vary when particular contexts are specified. For example, a snake and a raccoon were judged much less similar when no explicit context was given than when the context of *pets* was provided. The general idea is that the context tends to activate or make salient context-related properties, and, to the extent that examples being judged share values of these activated properties, their similarity is increased.

Linguistic Context

Linguistic contexts influence patterns of category extension in young children. In the procedure of interest, children are shown a set of objects varying in their similarity in terms of size, shape, and texture. In the control context, the experimenter points to an object and says "See this? Can you find another one?" In the linguistic context, the experimenter says "See this wug? Can you find another wug?" Relative to the control condition, the linguistic context is associated with an increased tendency to generalize in terms of shape rather than size or texture (Landau, Smith, & Jones, 1988). Three-year-olds apparently know that in the context of a noun, shape is likely to be the relevant aspect of similarity (see also Ward, Becker, Hass, & Vela, 1991).

Analogy and Relational Structure

Recent research in analogy suggests that what is crucial in analogical reasoning is not overall similarity but relational or structural similarity (e.g., Gentner, 1983, 1989). We need to define some terms to make this point clear. First, one may distinguish between attributional and relational similarity. Roughly, the distinction is as follows: Attributes are predicates taking one argument (e.g., X is red, X is large), whereas relations are predicates taking two or more arguments (e.g., X collides with Y, X is larger than Y). Attributes are used to state properties of objects; relations express relations between objects or propositions.

Gentner has argued that relational similarity has a special status in analogical reasoning (see also Hesse, 1963). Thus, when one suggests that "an atom is like the solar system," the intended meaning involves relations such as "revolves around," "more massive than," and "attracts" rather than attributes such as "hot" or "yellow." According to Gentner's structure mapping theory, interpreting an analogy is fundamentally a matter of finding a common relational structure. The objects in the two domains are placed in correspondence on the basis of holding like roles in the relational structure, not on the basis of intrinsic attributional similarity. Thus, matches and mismatches in object attributes can be neglected. One further observation is that people prefer to interpret analogies in terms of deep, cohesive systems of relational matches rather than sets of isolated relationships. That is, the presence of higher order relations between relations is an important determinant of the subjective appeal of an analogy (Rattermann & Gentner, 1987). The key to analogy is common systems of relations rather than sheer number of matching predicates or overall similarity.

Although there are differences between alternative theories of analogy, there appears to be a consensus that relational similarity is at the core of interpreting analogies (e.g., Holyoak & Thagard, 1989). The fact that people can correctly interpret analogies suggests that when cues indicate that a comparison involves an analogy, people realize that relational structure will be the relevant aspect of similarity.

We believe that these observations on analogy are relevant to the understanding of similarity. First, this view suggests that abstract relations contribute to similarity. Most important, however, we will argue that relational structures crucially determine the process of setting up correspondences between entities and that these correspondences are critical for determining similarity.

Implications

Context-specific changes in similarity reveal the further flexibility of similarity. Once again, however, the apparent flexibil-

ity is governed by systematic changes with context. Although two things may vary in their similarity depending on context, the full process–structure specification of a similarity comparison would take context into account. What seems to be random flexibility in item similarity becomes orderly when additional environmental variables are taken into account.

Similarity Is Fixed Externally and Arbitrarily

So far we have paid more attention to perceptual similarity than conceptual similarity, and so far similarity has not emerged as the sort of chameleon that Goodman suggested it might be. However, maybe we have only been looking where the light is good. Perhaps when we say that John is similar to Bill, no one understands us unless we say in what respects they are similar. Then we face the danger that similarity statements simply assert that one or more predicates apply to the entities being compared. That is, similarity structure would do no explanatory work, and the external specification of respects would have all of the burden. To the extent that this possibility is true, it would be very bad news indeed for psychologists who rely so heavily on the construct of similarity. In particular, as one moves from perceptual similarity to conceptual comparisons, this pessimistic view gains plausibility. Although we have good accounts of selective attention to brightness rather than shape, we are far from a credible account of selective attention in comparing abstract concepts such as *United States* and *Canada*.

INITIAL ASSESSMENT

It seems to us that there is cause for concern about the status of similarity. The idea of similarity as fixed is likely to be of narrow application, because there is quite clear evidence for the flexibility of similarity. Still, as long as similarity structures are linked to corresponding processing principles that address changes with presentation time, experience, and context, one retains a reasonably coherent notion of similarity. The central question is whether similarity is even more flexible than current theories allow, so much so that similarity is often arbitrary.

In this article, we present several experiments that demonstrate that similarity is highly flexible, even disturbingly flexible. Nonetheless, these observations will be used to argue for an important way in which respects are fixed: by the nature of the comparison process itself. Our thesis is that there are systematic and well-structured patterns to how multiple pieces of information are structured to yield similarity assessments.

The principle underlying our thesis is that similarity needs to be understood as a process. Previous work by Tversky and others suggests that researchers ignore the processing side of similarity at their peril. We argue that an analysis of similarity processing is crucial to providing respects for similarity.

FRAMEWORK FOR COMPARISON

Functions of Similarity

It may seem strange to raise the issue of what functions similarity serves, given that most researchers would concede the ubiquity of similarity in other cognitive tasks. We believe, however, that the issue merits a closer look, because the answers to this question provide the framework for a good deal of ongoing research.

Similarity as a Heuristic

One function of similarity is to allow people to make educated guesses in the face of limited knowledge. The general notion is that, in the absence of specific knowledge, people use similarity as a guideline for action. To give a simple example, suppose one sees an unfamiliar type of snake and wonders if it might possibly be dangerous. To the extent that the snake in question resembles a rattlesnake rather than a common garter snake, one might be cautious. Presumably, people's perceptual and conceptual processes have evolved such that information that matters to human needs and goals can be roughly approximated by a similarity heuristic (Medin & Ortony, 1989). That is, similar things may behave in similar ways, and the things people tend to be reminded of are useful to an extent that far exceeds what would be expected on the basis of random reminding. Again, note that what is relevant is often domain dependent. To continue with the snake example, it would not be particularly helpful to note that *snake* rhymes with *snowflake*.

Goodman's argument against similarity as an explanatory construct is that the "with respect to" specification is doing the explanatory work, and not similarity itself. This oversimplifies the role that similarity plays. Much of the work on similarity in cognitive psychology has focused on how multiple pieces of information are integrated into a single evaluation. Researchers in categorization have evidence that matching and mismatching properties are combined multiplicatively rather than additively (Medin & Schaffer, 1978; Nosofsky, 1992). This multiplicative rule corresponds to an exponential decay of similarity with psychological distance, and Roger Shepard (1987) has proposed that exponential-decay functions are a universal law of stimulus generalization. A deleterious effect of Goodman's equating "similarity" with "identical with respect to property X" is that the actual process for combining properties is not considered at all. The particular form of the integration process affects similarity judgments. As such, not all of similarity's explanatory work is performed by selecting the relevant properties. The integration process that is involved in most similarity judgments places constraints on the similarity judgment.

It is actually quite important that similarity judgments typically involve multiple properties. If the "with respect to" clause is filled in with a specific property, then similarity statements are of little use. If Items X and Y are similar with respect to only a single property, then very few inferences can be made about Y, even if a great deal is known about X. By having a similarity judgment that encompasses several properties, inductions can be made with more confidence. If X and Y are similar with respect to many properties, then what is known about X may well transfer to Y. In fact, one reason to say "X and Y are similar" instead of "X and Y are similar with respect to properties P_1, P_2, and so forth" is that one may wish to leave open the possibility that unknown properties are shared by X and Y. By making a nonspecific similarity claim about X and Y, one explicitly creates an expectation for new commonalities to be dis-

covered (Gelman & Wellman, 1991; Medin & Ortony, 1989; Wellman & Gelman, 1988).

Similarity as Comparison

Similarity is also a type of comparison. Explicit similarity comparisons are often made in normal discourse. One may say that John is similar to Bill or that A is like B. The first statement is a similarity comparison, whereas the second is a simile. We believe that similes and similarity statements show a number of common properties (see Ortony, 1979, and Glucksberg & Keysar, 1990, for discussion and relevant observations). First of all, both similes and similarity comparisons appear to be directional. To say that surgeons are like butchers means something different than to say butchers are like surgeons. The former criticizes surgeons and the latter compliments butchers.

Tversky (1977) has provided direct evidence that similarity judgments may be directional in that they can be asymmetrical. In his contrast model, the less salient stimulus is more similar to the salient stimulus than vice versa. Tversky observed, for example, that people rate the similarity of Red China to North Korea to be less than the similarity of North Korea to Red China. They also overwhelmingly prefer the second comparison to the first when given a choice between them. In Tversky's model, asymmetries arise because the distinctive features of *a* in an (*a*, *b*) comparison receive more weight than the distinctive features of *b*. Later, we suggest an alternative basis for these asymmetries; for now, the important point is that similarity comparisons may be asymmetrical.

Similarity comparisons also share with similes the property that certain comparisons are anomalous. We believe that, to some degree, each of the following comparisons is anomalous:

1. Robins are similar to robins.
2. Robins are similar to birds.
3. Robins are similar to questions.

One might say that the first statement is odd because robins are not similar to robins, rather, they are identical to robins. That is, the pragmatic principle of being informative is violated. The second statement may fall prey to the same criticism; robins are not similar to birds, they are birds. That is, comparisons seem to presuppose entities on the same level of abstractness. The third statement may seem strange because the term *similar* seems to presuppose some amount or type of similarity between robins and questions, neither of which is apparent.

The reason that observations about anomaly are important is that they suggest that similarity judgments involve something more than just a calculation. If one assumes that similarity judgments consist of counting and weighting matching and mismatching properties, then all three of the preceding comparisons are perfectly sensible in that they would yield some outcome. However, that outcome would not provide any indication of anomaly. To the extent that similarity judgments include intuitions about anomaly, they must involve more than a weighting function on common and distinctive properties.

Finally, one expects both similarity statements and similes to be at least somewhat informative. If a person says "Lemons are like bananas," he or she is correct in that they are both fruits and both yellow, but the comparison is hardly informative. On the other hand, to say "Butchers are like surgeons" is to assert that some (salient) property of surgeons, such as precise cutting technique, is also true of butchers (see Ortony, 1979). We claim that similarity comparisons, like analogical comparisons, may also involve assertions in which properties of one entity become candidate properties of the other. That is, similarity is more than identity in certain respects.

What we wish to emphasize is the observation that similarity is a type of comparison with properties that are distinct from those associated with the idea that similarity is a calculation. Consequently, similarity judgments may reflect both matching and mismatching properties as well as other processes associated with comparisons, such as directionality and implicit understandings about informativeness. These processes are crucial to providing respects for similarity.

Property Activation and Comparison

Suppose one is asked to rate the similarity of the United States and England on a 20-point scale. What information will enter into the comparison? It seems exceedingly unlikely that people would be able to access all of their knowledge about these two complex entities. Presumably, only a small subset of one's knowledge about the United States and about England will be activated. Furthermore, there is no guarantee that corresponding pieces of information will be activated. Suppose, for example, that a person started to think about sports. He or she might retrieve information about England playing in the World Cup soccer match but, for the United States, initially access information mainly concerning basketball and football. Are these noncorrespondences simply treated as mismatches? Sjöberg (1972) argued that similarity judgments are obtained by an active search process in which the judge looks for ways in which the objects under consideration are similar. In this spirit, we suggest that accessed information for one concept will tend to be carried over and tested for applicability to the other concept (see Ortony's, 1979, ideas about "attempted predication" and Clement & Gentner's, 1991, discussion of the "carryover" of properties in the domain of analogy). For example, having focused on the fact that England is noted for its soccer teams, one might recall that the United States recently qualified for and played in the World Cup soccer match in Italy. Having thought about basketball in the United States, one may be unable to retrieve any information about basketball in England. Nonetheless, rather than treating this information as a mismatching property, one might infer that it is very likely that basketball is played in England, in which case the difference becomes converted into a similarity (and perhaps, as well, a difference in terms of the prominence of basketball as a sport).

The idea that the comparison of two entities constrains the properties activated (and inferred) ought to extend to perceptual as well as conceptual stimuli. At the level at which features are commonly described, features may not be instantiated in an all-or-none manner. Whether or not one decides that an object is red may depend on what it is being compared with. Finally, one might expect that when similarity comparisons are stated directionally, as in A is similar to B, properties of the B term will receive more weight as candidate inferences than properties of the A term. This would follow from the idea that similarity comparisons are, at least in part, informative assertions.

