Title
The Differential Effects of Causes on Categorization and Similarity

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 20(0)

Authors
Rehder, Bob
Hastie, Reid

Publication Date
1998

Peer reviewed

The Differential Effects of Causes on Categorization and Similarity

Bob Rehder (rehder@psych.colorado.edu)
Department of Psychology
University of Colorado, Boulder, CO 80309

Reid Hastie (reid.hastie@colorado.edu)
Department of Psychology
University of Colorado, Boulder, CO 80309

Abstract

Does categorization involve more than the similarity of an item to a category prototype or other category members? Rips (1989) argues yes, because categorization and similarity ratings sometimes diverge, indicating that they are based on different factors. However, Smith and Sloman (1994) suggest that such categorization/similarity dissociations may be limited to special conditions. We examined the effect of causal relationships between category attributes on categorization and similarity, and found that causal knowledge had a much larger effect on categorization than on similarity. This result was obtained with stimuli rich in characteristic attributes and without participants thinking aloud, that is, in just those conditions where Smith and Sloman found categorization to be solely similarity-based. Thus, the categorization/ similarity dissociation demonstrated by Rips is alive and well, and the need for an account of categorization that goes beyond similarity is again highlighted.

The abilities to categorize and note similarities among objects and events are central to virtually all cognitive processes, including reasoning, problem solving, decision making, and memory storage and retrieval. Not surprisingly then, researchers have tried hard to unveil the relationship between categorization and similarity. One possibility that has received empirical support is that categorization largely consists of similarity comparisons: An item will be placed in a category to the extent that it is similar to a category prototype or to other category members (Smith & Medin, 1981). This reduction of categorization to similarity, if true, is important because it suggests that the focus of research should be on similarity itself rather than on categorization.

However, there is also evidence that categorization cannot always can be reduced to similarity. For example, Rips (1989) found that while most undergraduates classified a novel item (e.g., a circular object with a 3-inch diameter) into a variable category (e.g., pizzas) rather than a fixed-size one (e.g., quarters), they judged the item to be more similar to the fixed-size category (see also Fried & Holyoak, 1984; Rips & Collins, 1993). Rips also found that changes to "essential" properties of natural kinds that are causally-central (e.g., level of maturation) influenced categorization more than similarity (for related research with children, see Carey, 1985; Gelman & Wellman, 1991; Keil, 1989). Because categorization ratings did not always agree with similarity ratings, these studies demonstrate a categorization/ similarity dissociation that implies that categorization can be based on something besides (or in addition to) similarity comparisons. For example, categorization may involve the application of rules ("if it flies, it's a bird", Anderson, Kline, & Beasley, 1979; Nosofsky, Palmeri, & McKinley, 1994; Pazzani, 1991), or general reasoning processes (Barsalou, 1983; Murphy & Medin, 1985), such as "inference to the best explanation" (Harman, 1986; Rips, 1989).

In a replication of Rips' (1989) study of the effect of variability on categorization and similarity, Smith and Sloman (1994) asked under what conditions similarity-versus rule-based categorization is triggered. They found no evidence of a categorization/similarity dissociation when the novel item included a feature characteristic of the fixed category (e.g., a circular object with a 3-inch diameter that is silver colored), or when their participants did not think aloud while categorizing (as in Rips' study) and were asked to respond "as quickly as you can while being as accurate as possible". These results suggest that cases of categorization that involve more than simple similarity comparisons may be rare, limited perhaps to circumstances where the categorizer feels the need to justify his or her decision, and only with artificially sparse stimuli.

Categorization is so central to all of cognition that it is crucial to determine under what conditions categorizers employ rules and explanatory/causal reasoning in addition to similarity relations. As discussed above, demonstrations of categorization/similarity dissociations have also involved causally-central essential attributes. In this study we tested whether the conditions that eliminated the dissociation in Smith and Sloman (characteristic features and no think-alouds) would also eliminate it with categories laden with inter-attribute causal relationships. We found that the presence of causal relationships affected categorization much more than similarity, even with stimuli rich in characteristic features and without participants thinking-aloud.

Method

Materials

Our goal was to create materials that described categories that could really exist. We constructed six categories: two biological kinds (Kehoe Ants, Lake Victoria Shrimp), two nonliving natural kinds (Myastars, Meteoric Sodium Carbonate), and two artifacts (Romanian Rogos, Neptune Personal Computers). Each category had four binary attributes, the values of which were either abnormal or normal relative to its superordinate category. For each attribute, the base rates were described as 75% for the abnormal value and 25% for the normal value, which are henceforth referred to as the *characteristic* and *uncharacteristic* values. The attributes for Kehoe Ants are shown in Table 1.