Alignment and Structure

In models in which similarity involves some function of matching and mismatching properties, implicitly there must be some process that brings properties into correspondence. We refer to the process by which entities associated with the object of comparison are put into correspondence as *alignment.* A cornerstone of our thesis is that the alignment process needs to be considered explicitly and that it is not trivial.

We begin with a simple example. Suppose that Person 1 has a striped shirt and that Person 2 has both a striped shirt and striped pants. If one allows *striped* to be a feature, does one count just one match or two? If one decides to count just one match, then would Person 3, with a plain shirt and striped pants, also have one match on the feature *striped* when compared with Person 1? If so, would the match count exactly the same as a striped shirt match?

Now make the situation slightly more complicated. Person 1 has on a black and white striped shirt and red and green checkered pants, and Person 2 has on a black and white checkered shirt and red and green striped pants (any resemblance to actual people is strictly coincidental). Is one allowed to count both the red and green matches and the striped and checkered matches, or does a commitment to one exclude the other? That is, if one aligns striped with striped, does one get a mismatch for colors? Implicit in this example is the idea that structure and global consistency, rather than simple local matches, may be important in the process of determining similarity.

We suggest that the alignment process for similarity may be roughly the same as the process of structural alignment in analogical mapping (Falkenhainer, Forbus, & Gentner, 1990; Gentner, 1983, 1989; A. B. Markman & Gentner, 1991). For example, in Falkenhainer et al.'s (1990) structure-mapping engine (SME), an initial large, mutually inconsistent set of local feature matches is coalesced into global matches by imposing the constraint of structural consistency: one-to-one correspondences and consistent relational bindings. These globally consistent mappings constitute the possible interpretations of the comparison. The degree of match under a given interpretation depends not only on the number of matches but also on the structure of the common system (how deep, cohesive, and so forth; Forbus & Gentner, 1989). Matching features get more credit if they belong to larger connected systems (Clement & Gentner, 1991), and some nonmatching local correspondences will be accepted if justified by the common matching system. The net effect is to promote common systems of interconnected knowledge.

One implication of modeling similarity as alignment of connected structure is that (as discussed earlier) features of one domain may be hypothesized to be present in the other. For example, SME, as its final step, proposes *candidate inferences* that follow from the interpretation: predicates that belong to the common system in the base representation but that are not (yet) present in the target representation.

We have dealt extensively with the details of the alignment process for similarity in other articles (Gentner, 1989; Goldstone & Medin, in press; A. B. Markman & Gentner, 1991). Related models of the alignment process for analogy have also been proposed by Holyoak and Thagard (1989) and Hofstadter, Mitchell, and French (1987). For now, we wish chiefly to emphasize that alignment is a crucial aspect of the similarity comparison process and that it must operate over systems of interconnected features.

Summary and Implications

Our framework can be briefly summarized. First, similarity comparisons involve bringing aspects of the entities into correspondence. This alignment process is dynamic and is driven by multiple (global) constraint satisfaction rather than simply finding the best local matches in a piecewise manner. A further point is that just what gets aligned is not fixed a priori but depends on the particular comparison. Furthermore, the results of the alignment process are weighted so as to favor the "best" alignments. Most important, the entities being compared mutually constrain the features that are activated or inferred. Finally, comparisons may be directional and informative.

Shanon (1988) has also argued that similarity judgments involve constructive processes. He suggested that rather than similarity being determined by features, the features are themselves fixed by the similarity comparison. Although he presented no empirical data, Shanon offered the compelling example of aunts examining their newborn nephew. "Each Aunt sees in the baby facial features resembling one of her ancestors. The same face will be associated with different features depending on which other faces it is being compared with" (Shanon, 1988, p. 311). That is, just what features the baby's face "has" will vary with the comparison.

So far, our theoretical framework far outruns any data that would reinforce it. The next section aims to simultaneously redress this imbalance and to provide further observations bearing on the issue of how aspects associated with the comparison process may determine respects.

RESPECTS ASSOCIATED WITH THE COMPARISON PROCESS

The fact that respects can vary does not mean that similarity is slippery. We begin this section with some examples from Tversky's work in which respects are fixed by processes specific to the similarity judgment. Then we describe several new experiments bearing on the framework we have outlined. Finally, we consider the implications of this process orientation.

Context Effects

Diagnosticity

Tversky's (1977) diagnosticity hypothesis is that properties that are useful for grouping or categorization become more salient and consequently exert greater influence on similarity judgments. This hypothesis implies that grouping will affect similarity. To test the diagnosticity principle, Tversky (1977; see also Tversky & Gati, 1978) collected sortings from one group of participants and then used the same stimulus sets to collect similarity judgments from a separate group of participants. In one study, the stimuli were names of countries. Given the set consisting of Austria, Sweden, Poland, and Hungary, people

tended to put Austria and Sweden in one group and Poland and Hungary in the other. Given the set consisting of Austria, Sweden, Norway, and Hungary, people partitioned it into Sweden and Norway versus Austria and Hungary. This sets the stage for the test of the diagnosticity hypothesis. Participants in the similarity task had to choose which of Sweden, Poland, and Hungary was most similar to Austria or which of Sweden, Norway, and Hungary was most similar to Austria. The idea was that participants would implicitly group Sweden and Austria in the first case and Sweden and Norway in the second set. For the first set, Sweden was picked as most similar to Austria 49% of the time, compared with 36% for Hungary. In the second set, however, Sweden was picked only 14% of the time, compared with 60% for Hungary. That is, whether the third choice was Poland or Norway had a very large influence on whether Sweden or Hungary was seen as most similar to Austria.

The diagnosticity principle nicely accounts for the results just described. When Poland is part of the choice context, the natural grouping is in terms of political system (communist vs. noncommunist, at the time the study was conducted) and Austria and Sweden share the property of being democracies. When Norway is part of the choice context, geographical location becomes a more natural way of grouping, and Austria and Hungary share geographical proximity. That is, the similarities change in a manner that can be predicted by the grouping data. In short, although similarity is not independent of the choice set, it shifts in a systematic manner with context.

Extension Effect

The extension effect is conceptually related to the diagnosticity principle. The basic idea is that properties that are shared by all entities in some context have no diagnostic value, in the sense that they cannot be used to partition the set into (sub)categories. When the context is extended or broadened to include entities that do not share these common properties, these properties will acquire diagnostic value and become more salient. Therefore, the perceived similarity of two entities in the original context should be less than their perceived similarity in an extended context. For example, the perceived similarity of Italy and Switzerland is less in the context of other European countries than in a context that includes both European and American countries, even when one controls for response scale effects associated with people's tendency to produce the same average rating for any set of comparisons (Tversky, 1977).

Comparison and Respects

The comparison framework assumes that similarity assessment is a dynamic, context-specific process that determines the appropriate alignment of entities. That is, similarity comparison is less a computation over some feature space than it is a search process. We begin by describing two recent experiments that examined the role of the comparison process in feature activation and interpretation.

Experiment 1: Ambiguity and Context-Specific Features

According to the general framework we have been discussing, activated properties of one entity in a comparison are evaluated as candidate properties of the other entity. In the first experiment, participants were asked to compare a Stimulus B either with stimulus A alone or with Stimulus C alone. The stimuli were visual forms like those shown in Figure 1. Participants were asked to list common and distinctive properties in each comparison. Our aim was to show that the properties attributed to Stimulus B depend on whether it is being compared with A or C. In particular, the stimuli were constructed with the idea that a property attributed to B in an A-B comparison might be incompatible with a property attributed to B in its corresponding B-C comparison. For example, for the third triplet in Figure 1, B might be said to be three-dimensional when compared with A but two-dimensional when compared with C.

Method

Subjects. Twenty-four University of Michigan undergraduates participated in this experiment in partial fulfillment of a course requirement. Subjects were prescreened to have at least minimal familiarity with typing, but no subject failed to pass the prescreening.

Materials. Twenty-one picture triplets were created using a Macintosh computer graphics program. Eight representative triplets are shown in Figures 1 and 2. There were two types of triplets: ambiguous and nonambiguous feature triplets. Seventeen ambiguous triplets were created. Ambiguous triplets were constructed by first creating a picture (the B pictures in Figure 1) with two possible interpretations. For example, Picture B in the top row of Figure 1 can be interpreted as possessing three or four prongs, depending on whether the right-most protrusion is considered to be part of the base or a prong. Second, the other two pictures of a triple, A and C, were constructed so as to embody each of the two interpretations of B. For example, Picture A unambiguously represents a three-pronged object, whereas Picture C clearly possesses four prongs.

The ambiguous interpretations of B were chosen to be mutually exclusive, either directly or indirectly. Picture B from the first row of Figure 1 has direct mutually exclusive interpretations. It is impossible for an object to simultaneously possess three and four prongs. Picture B from the second row of Figure 1 has indirectly mutually exclusive interpretations. Although there is nothing logically impossible about being square shaped and being pincherlike, the two interpretations are not simultaneously possible in the particular instantiation of Figure 1. If B is seen as having a pincherlike component, then there is no hidden line behind the circle; if B is seen as having a square component, then there is a hidden line behind the circle. These two interpretations of B are mutually exclusive because they require properties that cannot occur simultaneously.

The A and C pictures for ambiguous sets were designed to (a) be fairly similar to B, (b) clearly reflect the two opposite interpretations of B, and (c) be approximately equally similar to B. The 17 sets were quite distinctive from each other, possessing different amounts of shading, different shapes, and different ambiguous features.

Of the 21 picture triplets, 4 belonged to the unambiguous set. The B pictures for the unambiguous sets were designed to have at least two critical features. One of these features was present in the A but not the C picture, and the other feature was present in the C but not the A picture. In the top row of Figure 2, B's peaks are at uneven heights (jagged). C shares this feature; A does not. B has three peaks. A also has three peaks, whereas C has four. The critical features were designed to be (a) perceptually nonambiguous, (b) reasonably salient, and (c) easily expressible in English. The unambiguous stimuli were included to provide an index of the listing of common and distinctive features in the absence of ambiguity.

Procedure. The subjects were instructed that they would see 21

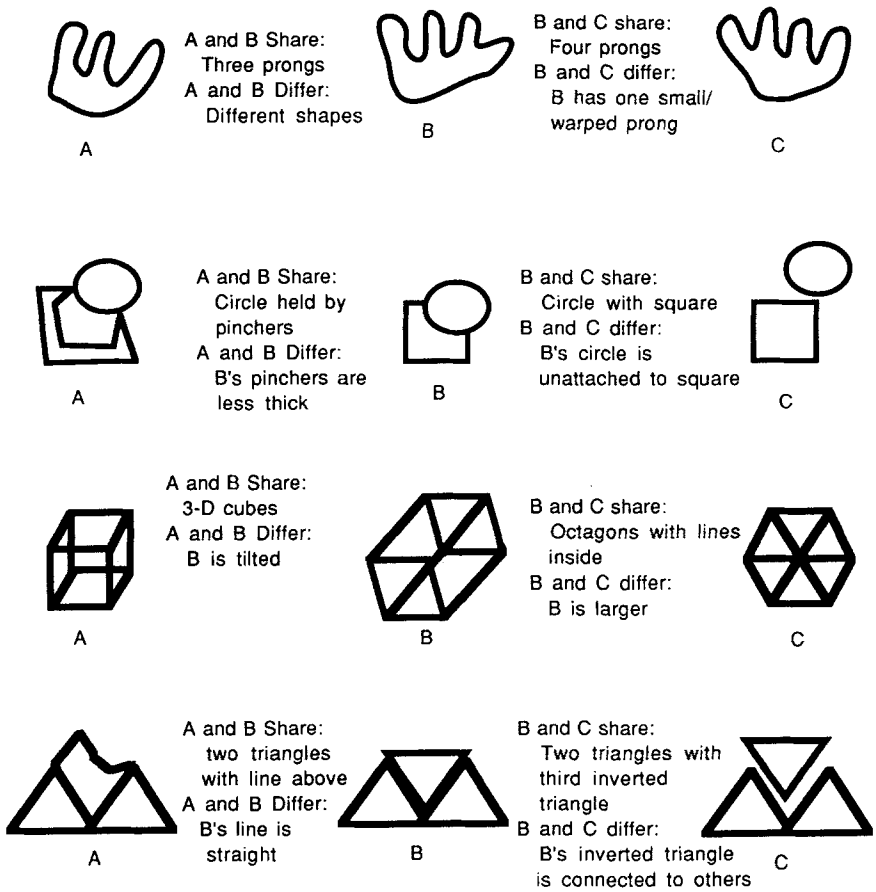

Figure 1. Representative stimuli from Set 1 of Experiment 1. (A and C's interpretations are "carried over" to B.)

pairs of pictures and that they were to type in the similarities and differences between the pairs. They were given a sample pair of a T-shaped object (A) and a T-shaped object with a tilted top bar (B). They were told that for similarities, they might list "Both have two lines. Both look like Ts. Both have a vertical line." For differences, they were told that possible differences might include "B's top bar is tilted upward on the right end. B looks like a children's slide. A's lines form a right angle."