Table 1

	Attribute	Characteristic Value	Uncharacteristic Value
A1	Iron sulfate in blood	High	Normal
A2	Immune system	Hyperactive	Normal
A3	Consistency of blood	Thick	Normal
A4	Nest building	Fast	Normal

While participants in the *control condition* were given no other information about the category, those in the *common-cause* and *common-effect conditions* were also taught about three causal relationships arranged in the patterns shown in Figure 1. In the common-cause condition, one attribute, A1, was the cause of the other three attributes, while in the common-effect condition, one attribute, A4, was caused by the other three attributes. These causal patterns have been useful in past research (Waldmann, Holyoak, & Fratianne, 1995), because they manifest the asymmetries inherent in causal (as opposed to simply associative) relationships: While the effect attributes in the common-cause pattern (A2, A3, A3) should be correlated with one another (because they have a common cause), the cause attributes of the common-effect pattern (A1, A2, A3) should not be correlated.

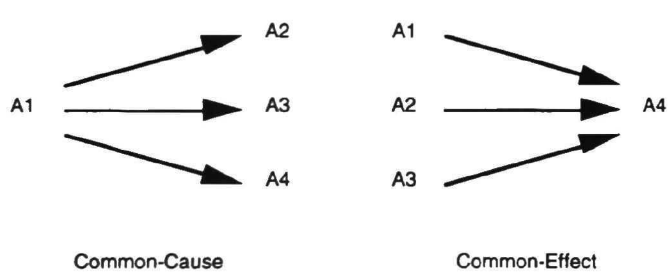

Figure 1

Each causal link was described as the characteristic value of one attribute causing the characteristic value of another attribute, and included some detail of the causal mechanism involved. For example, the cause linking A1 and A2 of Kehoe Ants was described as "Blood high in iron sulfate causes a hyperactive immune system. Iron sulfate molecules are detected as foreign by the immune system, and the immune system is highly active as a result".

Participants

252 University of Colorado undergraduates received course credit for participating in this experiment, and were randomly assigned in equal numbers to one of the six categories, and to either the common-cause, the common-effect, or the control condition.

Procedure

All phases of the experiment were conducted by computer. Participants first studied several screens of information about the category at their own pace. Participants in the control condition received a cover story, attributes, and attribute values and their base rates (75%/25% for the characteristic/uncharacteristic values). Participants in the common-cause and common-effect conditions also received a description of three causal relationships, and a diagram depicting those relationships. When ready, participants took a multiple-choice test that tested them on the knowledge they had just studied. Participants could request help which led the computer to re-present the information about the category. Participants were required to retake the test until they committed 0 errors and made 0 requests for help.

Participants then performed three tasks (counterbalanced for order): a categorization task, a similarity rating task, and an induction task. (Space limitations prevent us from reporting the induction results, which are not discussed further.) During the categorization task, participants rated the category membership of exemplars on a 20-point scale. There were 32 exemplars, consisting of all possible 16 examples that could be formed from four binary attributes, each presented twice. The attribute values of each exemplar were presented in order (1-4) on the computer screen.

During the similarity task, participants rated the similarity of two exemplars on a 9-point scale. Participant rated 32 pairs of exemplars randomly generated subject to the constraint that no pair was repeated and the pair never consisted of the same exemplar. No feedback was provided during either the categorization or similarity tasks.

Results

We report results of the categorization task and the similarity task. To facilitate comparisons, ratings from both tasks were first converted to a 0-100 scale.

Categorization Results

The categorization results were analyzed by performing a multiple regression for each participant. Four predictor variables (a_1, a_2, a_3, a_4) were coded as -1 if the attribute value was uncharacteristic, and +1 if it was characteristic. The regression weight associated with each a_i represents the influence of that attribute on a participant's categorization ratings. Positive weights will result if participants make use of the 75%/25% base rates for characteristic and uncharacteristic values, that is, if a characteristic value increases the categorization rating and an uncharacteristic value decreases it. An additional six predictor variables were constructed by computing the multiplicative interaction between each pair of attributes: a_1a_2, a_1a_3, a_1a_4, a_2a_4, a_3a_4, and a_2a_3. The resulting interaction terms are coded as -1 if one of the attributes has an uncharacteristic value and the other a characteristic value, and +1 if the values are both characteristic or both uncharacteristic. These interaction terms can be considered second-order configural properties like those assumed by feature-frequency models of categorization (Hayes-Roth & Hayes-Roth, 1977). For those attribute pairs on which a causal relationship was defined, the configural property can be interpreted as representing whether the causal relationship is confirmed (+1, cause and effect both present or both absent) or violated (-1, cause present and effect absent, or cause absent and effect present). The regression weight associated with such a configural property is the influence that confirmation or violation of a causal relationship had on categorization ratings, with a positive weight indicating that confirmation leads to a higher categorization rating and violation to a lower one.