On each of the 21 presentations, subjects saw one picture on the upper-left side of a Macintosh SE/30 screen and one picture on the upper-right side. There was a separate window in the bottom half of the screen for subjects to type in and edit their descriptions (using the delete key). For each pair, subjects were asked to list the similarities ("List the features that the pairs share") and the differences ("List the features that the pairs differ on"). Subjects were not constrained to list a certain number of features for each pair. Instead, they were instructed to type "END" when they were finished typing in features for a particular comparison. Whether they were asked to list similarities or differences first was randomized.

On each trial, subjects saw two pictures from a given triplet set. Subjects either saw Pictures A and B or Pictures B and C. Pictures A and C were displayed on the left side, whereas Picture B was always shown on the right side. For each of the 21 picture triplets, either A and B were displayed or B and C were displayed, but never both. Whether a given subject saw A or C was randomized. The pictures were approximately 3 cm × 3 cm and were separated from each other by 3 cm. The letter A was placed below the left object; the letter B was placed below the right object. These letters were provided to facilitate subjects' reference to the objects. The order of presentation of the 21 triplets was randomized.

Results

Scoring. The primary data of interest involve the number of times properties are listed of B that are true of A and C, as a function of whether B has been paired with A or C. Figure 1 illustrates the possible types of properties that could be listed for the last triplet of Figure 1.

The ambiguous item B may be interpreted in a manner that is consistent with its comparison object. When B is compared with A, A's unambiguous property of "a line crossing two triangles" may be carried across and used to describe B. Similarly, C's unambiguous property of "three triangles" can be applied to B when they are paired. Interpretations of B that are consistent with its comparison item also are possible when subjects list distinctive features. If subjects cite "B's line is straighter than A's" as a difference between A and B, then B is still being interpreted as having a line connecting two triangles (like A) and not as having three triangles (like C). The similarities and differences that are listed in Figure 1 show representative sub-

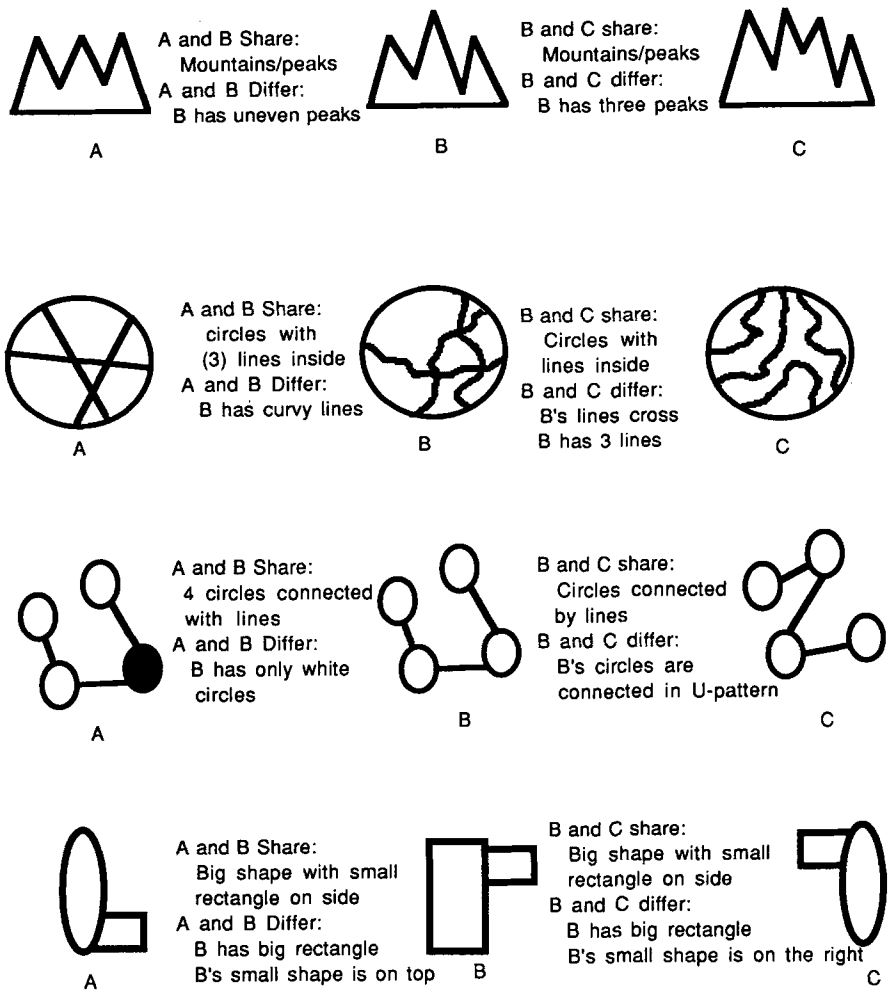

Figure 2. Stimuli from Set 2 of Experiment 1. (A and C's interpretations are contrasted with B's.)

ject responses; all of these descriptions indicate interpretations of B that are biased toward its comparison object.

Figure 1 also shows interpretations of B that are taken as evidence that its description is not biased toward its comparison item. If subjects respond that "Both B and C have (at least) two triangles," then B's description is not biased toward C's; in fact, the description indicates that B is interpreted as possessing only two triangles (like A). If a subject gives the description "B has one more triangle than A," then this is also clear evidence that B is being interpreted as showing three triangles (like C) and not as two triangles with a spanning line.

In short, Bs can be interpreted in a *comparison-consistent* or *comparison-inconsistent* manner. To determine the predominance of these two methods for interpreting the B term, the features listed for the 21 triads were analyzed. For this purpose, only subject responses for certain critical dimensions were tabulated. For the ambiguous triads, there was a single critical dimension corresponding to the dimension on which B was ambiguous and on which A and C had mutually exclusive values. For the four triplets in Figure 1, the critical dimensions were "three versus four prongs," "pincher versus square," "3-D versus 2-D," and "two versus three triangles," respectively. For the unambiguous triads, there were two critical dimensions: one for the feature that A and B shared that C did not, and one for the feature that B and C shared that A did not.

The values of the critical dimension were classified as "based on a property of A" or "based on a property of C." The description "both have three prongs" for the top triplet in Figure 1 would be classified as based on a property of A. The description "B's right-most finger is warped" would be classified as based on a property of C because the description implies that B is a four-fingered object and this property is unambiguously true of C. The categories refer to the properties used to interpret B. For every triad, the 12 subjects who received A-B pairs and the 12 subjects who received C-B pairs were classified into one of the two categories. If a subject did not mention the critical property, he or she was not placed into either category.

A judge rated each description as based on a property of A, based on a property of C, or "unsure." Because the judge evaluated only descriptions that were based on the critical dimension, she was highly confident in her classifications and rarely used the unsure category. Descriptions labeled unsure were not

included in the data analysis. The judge was not informed of the experimenter's hypothesis, and a random sample of the judge's answers was checked by a second judge. The correlation between the two judges exceeded .90.

Property listings. Descriptions based on a property of C when C was present and descriptions based on a property of A when A was present were combined and classified as comparison-consistent properties. Descriptions based on a property of C when A was present and descriptions based on a property of A when C was present were combined under comparison-inconsistent properties. For the ambiguous triads, the average number of comparison-consistent properties was 15.75. This figure is significantly greater than the average of 4.86 comparison-inconsistent properties, paired $t(16) = 9.42$, $p < .001$. For 15 of the 16 ambiguous triads, the comparison-consistent properties were more prevalent than the comparison-inconsistent properties. Figure 1 shows typical properties that are listed for similarities and differences. As an example, for the bottom triad of Figure 1, many people list "Two triangles with line above" for B when it is paired with A, but no one lists this as a property that distinguishes between B and C. Alternatively, many people list "Inverted third triangle between two triangles" as a property of B when it is paired with C, but nobody gives the "inverted third triangle" as grounds for distinguishing between A and B.

One alternate interpretation of the data is that (a) similarities are more likely to be listed than differences, (b) comparison-consistent properties are more likely to be listed as shared features than as distinctive features, and (c) these points alone account for the relative predominance of comparison-consistent properties. Significantly more similarities were listed than differences, paired $t(23) = 1.9$, $p < .05$. The average number of features listed as shared between two pictures was 2.5; the average number of differences listed was 1.9. Although there were somewhat more properties listed for similarities than differences, the relative predominance of comparison-consistent properties is found even for the "different" property listings. Just considering the listing of different properties, comparison-inconsistent properties were given an average of 4.76 times and comparison-consistent properties were listed 7.20 times, paired $t(16) = 3.43$, $p = .003$. Thus, for example, people often list "Right-most prong is warped in B" as a difference between B and C in the top triad of Figure 1, giving B an interpretation that is consistent with C's property of having four prongs.

A very different pattern of results was found for the unambiguous triads. On these triads, the comparison-consistent description was listed an average of 6.25 times, whereas the comparison-inconsistent description was listed an average of 19.25 times, paired $t(3) = -5.54$, $p < .001$. All four of the unambiguous triads yielded more comparison-inconsistent descriptions than comparison-consistent descriptions. Roughly, when presented with A and B, people list shared properties that are also shared by C; people list distinctive properties of A and B that B shares with C. Figure 2 shows typical features that are listed by subjects. For example, in the third triad, people seldom list "U-shape formation" as a shared property of A and B, even though they regularly list it as a property of B when distinguishing it from C. The result for the unambiguous stimuli is consistent with other evidence that suggests that distinctive features are more prominent than common features with pictorial stimuli (Gati & Tversky, 1984). Comparison-inconsistent features are distinctive features and comparison-consistent features are common features.

Additional observations. One other finding, not directly tested for but clearly present in the data, was the relative predominance of metaphors in similarity versus difference listings. For each triad, the number of unique metaphors was tabulated for similarity and difference listings. A unique metaphor was defined as any nonliteral term that was listed as a similarity but not a difference or as a difference but not a similarity. Thus, listing as a shared feature for A and B in the top triad of Figure 1 "Both look like combs" would be categorized as a metaphor. A liberal criterion for metaphoricity was used. Only unique metaphors were counted because these tended to be unusual and uncontroversially metaphorical. Thus, "Both have three prongs" was not considered, because *prongs* was mentioned at least once in the listings of all 24 subjects as a similarity and a difference.

An average of 4.86 unique metaphors per triad was found in the similarity listings. The average number of unique metaphors found in difference listings was significantly less (0.90), paired $t(20) = 3.66$, $p < .002$. For 20 of 22 triads, there were more unique metaphors found in similarities listings than in difference listings (of the 2 exceptions, 1 was an ambiguous triad and the other was an unambiguous triad). Although it is true that there were slightly more properties listed as similarities than differences (60.0 vs. 45.6 features per triad), this difference is not close to the fivefold increase in metaphorical properties when switching from listing different properties to listing similar properties. As an example of the effect, consider the first triad of Figure 1, where 15 mentions of *forks, teeth, combs, valleys,* and *hairs* were recorded when subjects were asked to find the similarities between B and either A or C. The only metaphors given when subjects were asked to list differences were *hands, prongs,* and *fingers,* none of which is a unique metaphor.

Discussion

These results are consistent with the proposal that the interpretation of ambiguous stimuli (B pictures) is based on properties borrowed from the unambiguous stimuli with which they are compared. If the properties attributed to B were independent of its comparison picture, then comparison-inconsistent properties should be as commonly listed as comparison-consistent properties. The fact that A and C bias B's interpretation in mutually exclusive directions speaks against the possibility that B has properties associated with both A and C and that properties are listed that are consonant with the unambiguous comparison item. Such a claim would require the top triad of Figure 1's B to have the properties "has three prongs" and "has four prongs." The point is that the figure may support mutually exclusive interpretations.

Note that the results with ambiguous stimuli are directly opposite of those in which the features for stimulus B are unambiguous (Figure 2). Unambiguous features of B are more likely to be mentioned when they differ from those of the comparison stimulus than when they match it. This result is consistent with

Gati and Tversky's (1984) evidence that distinctive features are weighted more heavily than common features for visual stimuli.

Other experiments have shown an influence of context on the representation of ambiguous shapes. Carmichael, Hogan, and Walter (1932) demonstrated that drawings are recalled so as to be consistent with their accompanying label. Selfridge (1959) showed that a shape halfway between an A and an H is interpreted as an A when surrounded by C and T and as an H when surrounded by T and E. The current results differ from these findings in two ways. First, our subjects were not led to believe that the objects came from the same class or referred to the same thing. Our subjects listed both common and distinctive features of the object pairs. Still, they were biased to interpret the objects so as to highlight their commonalities. Second, our results show that context effects extend to contexts defined by comparison items and have opposite effects depending on the clarity–ambiguity of an object.

It is not clear exactly what to make of the observation that metaphors were five times as likely to be mentioned for similarities as for differences. One possibility is that differences are derivative of similarities (see also A. B. Markman & Gentner, 1991). For example, one might note that a red circle and an orange circle are both circles and both colored and then describe the specific colors as differences. If differences tend to be derivative of (aligned) similarities, then the features listed for differences will tend to be more specific and less abstract than those listed for similarities. The relative prominence of metaphors for similarities would just be a specific instance of the general tendency for similarities to be more abstract than differences.

In any event, this experiment provides a clear demonstration that the properties or features of an object depend on what it is being compared with, even to the extreme in which the properties attributed to it in one context conflict with properties attributed to it in another context. We turn now to a further context effect, in this case with conceptual stimuli.

Experiment 2: Context and Asymmetries

The principle that properties activated for one entity of a comparison are evaluated with respect to the other entity should be especially true for conceptual entities that are richly structured. It seems likely in such cases that only a subset of the associated information enters into a comparison (recall the earlier example of the United States and England). The present experiment attempted to provide evidence supporting this conjecture by exploiting the potential asymmetry of comparisons. The idea is that properties of the base or standard are more likely to become activated than properties of the target. This leads to the prediction that the common properties associated with a similarity comparison differ depending on the direction of the comparison (see also Glucksberg & Keysar, 1990; Ortony, 1979). In particular, the common properties listed when A and B are being compared might be more closely associated with B when A is being compared with B and more closely associated with A when B is being compared with A.

Method

Subjects. Thirty-six University of Michigan undergraduates participated in this experiment in partial fulfillment of a course requirement.

Materials. The 17 pairs of words listed in Table 1 were used as materials. The word pairs were chosen to sample a variety of domains, to be fairly similar, and to have a variety of salience differences. Some words were relatively similar in their salience (*squirrels* and *mice*) and some had a relatively large salience imbalance (for our participants, *United States* was likely to be more salient than *England*).

Procedure. Subjects were presented with a seven-page booklet, each page containing three pairs of words. Subjects were instructed to (a) rate, on a 20-point scale ranging from *not similar at all* (1) to *highly similar* (20), the similarity of the words, and (b) list the features that the two words had in common. They were told to list as many similarities as they could, without spending more than 1 min on any word pair. The word pairs were displayed in the form "How similar is X to Y?" followed by the statement "Consider Y. List the properties that X has in common with Y." Eighteen subjects received booklets that had "X" filled in with the left words in Table 1 and "Y" filled in with the right words. The other 18 subjects received booklets with the opposite assignments.

Approximately 12 cm of white space was left between word pairs for subjects to write down their properties. The order of the pages (except the first) was randomized. Subjects took approximately 20 min to complete the booklet.

Results

Scoring. All feature listings were transcribed into a computer spread sheet, randomized, and given to two naive judges to score. They were given the word pairs and instructed to place each feature that was listed by a subject into one of the following five categories: (a) equally applicable to both concepts, (b) biased toward the meaning of the left concept, (c) biased toward the meaning of the right concept, (d) true of the left concept and not true of the right concept, or (e) true of the right concept and not true of the left concept.

Judges were told that the second and third categories were to be used if the feature was true of both words but seemed more appropriate for or applicable to one of the words. For example, the feature "both are scientists" is true of Einstein and Franklin, but both judges decided that the feature was more appro-

Table 1
Word Pairs and Similarity Ratings: Experiment 2

X, Y	Similarity of X to Y	Similarity of Y to X
England, United States	12.84	11.40
Prunes, apples	9.84	8.77
Albert Einstein, Benjamin Franklin	11.16	11.88
Blimps, cars	5.00	6.12
Squirrels, mice	10.84	11.94
Physics, philosophy	5.53	5.65
Doctors, engineers	9.79	10.94
China, Korea	12.21	11.35
Cows, dogs	8.37	7.47
Chocolate bars, popcorn	7.11	7.47
Wisconsin, Michigan	11.26	11.65
Pencils, crayons	12.68	15.13
Spanish, English	8.16	7.94
Skateboards, bicycles	10.37	11.24
Frisbees, boomerangs	11.16	12.94
Russia, Poland	11.21	10.39
Gorillas, elephants	8.42	7.50

priate to Einstein. Similarly, "eaten at movie theatres" was judged to be more appropriate to popcorn than chocolate bars, and "aids digestion" was judged to be more appropriate to prunes than apples. The fourth and fifth categories were reserved for clear-cut cases in which a feature was true for only one of the words. For example, the feature "both famous Americans" was rated by the judges to be true of Benjamin Franklin but not Albert Einstein.

Rating. A grand total of 2,103 features was listed by the subjects, counting each token separately. In this first analysis, no attempt was made to group similar descriptions together. The two judges agreed on 87% of their assignments of these features to the rating categories. The judges' responses were then collapsed into three categories, depending on whether the feature was biased toward neither word, toward the base word (Y in "How similar is X to Y?"), or toward the target word. Judge 1 rated 432 features as biased toward the base, 377 as biased toward the target, and 1,294 as equal or neutral. Judge 2 rated 413 features as biased toward the base, 363 as biased toward the target, and 1,327 as equal. For both judges, the number of features biased toward the base was (marginally) significantly greater than the number of features listed that were biased toward the target (Judge 1: sign test $Z = 1.88$, two-tailed $p = .06$; Judge 2: $Z = 1.79$, two-tailed $p = .073$). If only features are used for which the two judges gave the same code, then the number of second (base) word biased features is significantly greater than the number of first (target) word biased features (368 vs. 312; $Z = 2.16$, two-tailed $p < .05$). An example of the bias is as follows: Subjects are more likely to mention the property "both are found on farms" when asked "How similar are dogs to cows?" than when asked "How similar are cows to dogs?"

This asymmetry also held for an analysis of tokens rather than types. To create equivalence classes, a third judge grouped the 2,103 descriptions into semantically similar groups, resulting in 335 categories. If over 75% of the descriptions of a category were judged to be biased toward one comparison word's meaning, then the category was considered to be biased toward the word. Finally, the judge determined whether the category was more often invoked in descriptions when the word that the category was biased toward was in the base or the target. As an example, subjects were more likely to refer to the category "both make a lot of money" (which included descriptions such as "both earn a good living" and "both are rich") when asked to compare engineers with doctors rather than doctors with engineers. Overall, categories were more frequently invoked when the biased word was in the base position (binomial $Z = 1.98$, $p < .05$).

The fourth and fifth categories were given only 30 times by Judge 1 and 19 times by Judge 2. Although the results are in the direction of favoring the base, no significant differences are observed between the base (18 cases for Judge 1 and 10 for Judge 2) and target concept (12 cases for Judge 1 and 9 for Judge 2) biasing if only these codes are considered.

Similarity judgments. Although asymmetries in similarity judgments were not the focus of this study, the ratings are summarized in Table 1. The average absolute difference in ratings for a pair was 0.96, which is roughly of the magnitude observed by Tversky and Gati (1978).

One other analysis was conducted that was inspired by casual observation of the data. Most of the subjects completed the booklet in pen. It was noticed that there were a number of crossed-out ratings: ratings that were written by subjects and then crossed out and replaced by another rating. There were 34 such crossed-out ratings in total. In 22 of these cases, the crossed-out rating was replaced by a higher rating (e.g., "7" was crossed out and replaced with "8"). In 12 cases, the crossed-out rating was replaced by a lower rating. This marginally significant difference (sign test $Z = 1.71$, two-tailed $p = .088$) is consistent with the hypothesis that similarity ratings reflect an active process of searching for commonalities. Originally low ratings may be raised as subjects discover new features shared by the words.

Discussion

The results support the prediction that the properties may be more closely associated with one concept than another and that the common properties activated for a similarity comparison depend on the direction of the comparison. Note that this asymmetry requires two things to be true: that activated properties of one concept are evaluated with respect to the other concept and that the activation is biased toward the base concept. Therefore, the magnitude of the second process may represent a serious underestimate of the likelihood of the first process. An example may serve to underline this observation. We have informally asked a number of people to list numerous properties of candy bars, and no one has said that one property is that people can buy them in movie theaters. However, that is far and away the most frequently mentioned common property when people are asked to compare candy bars and popcorn. This particular property did not, however, yield an asymmetry, because it was very likely to be mentioned regardless of the direction of the comparison.

A critic might properly note that we used a fairly strong experimental manipulation aimed at asymmetries. In addition to the directional comparisons, we explicitly asked participants to focus on the base term before beginning to list common features. We believe our experimental strategy is justified as a test of the idea that (common) features may vary with respect to how closely they are associated with different concepts. That is, if common features for comparisons are equally true or equally linked with both entities, then our "strong" experimental manipulation would not have succeeded. In follow-up work by Cynthia Aguilar, Evan Heit, and Douglas L. Medin, participants were simply asked to list common and distinctive features (with no instruction to consider the base term), and consistent asymmetries of common features favoring the base term were found.

The results support two conclusions in regard to similarity comparisons. One is that the common features of two concepts may differ as a function of the direction of a comparison, and the second is that the difference is in the direction of favoring properties that are more closely associated with the base or ground of a comparison than with the target. These findings are consistent with Ortony's (1979) theory of salience imbalance and his analysis of asymmetries of similarity (see also Ortony, Vondruska, Foss, & Jones, 1985). They also fit nicely with

our general argument that respects are fixed by the comparison process itself.

Experiment 3: Context and Changes in Respects

To give a similarity rating, a subject must implicitly or explicitly decide what levels of similarity correspond to the different numeric ratings. Although typically instructed that a rating of 1 means *not similar at all* and a rating of 9 means *highly similar*, subjects must decide what counts as highly similar. Parducci (1965) has persuasively argued that the entire set of comparisons determines what rating a particular comparison will receive. A moderately similar pair of objects may receive a rating of 7 in a context of many highly dissimilar comparisons but a rating of 3 in a context of highly similar comparisons.

The third experiment investigated a more radical context sensitivity. We suggest that "highly similar" will depend both on potential contrast comparisons and on deciding what respects are relevant. The scale used to rate a particular comparison depends on the comparison itself and the respects it suggests, and not just the other comparisons. Some evidence for this possibility comes from Fillenbaum and Rapoport's (1974) analysis of Charles Fillmore's similarity ratings of verbs and accompanying protocol. The researchers reported the following: "[Fillmore] judges that 'Acquit' is similar to 'Clear' because they are *synonyms,* 'Acquit' is similar to 'Convict' because they are *alternatives from the same class,* and 'Apologize' is similar to 'Forgive' because *one has the other as its goal* (p. 57)." If one thinks of similarity ratings as judgments of "How similar is X to Y, with respect to aspects {Z}?" then the "with respect to aspects {Z}" clause is not completely determined until the actual comparison items X and Y are presented.

In the third experiment, we systematically examined the possibility of comparison-defined respects by presenting subjects with *separate-context* and *combined-context* comparisons. In the separate-context comparisons, stimuli A and B are presented, and stimuli A and C are presented in separate contexts. In the combined-context comparisons, both pairs are presented in the same context. Words are compared with antonymically, metaphorically, and categorically similar words. When the separate-context similarity ratings are contextually separated, they can be based on different respects. For example, in antonymic comparisons such as black–white, there may be a tendency to focus on differences. Categorical comparisons (e.g., black–red) may focus on shared respects. By conjecture, when the pairs appear in the same context, there may be a tendency for the same pool of respects to be involved. In the case of antonyms, this should shift attention to shared respects and boost similarity. In the case of metaphors, the natural respects should tend to be abstract shared features, and the combined context may highlight mismatching properties. Overall, antonymic comparisons should receive higher similarity ratings and metaphoric comparisons lower ratings in the combined context, relative to categorical comparisons.