The regression weights averaged over participants in the common-cause, common-effect, and control conditions are presented in Figure 2. As expected, in all conditions the weights given to the four attributes (a_1, a_2, a_3, a_4) were positive, indicating that participants used the base rate information when generating categorization ratings. In addition, in the common-effect condition, the common-effect (a_4) had greater weight (8.11) than the three causes (7.02, 5.64, and 4.32 for a_1, a_2, and a_3, respectively). In the common-cause condition, the common-cause (a_1) carried greater weight (8.54) than the three effects (3.94, 4.19, and 3.82 for a_2, a_3, and a_4). Note that in the control condition the first attribute also had greater weight (8.62) than attributes 2-4 (6.77, 6.81, and 6.52), an effect we believe is due to the first attribute always appearing first in an exemplar's list of attribute values. To some extent, the greater salience of the first attribute explains the greater weight associated with the common-cause in the common-cause condition. However, the difference between a_1 and a_2, a_3, a_4 is greater in the common-cause condition (4.56) than in the control condition (2.20), so the greater weight of the common-cause cannot be explained solely by the greater salience of the first attribute.

These conclusions are supported by statistical analysis. We carried out a 3 (causal pattern condition: common-cause, common-effect, control) by 6 (category) by 3 (order of category task: first, second, or third) by 4 (attribute: 1, 2, 3 or 4) analysis of variance on the regression weights, with repeated measures on the last factor. The pattern of weights for the four attributes differed significantly between the common-cause and common-effect conditions (significant interaction between the common-cause/common-effect contrast and the attribute factor: $F(3, 696)=11.50$, $MSE=22.3$, $p<.0001$), and the pattern of weights in each of these conditions differed significantly from the control condition ($F(3, 696)=2.67$, $MSE=22.3$, $p<.05$ for common-cause, $F(3, 696)=6.14$, $MSE=22.3$, $p<.001$ for common-effect). In the common-cause condition, the weight associated with the common cause was significantly greater than the weights of the three effect attributes ($F(1, 76)=40.73$, $MSE=43.5$, $p<.0001$). In the common-effect condition, the weight associated with the common effect was significantly greater than the weights of the three causes ($F(1, 76)=20.47$, $MSE=12.4$, $p<.0001$).

Figure 2 also demonstrates that causal knowledge affected the weight given to configural properties. The configural properties corresponding to those attribute pairs assigned causal relationships had substantial positive weights in both the common-cause condition (4.66, 4.15, and 3.87 for a_1a_2, a_1a_3, and a_1a_4), and the common-effect condition (3.02, 2.89, and 3.21 for a_1a_4, a_2a_4, and a_3a_4). In other words,

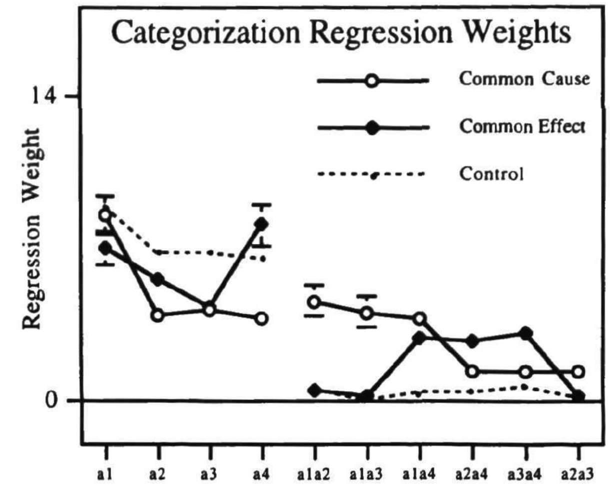

Figure 2

categorization ratings were higher if a pair of attributes confirmed a causal relationship that participants had learned, and lower if the pair violated that relationship. A 3 (causal pattern) by 6 (category) by 3 (order of category task) by 6 (configural property) ANOVA on the regression weights revealed that the pattern of configural property weights in the common-cause and common-effect conditions differed from one another ($F(10, 1160)=26.79$, $MSE=10.0$, $p<.0001$) and each differed from the control condition ($F(5, 1160)=10.56$, $MSE=10.0$, $p<.0001$; $F(5, 1160)=7.65$, $MSE=10.0$, $p<.0001$). Weights of the configural properties corresponding to the three causal relationships differed significantly from the control condition (all p's<.0001), in both the common-cause and common-effect conditions.