Note that the predicted interaction just described differs from expectations associated with a range-frequency account of context effects. Consider antonymic pairs. They may receive somewhat lower similarity ratings than do categorical comparisons when presented in separate contexts. The tendency to use the full range of the scale may attenuate these differences. In the combined context, the categorical comparisons should be boosted relative to the antonymic comparisons because the full range of the scale can be used (antonyms will take up the low end and categorical comparisons the high end). In general, differences that appear in the separate-context condition should be amplified in the combined-context condition.

Method

Subjects. Sixty-three University of Michigan undergraduates participated in this experiment in partial fulfillment of a course requirement.

Materials. Thirty-eight sets of three phrases (in most cases, phrases were single words) were constructed. Two different types of sets were used.

In *antonymic* sets, the standard phrase is an antonym of one word and is categorically related to the other phrase. The sets {sunrise, sunset, sunbeam} and {black, white, red} are antonymic sets. There were eight such sets.

In *metaphorical* sets, the standard is metaphorically related to one word and categorically related to the other phrase. The set {skin, bark, hair} is a metaphorical set. *Skin* and *bark* are metaphorically related; both are the outside covering of living objects. *Skin* and *hair* belong to the category *body parts.* Many of the metaphorically related phrases were borrowed from Gentner (1988). There were 30 metaphorical sets. Examples include {Rolls Royce, champagne, Volkswagon}, {Monday, January, Sunday}, {sun, lightbulb, moon}, and {insult, slap, promise}.

Procedure. All comparisons were presented on Macintosh computers. Subjects saw a comparison for each of the 38 sets of phrases. Separated- and combined-context comparisons each composed half of the trials. For all comparisons, subjects were instructed to rate phrases for similarity. Subjects were told, "A rating of 1 means that phrases are not similar at all. A rating of 9 means that the phrases are *highly* similar. Use the numbers 2–8 for intermediate degrees of similarity." For the separate-context condition, two words were shown next to each other; one word was the standard word and the other word belonged in the standard word's transformational, antonymic, or metaphorical set. For combined-context comparisons, all three phrases were presented simultaneously in the form of an isosceles triangle. Subjects rated the similarity of the top word to the left word and the top word to the right word. Subjects were instructed to look at all three words before rating the two pairs.

Subjects received one block of combined-context comparisons and one block of same-context comparisons. The order of blocks was randomized, as was the position of the lower words in the combined-context comparison.

Results

The question of greatest importance is whether similarity ratings vary systematically with context. That is, does the similarity between a pair of words depend on whether the words were presented in separate pairs or in a combined context? Several measures indicated that the presentation context does influence the similarity of antonymic, categorical, and metaphorical comparisons.

Antonymic sets. Eight antonymic sets were presented to subjects. The data suggest that antonyms are judged to be more similar to each other in the combined-context comparison than in the separate-context comparison. There is not an equally significant effect for categorically related phrases. Collapsing

over the eight sets reveals an interaction. For separate-context comparisons, the standard phrase and its antonym received an average similarity rating of 4.5, compared with a rating of 5.0 between the standard and the categorically related phrase, paired $t(7) = 2.47$, $p < .05$. For combined-context comparisons, the average standard–antonym rating was 5.5, compared with a mean standard–categorical rating of 5.3 ($p > .05$). The interaction between comparison type and presentation type was significant, $F(1, 28) = 4.8$, $p < .05$. The crossover interaction was primarily caused by the substantial increase in similarity for antonymically related phrases when they were judged in the context of a categorically related phrase. For example, *black* and *white* received a similarity rating of 2.2 when presented by themselves; this rating increased to 4.0 when *black* was simultaneously compared with *white* and *red* (*red* only increased 4.2 to 4.9).

Two items produced rating reversals with the change in context. The separate-context comparison for one of these items was sunrise–sunset = 4.6 and sunrise–sunbeam = 5.2; the combined-context comparison was sunrise–sunset = 6.4 and sunrise–sunbeam = 5.7. As such, whether sunset or sunbeam is more similar to sunrise depends on whether the words are presented in two pairs or simultaneously. For seven of the eight antonymic sets, standard–antonym similarity increased more than standard–categorical similarity in the move from separate- to combined-context comparisons.

Significant results were also obtained in analyzing the data by classifying subjects. For 49 of 63 subjects, the average similarity ratings for the eight antonymic sets showed a greater increase in standard–antonym similarity than standard–categorical similarity when going from separate to combined comparisons.

Metaphorical sets. Thirty metaphorical sets were presented. Categorically related phrases become more similar to one another when they are presented in the context of metaphorically related phrases. Again, there was an interaction between comparison and presentation type. For separate-context comparisons, the standard phrase and its metaphor received an average similarity rating of 4.4, compared with a rating of 5.0 between the standard and the categorically related phrase. For combined-context comparisons, the standard–metaphor rating was 4.5, compared with a standard–categorical rating of 5.8. This interaction was significant, $F(1, 116) = 4.92$, $p < .05$. As an example of this interaction, for the separate-context comparison, skin–hair = 4.7 and skin–bark = 6.6; for the combined-context comparison, skin–hair = 5.5 and skin–bark = 5.3. Such rating reversals were found for only two sets; such a reversal indicates that which of two concepts is more similar to a third depends on the presentation format of the comparison. For 20 of the 30 sets, standard–categorical similarity increased more than standard–metaphor similarity when going from separate- to combined-context comparisons.

The results for phrase sets are supported by the data broken down by subjects. Forty-five of 63 subjects showed greater separate- to combined-context similarity increases for categorically related phrases than for metaphorically related phrases.

Discussion

Different standards of comparison and different respects seem to be used depending on the particular phrases being compared. A standard phrase does not seem very similar to its antonym when they are compared in isolation because the focus is on their dimensional difference. For example, black and white seem very different when presented as a separate pair. The standard becomes much more similar to the antonym when the standard phrase is simultaneously compared with a categorically related word. Presenting a black to red comparison in the same context as the black to white comparison substantially increases the rated similarity of black to white. Our interpretation of this effect is that separate-context comparisons use different standards of similarity and focus on different respects. When antonyms are compared by themselves, the single difference between them is highly salient. The properties that are shared by the antonyms fall into the background. The fact that white and black are at opposite ends of the gray scale is particularly salient. The fact that black and white are both monochrome colors on endpoints of the gray scale is not as salient. This commonality becomes more important when red is included in the context.

For metaphorical sets, the standard phrase seems more similar to a categorically related phrase when it is simultaneously compared with a metaphorically related phrase. The ratings for metaphorical pairs did not tend to vary with context. Although we will not pretend that we expected this exact pattern of results, they are generally consistent with our arguments. For a metaphorical comparison such as skin versus bark, understanding the metaphor involves fixing respects. For metaphors, it is not clear what the contrast set would be (presumably, alternative metaphors involving the same respects), and the comparison should be less susceptible to contextually suggested alternative respects. For the categorical comparisons, the three-way context may increase the size of the contrast set and highlight additional shared respects. Although the precise interpretation of these results is uncertain, it is clear that context produces differential effects on antonyms, metaphors, and categorical comparisons that can be understood in terms of varying contrast sets and respects.

Nonindependence of Feature Weighting

A recent series of studies by Goldstone, Medin, and Gentner (1991) has indicated that feature weighting is not independent of the outcome of the comparison process. An example of their stimulus materials is shown in Figure 3. Participants were shown a standard (T) and different pairs of alternative stimuli and asked to judge which of the alternatives was more similar to the standard. For example, A contains the most attributional matches (triangle, circle, shading), whereas D contains the most relational matches (two figures with the same shape, all figures have the same shading). B and C are intermediate in both attributional and relational matches.

Note that, as one moves from A and C to B and D, one attributional match (shading) is removed and one relational match is added (same shading). Furthermore, as one moves from A and B to C and D, one relational match (two figures with the same shape) is added and one attributional match (triangle shape) is removed. If features are evaluated independently, then A should be picked as more similar than C to T to the same degree that B is picked as more similar than D to T (similarly, A should be picked over B to the extent that C is picked over D).

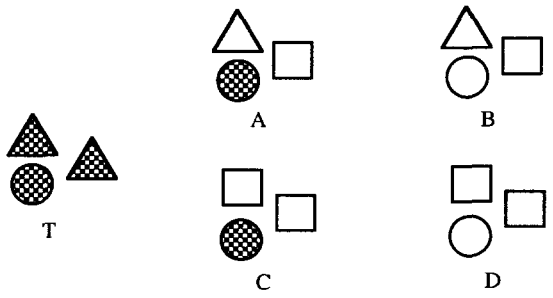

Figure 3. Sample stimuli from the Goldstone, Medin, and Gentner (1991) experiments. (As one moves from A and C to B and D, an attributional match with T is removed [shading] and a relational match is added [same shading]. Similarly, in going from A and B to C and D, an attributional match is [triangle] deleted and a relational match is [same-shape figures] added. From "Relational Similarity and the Nonindependence of Features in Similarity Judgments" by R. L. Goldstone, D. L. Medin, and D. Gentner, 1991, *Cognitive Psychology, 23,* p. 255. Copyright 1991 by Academic Press. Reprinted by permission.)

The results were that people's judgments departed systematically from independence. Specifically, the violations of independence were in the direction of choosing A and D over B and C on the paired tests. Goldstone et al. (1991) referred to this tendency as a "Max" effect: If the choice stimuli are attributionally similar to the standard, then an extra attributional match has more weight than an extra relational match; if the alternatives are relationally similar, then an extra relational match has more weight than an extra attributional match. It is as if attributes and relations form distinct "pools" and shared features affect similarity more if the pool they are in is already relatively large.

The correct interpretation of these Max effects is not entirely clear. One might argue that participants find the judgment task to be ambiguous and assume that whichever type of similarity is maximized must be what the experimenter had in mind. An equally plausible alternative is that Max effects are directly tied to the alignment and comparison process and do not depend on pragmatic factors. Whatever the correct interpretation, it is clear that the weight given to a particular match depends on the other matches in the scene.

Nor are these Max effects restricted to attributes and relations. Figure 4 shows another of a variety of stimulus sets that can produce nonindependence of features. The A, B, C, and D stimuli can match the standard T in either global letter form (A and T share the global letters B and F) or local letter form (e.g., T and B both have a global letter constructed from S's). Again, participants systematically choose both A and D as more similar to T than B and C. That is, it is better to maximize either global matches or local matches than to have a mixture of local and global matches. Other studies reveal similar effects for shapes versus shading, curved shapes versus straight shapes, and within- versus across-part matches. Therefore, the Max principle appears to have some generality.

Alignment

As we noted earlier, in models in which similarity involves some function of matching and mismatching properties, implicitly there must be some process that brings properties into correspondence. A cornerstone of our thesis is that the alignment process needs to be considered explicitly and that it is not trivial.

A. B. Markman and Gentner (1990) have investigated the relation between similarity and alignment using scenes such as the one shown in Figure 5. The two women are highly similar perceptually, but they play different roles in the scenes. In the scene on the left, the woman is the *recipient;* in the scene on the right, the woman is the *donor.* In this sense, the scenes involve cross mapping in that the most natural perceptual correspondences conflict with the relational correspondence (Gentner & Toupin, 1986).

In A. B. Markman and Gentner's (1990) study, the experimenter pointed to the cross-mapped object and asked participants to point to the object in the other scene that "went with" the cross-mapped object. One group of participants was first asked to make similarity judgments and then given the mapping questions, and another group was asked only the mapping questions. The group that first made similarity judgments was far more likely to map according to the relational structure [from the woman in scene (a) to the squirrel in scene (b)] than was the control group. Other control conditions ruled out amount of exposure to the materials as the reason for the difference. In short, A. B. Markman and Gentner's study showed that similarity judgments involve determining the best global alignment and are sensitive to relational structure.

The interactive nature of comparison processes and reasoning was nicely described by John Turner (1987):

> The assumption here is that categorization and comparison depend on each other and neither can exist without the other: the division of stimuli into classes depends upon perceived similarities and differences (comparative relations), but stimuli can only be compared in so far as they have already been categorized as identical, alike, or equivalent at some higher level of abstraction, which in turn presupposes a prior process of comparison and so on. (p. 46)

Turner's ideas anticipated the results of further research by A. B. Markman and Gentner (1991), showing the importance of alignment in determining commonalities and differences. They presented participants with pairs of concepts varying in their similarity and asked them either to list commonalities or to list differences between the concepts. As one might expect, the number of commonalities listed decreased systematically as the similarity of the concept pairs decreased. Surprisingly, however, the number of differences listed did not change at all as a function of the similarity of the pair. Subjects listed just as many differences for very similar pairs as for very different pairs. However, as similarity decreased, the nature of the differences changed from what one would call alignable differences to nonalignable differences.

A. B. Markman and Gentner (1991) argued that differences are more easily accessed when two concepts are aligned and that differences often emerge out of commonalities. For example, people list as a similarity of *hotels* and *motels* (a highly similar pair) that one can stay overnight in them; then they list as a difference that one usually stays for only one night in a motel but for multiple nights in a hotel. The difference is associated with an aligned similarity. In contrast, when comparing

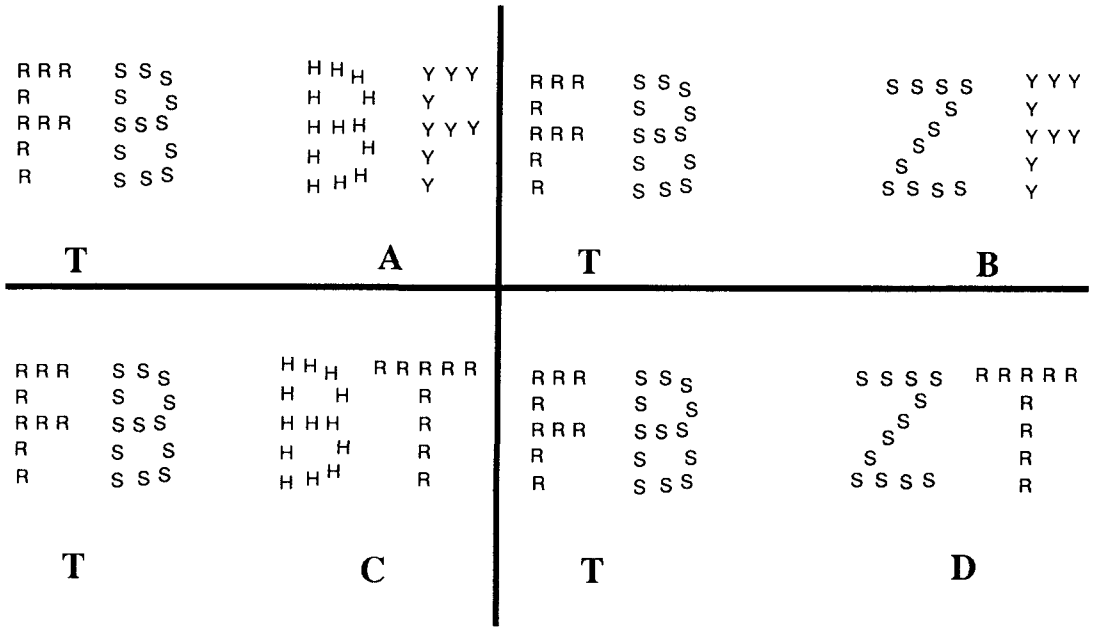

Figure 4. Another set of stimuli that produce Max effects. (A and D are judged to be more similar to T than B or C. That is, two global or two local matches are more effective than one local and one global match. From "Relational Similarity and the Nonindependence of Features in Similarity Judgments" by R. L. Goldstone, D. L. Medin, and D. Gentner, 1991, *Cognitive Psychology, 23,* p. 255. Copyright 1991 by Academic Press. Reprinted by permission.)

motels with a very different concept such as *hammers,* subjects do not tend to list aligned differences but tend to name things that apply to one item and not the other (e.g., one can stay in a motel but not a hammer and one can hit things with a hammer but not a motel).

Alignment works to fill in respects: Two scenes are similar with respect to features (both attributes and relations) that belong to matching scene parts. In a separate article (Goldstone & Medin, in press), we have dealt extensively with the alignment process for similarity comparisons and have described a computational model for alignment, SIAM (similarity, interactive activation, and mapping), based on interactive activation of local information to achieve global constraint satisfaction. SIAM successfully captures the relative contributions of aligned and nonaligned matches to similarity, including the increasing importance of aligned matches as processing time increases. SIAM has also received support for a number of other counterintuitive predictions; for example, SIAM correctly predicts that matching nondiagnostic features (features shared by all objects in scenes) increases the ease with which across-scene mappings are made (see Goldstone, 1991, and Goldstone & Medin, in press, for other examples). These observations underline the point that whether a feature match between two scenes will count as a feature match (and how much it will count) depends on whether the feature match belongs to corresponding (aligned) parts (see also Clement & Gentner, 1991).

Summary

The observations and experiments just described not only suggest that the similarity comparison process needs to be systematically studied, they also provide important constraints on respects. Although we are far from a full account of how similarity comparisons fix respects, it is clear that comparison is crucially involved in determining what properties will be accessed, inferred, or discovered and how properties will be weighted. Similarity is not a nondecomposable construct so much as it is a dynamic process. We now turn to some implications of these findings.

IMPLICATIONS

Views of Similarity

It may be useful to review the points of view with which we began this article. First, there is the position that the perceptual system provides an unchanging bedrock for similarity. There are three major problems with this view. One is that even within a single procedure for assessing similarity, performance may vary with factors such as processing time and previous experience. The second problem, discussed at length by L. B. Smith and Heise (1992), is that this position creates a dichotomy between perceptual and conceptual similarity and tacitly concedes that beyond hard-wired perceptual similarity, everything is conceptual. Furthermore, perceptual similarity may be stable only in contexts with relatively familiar, simple stimuli where the alignment process is straightforward. With more complex and unfamiliar entities, however, the simple story may not go through, and implicit assumptions about comparison processes may need to be made explicit. We concur with L. B. Smith and Heise (1992) in thinking that (perceptual) similarity should be more ambitious.

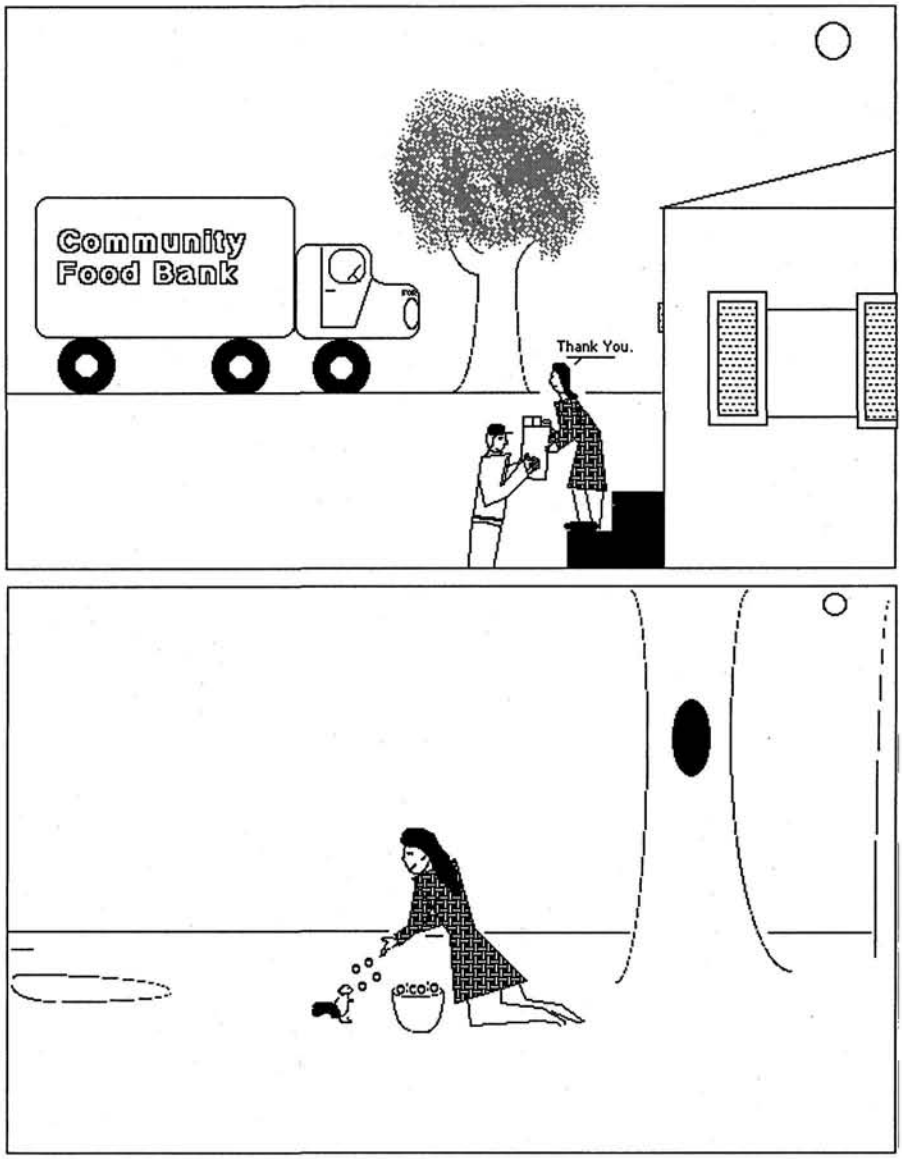

Figure 5. Sample stimuli from the A. B. Markman and Gentner (1990) experiments. (On superficial grounds, the woman in [a] should be matched with the woman in [b]. With respect to roles, however, the squirrel in [b] is the recipient and should be matched with the woman in [a], who is also the recipient. From "Analogical Mapping During Similarity Judgments" (p. 39) by A. B. Markman and D. Gentner, 1990, in *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society,* Hillsdale, NJ: Erlbaum. Copyright 1990 by Cognitive Science Society Incorporated. Reprinted by permission.)

The third major problem with viewing similarity as fixed is that it leads researchers to ignore the processing side of similarity. Our studies, as well as others, show similarity to be dynamic and context dependent. Experiment 1 used perceptual stimuli and showed that the properties of a given stimulus can vary as a function of what it is compared with. Experiment 2 used conceptual stimuli and demonstrated that the common properties that are instantiated between two stimuli depend on the direction of the comparison. Experiment 3 observed that similarity judgments depend on the respects that are highlighted by potential contrast sets. Finally, the series of studies by Goldstone et al. (1991) have indicated that the weight given to a particular perceptual property depends on the distribution of other properties present (the Max effect). Similarly, Clement and Gentner (1991) found that the importance of a predicate match depended on its connectivity to other matches. This adds up to a similarity that is too dynamic to be treated as fixed.

The second and third positions on similarity, namely, that it may vary with experience and with context, are steps in the right direction, in our opinion. These positions are consistent with the goal of identifying combinations of similarity structures and processing mechanisms that serve to fix respects. We

believe, however, that these views are incomplete because they have ignored the role of the comparison process in fixing respects.

Of course, one could broaden the definition of context to include comparisons and thereby encompass our results. The risk associated with an ever broader definition of context is that it will become a vague appeal when precision and attention to process is needed (E. R. Smith & Zarate, 1992, for a positive example of such precision). In fairness to those who emphasize the role of context, it must be said that the results of the first experiment do not discriminate between a general context effect and an effect tied to implicit or explicit comparison. On the other hand, the opposite pattern of results would also have been consistent with context effects. The second experiment would need to expand the notion context to include asymmetries. The third study was specifically organized around an anticipated context effect; note, however, that range-frequency theory, a model for context effects on judgment, did not capture the differential effects of context on antonyms, metaphors, and categorically related items. Nor can either the mapping results of A. B. Markman and Gentner (1990) or the temporal aspects of alignment described by Goldstone and Medin (in press) be captured by a general context effect. Although it is difficult to draw a sharp line between context effects and comparison-specific effects, we believe that our framework for comparison points us in the direction of particular process models.

The fourth position we mentioned at the beginning of this article is that similarity is so unconstrained that it may play little or no role in fixing respects. This position will remain viable for at least as long as the understanding of similarity is incomplete. At the same time, however, the present observations reduce the scope of this negative thesis by showing that respects are at least partially fixed by the similarity (comparison) process.

Our framework provides a fifth perspective on similarity. On a superficial level of analysis, one could take our review of respects as showing that similarity is completely unruly. That is, our results demonstrate that similarity, even perceptual similarity, is highly variable and comparison dependent. However, such a conclusion misses the central point. We have argued that similarity cannot be defined in a manner that ignores the processing side of similarity. Similarity judgments are highly variable but bound to the details of the comparison process.

Goodman's Thesis Revisited

Nelson Goodman (1972) called similarity a chameleon, but we believe that similarity is more like two yoked chameleons: The entities entering into a comparison jointly constrain one another and jointly determine the outcome of a similarity comparison. (That is, paradoxically, two chameleons may behave in a more orderly manner than one.) Thus, similarity is changeable and context dependent but systematically fixed in context. This systematicity allows a person to understand what other people mean when they say "X is similar to Y," even when they do not explicitly mention respects. Our studies show that people's judgments of similarity reflect constraints that serve to fix respects. In short, although the comparison process is dynamic, it is also lawful.

Goodman's thesis also lends itself to an implicit theory of similarity comparison that is inadequate. It seems to presuppose a set of independent features whose shared status may be affirmed when a similarity statement is made. This theory has at least three serious, related problems. First of all, it does not address the issue of how multiple respects are integrated to determine similarity, a critical aspect of the inference function of similarity. Second, it misses the distinction between attributes and relations and consequently ignores structure. Gentner and her colleagues (e.g., Gentner, 1983; Gentner & Rattermann, 1991) have, in contrast, stressed the importance of structure for guiding inferences. For example, consider what one might know about *quaggas* (some hypothetical or unfamiliar entity) from the statement "Quaggas are similar to zebras." Although we have no data bearing on this question, our intuitions are that people might be at least modestly confident that quaggas are hooved animals and not especially certain about whether or not they are striped. *Hooved* is (by conjecture) part of an interrelated set of properties associated with zebras, whereas *striped* seems more to be an isolated property. In any event, Goodman's framework would find the comparison completely uninformative until the respects were specifically mentioned. The third problem is that this view does not address the possibility of comparison-dependent and dynamically constructed features. In brief, although we believe that attention to comparison processes provides an important source of constraints on respects, we do not endorse the implicit theory of similarity associated with Goodman's challenge. (In fairness to Goodman, we should add that he might not endorse it either; his goal was simply to point out some of the vagaries of similarity statements.)

Implications for Similarity Theories

Although we have claimed that the comparison process serves to fix respects, we have not offered a specific computational or mathematical model to account for the full range of results described. A summary of the principle ideas associated with our comparison framework is as follows:

1. Similarity comparisons involve mutually constraining property instantiation and interpretations.
2. Similarity comparisons are informative and may be directional.
3. The respects associated with similarity assessments are influenced by the comparison context.
4. Similarity comparisons involve alignment driven by global constraint satisfaction.
5. The contribution of a match to similarity comparisons depends on the overall pattern of correspondences between entities.

Each of these statements can be directly or indirectly linked with related evidence. Experiment 1 demonstrated comparison-specific property interpretation. Experiment 2 showed that the common features associated with directional comparisons tend to be more closely associated with the base term than the target term. This observation is consistent with the idea that directional comparisons involve assertions or attempted predications (see also Ortony, 1979). Tversky's (1977) diagnosticity and extension effects, along with the results of our Experiment

3, show that comparison contexts serve to fix respects. The A. B. Markman and Gentner (1990) study (refer again to Figure 5) has provided direct evidence on global constraint satisfaction and alignment, and their observations on similarities and differences (A. B. Markman & Gentner, 1991) have yielded important indirect evidence for alignment (see also Goldstone, 1991). Finally, the Goldstone et al. (1991) Max effects and the Goldstone and Medin (in press) results on the differing contributions of aligned and nonaligned feature matches to similarity have clearly shown nonindependent feature weighting (see also Clement & Gentner, 1991).

These observations do not add up to a computational model, but they place important constraints on theories of similarity. Specifically, it seems clear that models that attempt to capture the process of alignment and comparison, such as SME, ACME (Analogical Constraint Mapping Engine), and SIAM, are promising candidates worthy of further development and testing. All three models rely on relational structure to achieve global constraint satisfaction, and, therefore, all three reject the idea of independent, additive features. In our opinion, what these models currently lack is a well-specified mechanism for feature construal and context- and comparison-specific feature construction (see Hofstadter & Mitchell, in press, for some interesting ideas along these lines). Ideally, there would be some level of abstraction intermediate between our general framework and our two computational implementations, SME and SIAM. The goal would be a general model with some allowance for task-specific implementational details. Still, we think that the combination of a framework and computational examples shows that when it comes to comparison processes, we are not proposing to replace silence with vagueness.

Our results are more striking from the perspective of geometric and featural models of similarity. These models do not address either relational structures or the processes by which correspondences are achieved (they are also silent about context- and comparison-specific feature construction). Consequently, they do not address many of the results described in this article. Tversky's (1977) original work on diagnosticity and the extension effect pointed to the significance of similarity processes, and the present article continues in this direction. Featural and geometric models have been and will continue to be very useful tools for psychologists, but they are limited in critical ways.

Is Similarity a Unitary Construct?

Are the various measures or uses of similarity more or less the same? One way to assess whether these alternative measures are getting at a unitary construct of similarity is to see whether or not they agree with one another. Although it would stray from our overall purposes to provide a general review, the evidence is mixed and should prompt caution. Tversky and Gati (1982) used a variety of measures in their studies of the coincidence hypothesis: the idea that the dissimilarity between two objects differing on two dimensions may be larger than would be predicted on the basis of their unidimensional differences. They found support for the coincidence hypothesis using judgments of similarity and dissimilarity, recognition memory errors, and classification and inference decisions as dependent variables. This result suggests that the various measures involve a common underlying component.

On the other hand, different measures of similarity often yield lower correlations than one would expect if the measures were solely an index of similarity plus some noise. Podgorny and Garner (1979) compared similarity ratings with same–difference reaction times and observed only a modest correlation (−.588). Sergent and Takane (1987) collected similarity judgments, derived a multidimensional scaling solution, and then attempted to use this solution to predict speeded same–difference judgments. They observed some systematic discrepancies between the two measures. Keren and Baggen (1981) noted that (unspeeded) similarity judgments did not accurately predict the pattern of confusions under speeded identification (see also Beck, 1966; Klein & Barresi, 1985). Dissociations have also been found in memory studies. For example, Rattermann and Gentner (1987) found that subjects' rated similarity of stories varied independently of the degree to which one story reminded them of the other.

One might take these observations as evidence that direct and indirect measures of similarity are fundamentally different. Before endorsing such a conclusion, one should note that different measures of the same type may not agree with each other. For example, the pattern of performance may vary as a function of processing time. Goldstone (1991) collected same–difference judgments under three different deadlines and observed that the correlation between reaction times at various deadlines decreased as a function of the difference in deadline. Ratcliff and McKoon (1989) found that attributional similarity is available earlier in processing than relational information (see also Goldstone et al., 1991). Studies comparing unspeeded similarity judgments and speeded reaction times vary both the type of measure and the processing time. Therefore, it is unclear how much of the difference depends on task as opposed to time pressure. What these observations do show is that similarity (as measured) varies systematically across conditions.

As a further complication, Medin et al. (1990) noted a divergence between two direct judgment measures. Although similarity judgments sometimes correlate very highly with dissimilarity judgments (e.g., −.98; Tversky, 1977), Medin et al. found a much reduced association. For example, Stimulus B in Figure 6 tended to be selected as more similar than Stimulus A to the standard T by people making similarity judgments and as more different by people making dissimilarity judgments. Overall, the correlations between similarity and difference judgments were −.67 and −.70 for two different stimulus sets.

The available data suggest that different measures of similarity are clearly significantly correlated and that they are also tapping different phenomena to some extent. Measures of similarity are not highly enough correlated that one could successfully argue that similarity is fixed or invariant. Different measures of similarity may engage different processes. In the same versus different judgment task, spotting one difference is sufficient to elicit a "different" response; it is unlikely that similarity ratings proceed by hunting for a single featural difference between compared items. Conversely, similarity rating tasks involve processes that are not required in a same versus different judgment task. For example, according to the range-frequency adjustment process (Krumhansl, 1978; Parducci, 1965), people

who are asked to use a numeric scale tend to (a) divide the range into a fixed number of intervals of equal breadth and (b) establish intervals that are used with equal frequency. Still, different measures of similarity are correlated. Although different tasks involve different processes, there would also appear to be shared components. A description of how measures are associated and dissociated will require an analysis of the processes and structures required for the tasks. We see no alternative to this empirical–theoretical prescription.

The experiments on which we have focused have involved direct judgments of similarity in almost every case. Again, we are agnostic about the generality of our results for contexts in which similarity is indirectly evaluated. For example, we do not know whether patterns of false alarms on new–old recognition memory tests will show any of the comparison-dependent processes described here (but there is evidence for strong directional asymmetries in memory comparisons [e.g., Agostinelli, Sherman, Fazio, & Hearst, 1986] and Max-like effects in patterns of false alarms [Tversky & Gati, 1982]). The key point is that one needs to see similarity as not only structure but also process, and a large majority of the work on similarity processes remains to be done.

Relation to Other Ideas

Our results are perhaps most compatible with research on analogy. Gentner and others have stressed the importance of alignment processes in understanding analogies, and there is mounting evidence that alignment is a central aspect of similarity comparisons (e.g., Goldstone, 1991; Goldstone & Medin, in press; A. B. Markman & Gentner, 1990). Although early work on analogy treated the representation of the base and target of an analogy as fixed, more recently ideas about comparison-dependent inferences and dynamic rerepresentation have been advanced (e.g., Gentner, 1989; Gentner & Rattermann, 1991; Hofstadter & Mitchell, in press). Again, the upshot of these innovations is to focus attention on the comparison process.

The present framework is also paralleled by findings in the literature on decision making. An early view of decision making was that preferences preexisted and were revealed by choices. There is mounting evidence, however, that preferences are often constructed during the judgment process itself (e.g., Simonson & Tversky, 1991; Tversky & Kahneman, 1986; Tversky, Sattath, & Slovic, 1988). For example, what is objectively the same situation may lead to different choices depending on how the problem is framed. We believe that closer analyses will reveal some close correspondences between findings on decision making and phenomena associated with similarity judgments.

By conjecture, the present framework might also extend to the categorization literature. Recently, researchers have attempted to integrate similarity-based learning with explanation-based or knowledge-driven learning. One approach has been to assume that knowledge selects and weights features from a preexisting pool of features (e.g., Lebowitz, 1986a, 1986b). Wisniewski and Medin (1991) argued that this approach is inadequate and that knowledge determines which features will be constructed and inferred, as well as how features are weighted (see also Barsalou, 1987, 1989, for further exposition of the idea that categorization is dynamic and constructive). Lee Brooks has also recently suggested that categorization at least partially determines which features are perceived or inferred to be present in the situation (Brooks, Allen, & Norman, 1991). The present results and framework are very compatible with these results on the role of knowledge in category learning.

What is common to each of these areas is a particular view of constructive processes. Constructivism is not a prescription for chaos but for giving proper attention to process. Although the attention given to representation and structure has been beneficial, one should never think that representations come out of the mind in pure form without any influence of process. Much of the orderliness seen may reflect orderliness of processing principles.

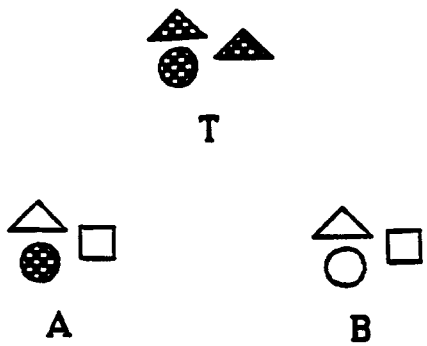

Figure 6. A sample stimulus set in which similarity and difference judgments are not mirror images of each other. (The standard T is attributionally similar to A because they both have a checkered circle. The stimulus B does not have this attributional match but instead shares a matching relation, "same shading," with T.)

Similarity as Ground

Does the fact that respects can be specified mean that similarity can ground other cognitive processes? There is no simple answer to this question. The observation that children seem to be less flexible about similarity than adults and that noun contexts bias even young children in the direction of object shapes suggests that early noun categories may be organized by similarity principles. The further suggestion that children tend to be holistic rather than analytic in their object perception (e.g., Kemler-Nelson, 1989) fits nicely with other evidence showing that children tend to assume that labels apply to whole objects (E. M. Markman, 1990). The cooperation of these constraints would lead to a coherent account of the acquisition of object categories, and similarity would be an important part of the story (e.g., see Gentner & Rattermann, 1991).

The picture with respect to complex perceptual categories involving actions, events, or concepts that are partially or not at all perceptual in character seems much more unclear. The constraints needed to specify respects may end up doing most of the explanatory work. What is needed is a systematic analysis of the conditions of learning in terms of what information is available in what contexts.

Recently, researchers have questioned the role of similarity in categorization. Rips (1989) argued against what he called the resemblance theory, the theory that "the object is a member of a category if it is sufficiently similar to known category members" (p. 21). In one of his experiments, subjects were told to imagine a 3-in. (7.62-cm) object and were asked whether it was more similar to a quarter or a pizza, and which it was more likely to be. Subjects considered the object more similar to a quarter but believed that it was more likely to be a pizza, presumably because of the greater variability in pizza sizes. Similarly, Gelman and Markman (1986) and Carey (1984) found that even children as young as 4 years make inductive inferences on the basis of category knowledge and not visual similarity. Given these results, to what extent does the comparison process discussed here influence categorization and induction?

These researchers treat similarity as physical similarity. Rips (1989) argued that if nonphysical resemblances are considered, then the force of the resemblance theory is weakened. If the possibility of abstract similarities is included, then resemblance theory is "cut from its moorings" and cannot serve as a constraint on categorization. However, we argue that if similarity is viewed as the outcome of a particular comparison process involving alignment, carryover of properties, and an active search for commonalities, then it is still constrained even if abstract properties are admitted into the analysis. Abstract commonalities are often mentioned in our subjects' comparisons, and we see no reason to limit similarity to physical similarity.

Whether or not the comparison process that we describe plays a key role in categorization is still an open question. We do have suggestive evidence for some parallels. In one experiment, one group of subjects judged Doberman pinschers to be more like raccoons than sharks. Other subjects judged Doberman pinschers to more likely be members of Group A {boar, lion, shark} than Group B {boar, lion, raccoon}. These results, like Rips's, are problematic for accounts in which a fixed, context-independent similarity computation determines categorization. Taken one category member at a time, Doberman pinschers are less similar to Group A animals than Group B animals. The results can, however, be explained by a comparison-specific interpretation process similar to the one we previously invoked. Doberman pinschers seem to be placed in Category A because all Category A animals and Doberman pinschers are ferocious. "Ferocious" becomes important because it emerges as a similarity between all three category members. Earlier, we found that similarity comparisons cause properties of one item to be carried over to the other item. This process would also explain how the presence of *boar* and *lion* would increase the likelihood of categorization on the basis of ferocity.

Basically, the same argument holds for the role of similarity in other cognitive functions. The agenda is to specify respects and then determine whether similarity will hold up under detailed examination of acquisition and use conditions. At present, we cannot provide a definitive answer to the question of how well similarity can support reasoning and other cognitive functions. We have shown that similarity is not the chaotic thing that its severest critics claim, and we have provided a basis for looking in the right place. Similarity is far from an empty concept with no explanatory power.

In fairness to similarity critics, it must be said that a chaotic similarity is only part of the problem. To the extent that similarity is fixed by goals, theories, and belief systems, it is these factors that have explanatory power. Our goal in this article has been to point to a major source of (previously unexplored) constraints on similarity rather than to resolve the issue of whether and how well similarity can ground other processes. If similarity is chaotic, it cannot ground anything; if similarity is not chaotic, then it may have a role to play. Frank Keil has suggested (personal communication, October 18, 1991) that similarity may take over where theories leave off. The general idea is that knowledge, theories, and belief systems do provide critical constraints but that they inevitably run into ignorance when the reasoner is forced to resort to heuristics and strategies to proceed. The heuristic function of similarity may do important work in learning and induction. From this perspective, a critical research agenda is to describe the integration and coordination of similarity-based and knowledge-driven processes.

Similarity as Comparison

Finally, to focus on similarity solely with respect to its ability to constrain other cognitive functions may be to miss the central point. Similarity is a comparison process that itself is a fundamental cognitive function. Similarity needs to be understood on its own, just as do other comparisons such as simile, metaphor, and analogy. From this perspective, similarity involves far more than a simple computation over a set of fixed features. Instead, similarity is always dynamic, is often inherently asymmetrical, and discovers and aligns features rather than just adding them up. Similarity has its own mysteries that we are only beginning to understand.

Conclusion

It is natural to assume that, to constrain similarity comparisons appropriately, the representation of each of the constituent terms must be rigid. In contrast, our observations suggest that the effective representations of the constituents are determined in the context of the comparison, not prior to it. It is as if the two terms were dancers: Each dancer may have a repertoire of stylistic preferences, but the actual performance depends on an interaction between the two. For asymmetrical comparisons, the "base dancer" takes the lead and the "target dancer" follows. The result is appropriately constrained even though the constituents are quite flexible.

Clearly, the framework we have provided for similarity comparison is more of an agenda than a set of answers. Nelson Goodman (1972) was correct to argue that respect for similarity requires specifying respects. It is incorrect to think that the answers to the respects question have little to do with similarity other than to expose it as an empty notion. Respects are determined by factors that are intrinsic to the comparison process. When researchers ask people "How similar are X and Y?" it is as if people are answering the subtly different question "How are X and Y similar?" That is, a critical aspect of similarity comparisons is the procedure for fixing respects.

References

Agostinelli, G., Sherman, S. J., Fazio, R. H., & Hearst, E. S. (1986). Detecting and identifying change: Additions versus deletions. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 445–454.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition, 10,* 82–93.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge, England: Cambridge University Press.

Barsalou, L. W. (1989). Intra-concept similarity and its implications for inter-concept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76–121). Cambridge, England: Cambridge University Press.

Beck, J. (1966). Effect of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics, 1,* 300–302.

Brooks, L. R., Allen, S. W., & Norman, G. R. (1991). The role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 279–287.

Brown, A. L. (1989). Analogical mapping and transfer: What develops? In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 369–412). Cambridge, England: Cambridge University Press.

Brown, A. L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science, 14,* 107–133.

Carey, S. (1984). Are children fundamentally different kinds of thinkers and learners than adults? In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Current research and open questions* (Vol. 2, pp. 485–517). Hillsdale, NJ: Erlbaum.

Carmichael, L. L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology, 15,* 73–86.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Clement, C. A., & Gentner, D. (1991). Systematicity as a selectional constraint in analogical mapping. *Cognitive Science, 15,* 89–132.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1990). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41,* 1–63.

Fillenbaum, S., & Rapoport, A. (1974). Verbs of judging judged: A case study. *Journal of Verbal Learning and Verbal Behavior, 13,* 54–62.

Fisher, M. A., & Zeaman, D. (1973). An attention-retention theory of retardate discrimination learning. In N. R. Ellis (Ed.), *International review of research in mental retardation* (Vol. 6, pp. 169–256). San Diego, CA: Academic Press.

Forbus, K. D., & Gentner, D. (1989). Structural evaluation of analogies: What counts? In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 341–348). Hillsdale, NJ: Erlbaum.

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology, 16,* 341–370.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23,* 183–208.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the nonobvious. *Cognition, 38,* 213–244.

Gentner, D. (1977). Children's performance on a spatial analogies task. *Child Development, 48,* 1034–1039.

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). Cambridge, England: Cambridge University Press.

Gentner, D., & Rattermann, M. J. (1991). The career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). Cambridge, England: Cambridge University Press.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science, 10,* 277–300.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

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

Goldstone, R. L. (1991). *Similarity, interactive activation, and mapping.* Unpublished doctoral dissertation, University of Michigan.

Goldstone, R. L., Gentner, D., & Medin, D. L. (1989). Relations relating relations. In *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). Similarity, interactive activation, and mapping: An overview. In K. Holyoak & J. Barnden (Eds.), *Advances in connectionist and neural computation theory, Vol. 2: Connectionist approaches to analogy, metaphor, and case-based reasoning.* Norwood, NJ: Ablex.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational Similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology, 23,* 222–262.

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

Hesse, M. B. (1963). *Models and analogies in science.* Notre Dame, IN: Notre Dame University Press.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hofstadter, D., & Mitchell, M. (in press). The copycat project: An overview. In K. Holyoak & J. Barnden (Eds.), *Advances in connectionist and neural computation theory, Vol. 2: Connectionist approaches to analogy, metaphor, and case-based reasoning.* Norwood, NJ: Ablex.

Hofstadter, D. R., Mitchell, M., & French, R. M. (1987). *Fluid concepts and creative analogies: A theory and its computer implementation* (Technical Report No. 87-1). Ann Arbor: University of Michigan, Fluid Analogies Research Group.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355.

Hutchinson, J. W., & Lockhead, G. R. (1977). Similarity as distance: A structural principle for semantic memory. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 660–678.

Keil, F. C. (1989). *Concepts, kinds and development.* Cambridge, MA: Bradford Books/MIT Press.

Keil, F. C., & Batterman, N. A. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23,* 221–236.

Kemler-Nelson, D. G. (1989). The nature and occurrence of holistic processing. In B. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 357–386). Hillsdale, NJ: Erlbaum.

Keren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. *Perception & Psychophysics, 29,* 234–246.

Klein, R. M., & Barresi, J. (1985). Perceptual salience of form versus material as a function of variations in spacing and number of elements. *Perception & Psychophysics, 37,* 440–446.

Kolstad, V., & Baillargeon, R. (1991). *Appearance- and knowledge-based responses to containers in infants.* Unpublished manuscript.

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

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3,* 299–321.

Lebowitz, M. (1986a). Concept learning in a rich input domain: Generalization-based memory. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: Vol. II* (pp. 193–214). Los Altos, CA: Morgan Kaufman.

Lebowitz, M. (1986b). Integrated learning: Controlling explanation. *Cognitive Science, 10,* 219–240.

Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior, 16,* 395–418.

Markman, A. B., & Gentner, D. (1990). Analogical mapping during similarity judgments. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 38–44). Hillsdale, NJ: Erlbaum.

Markman, A. B., & Gentner, D. (1991). Communalities, differences and the alignment of conceptual frames during similarity judgments. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 287–292). Hillsdale, NJ: Erlbaum.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14,* 57–77.

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

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

Melara, R. D., Marks, L. E., & Lesko, K. (1992). Optional processes in similarity judgments. *Perception & Psychophysics, 51,* 123–133.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

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

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

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and classification. *Perception & Psychophysics, 45,* 279–290.

Nosofsky, R. M. (1992). Exemplars, prototypes and similarity rules. In S. Kosslyn, R. Shiffrin, & A. Healy (Eds.), *Essays in honor of William K. Estes* (Vol. 1, pp. 146–167). Hillsdale, NJ: Erlbaum.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review, 86,* 161–180.

Ortony, A., Vondruska, R. J., Foss, M. A., & Jones, L. E. (1985). Salience, similes, and the asymmetry of similarity. *Journal of Memory and Language, 24,* 569–594.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review, 56,* 132–143.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200.

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.

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology, 21,* 139–155.

Rattermann, M. J., & Gentner, D. (1987). Analogy and similarity: Determinants of accessibility and inferential soundness. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 23–35). Hillsdale, NJ: Erlbaum.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382–407.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, England: Cambridge University Press.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45,* 1043–1056.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15,* 346–378.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1.* Cambridge, MA: MIT Press.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *Mechanization of thought processes* (pp. 511–526). London: Her Majesty's Stationery Office.

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.

Shepp, B. E. (1983). The analyzability of multi-dimensional objects: Some constraints on perceived structure, the development of perceived structure, and attention. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development* (pp. 39–73). Hillsdale, NJ: Erlbaum.

Simonson, I., & Tversky, A. (1991). *Context effects in choice.* Unpublished manuscript, University of California at Berkeley.

Sjöberg, L. (1972). A cognitive theory of similarity. *Göteborg Psychological Reports, 2*(10), 1–23.

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

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99,* 3–21.

Smith, J. E. K. (1980). Models of identification. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 129–158). Hillsdale, NJ: Erlbaum.

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, England: 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). New York: Elsevier.

Smith, L. B., & Kemler, D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology, 24,* 279–298.

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning.* San Diego, CA: Academic Press.

Thorndike, E. L. (1931). *Human learning.* New York: Century.

Turner, J. C. (1987). *Rediscovering the social group: A self-categorization theory.* Oxford, England: Basil Blackwell.

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

Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 79–98). Hillsdale, NJ: Erlbaum.

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

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* 251–278.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95,* 371–384.

Ward, T. B., Becker, A. H., Hass, S. D., & Vela, E. (1991). Attribute, availability, and the shape of bias in children's category generalization. *Cognitive Development, 6,* 143–167.

Watanabe, S. (1969). *Knowing and guessing: A formal and quantitative study.* New York: Wiley.

Wellman, H. M., & Gelman, S. A. (1988). Children's understanding of the nonobvious. In F. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 99–135). Hillsdale, NJ: Erlbaum.

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

Zeaman, D., & House, B. J. (1963). The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency* (pp. 159–223). New York: McGraw-Hill.

Received September 9, 1991
Revision received March 2, 1992
Accepted April 21, 1992 ■