A further result of note is that in the common-cause condition the regression weights for the interactions between the effect attributes (1.39, 1.47, and 1.48 for a_2a_4, a_3a_4, and a_2a_3) were significantly different from the control condition ($p<.05$). In other words, categorization ratings were higher if two effect attributes were both uncharacteristic or both characteristic, and lower if one was uncharacteristic and the other characteristic. A property of common-cause causal patterns is that effect attributes will be correlated, because of their common cause. In fact, participants in the common-cause condition exhibited this knowledge by weighting the correlation between the effects when making categorization decisions. This is evidence that they were treating the causal relationships as more than simple associations and engaging in a form of causal reasoning. In contrast, the common-effect pattern of causal relationships does not imply the existence of correlations between cause attributes, and, in fact, in this study the parameter estimates for these interaction terms (a_1a_2, a_1a_3, and a_2a_3) did not differ significantly from the control condition (all F's<1). This result replicates other findings that people take into account the asymmetries inherent in causal relationships (Waldmann & Holyoak, 1992; Waldmann et al., 1995).

We asked whether the pattern of weights shown in Figure 2 was representative of most participants. Of the 168 common-cause and common-effect participants, 5% exhibited the pattern of attribute weights shown in Figure 2 but not the pattern of configural feature weights, 13%

exhibited the pattern of configural feature weights but not the pattern of attribute weights, and 20% exhibited both. 45% exhibited a similarity-based, or "family resemblance", strategy in which there was no discernible use of the causal knowledge. Thus, the group means in Figure 2 do not represent the categorization strategy used by most participants, i.e., substantial individual differences existed.

Similarity Results

The similarity results were also analyzed with a multiple regression per participant. Four predictors (ma_1, ma_2, ma_3, ma_4) were formed by multiplying a_i of the first exemplar with a_i of the second, the result being that each ma_i was coded as -1 if the values of attribute i in the two exemplars mismatched (i.e., one uncharacteristic and one characteristic) and +1 if they matched (i.e., both characteristic or both uncharacteristic). Six predictors (ma_1a_2, ma_1a_3, ma_1a_4, ma_2a_4, ma_3a_4, and ma_2a_3) were formed by multiplying a_ia_j of the first exemplar with a_ia_j of the second, the result being that each ma_ia_j was coded as -1 if the configural property in the two exemplars mismatched and +1 if they matched. Matches (either between attributes or configural properties) that increase the rated similarity of two exemplars will manifest themselves with positive regression weights.

Figure 3 presents the regression weights averaged over participants in the three causal conditions. The pattern of weights given to matches of attributes (ma_{1-4}) mirrors the attribute weights found for the categorization results (Figure 2). In the common-cause condition, whether the common-cause attributes matched in the two exemplars (ma_1) influenced similarity ratings more (weight of 14.24) than matches on the three effect attributes (weights of 11.86, 11.50, and 9.86 for ma_{2-4}). In the common-effect condition, matches between the common-effects (ma_4) had greater weight (11.76) than matches on the three cause attributes (11.13, 10.10, and 9.83 for ma_{1-3}). Statistical analysis confirmed that the pattern of weights differed significantly between the common-cause and common-effect conditions ($F(3, 696)=11.88$, $MSE=16.5$, $p<.0001$) and each of these conditions differed from the controls ($F(3, 696)=5.15$, $MSE=16.5$, $p<.005$; $F(3, 696)=2.30$, $MSE=16.5$, $p<.10$). In the common-cause condition, the weight associated with matching common causes was significantly greater than effect matches ($F(1, 76)=13.72$, $MSE=58,2$, $p<.0005$). In the common-effect condition, the weight associated with matching common effects was significantly greater than the cause matches ($F(1, 76)=32.36$, $MSE=18.4$, $p<.0001$).

The weights for matches between configural features (Figure 3) were small and not appreciably different from the control condition. While the pattern of these weights differed significantly between the common-cause and common-effect condition ($F(5, 1160)=4.64$, $MSE=7.5$, $p<.0005$), neither of these conditions differed significantly from the control condition ($F(5, 1160)=1.46$, $MSE=7.5$, $p=.21$ for common-cause; $F(5, 1160)=1.72$, $MSE=7.5$, $p=.13$ for common-effect). Thus, while the causal relationships that participants learned affected similarity ratings by changing the weight, or salience, of matches between attributes, matches between configural features that code whether a causal relationship was confirmed or violated had no effect on similarity ratings.

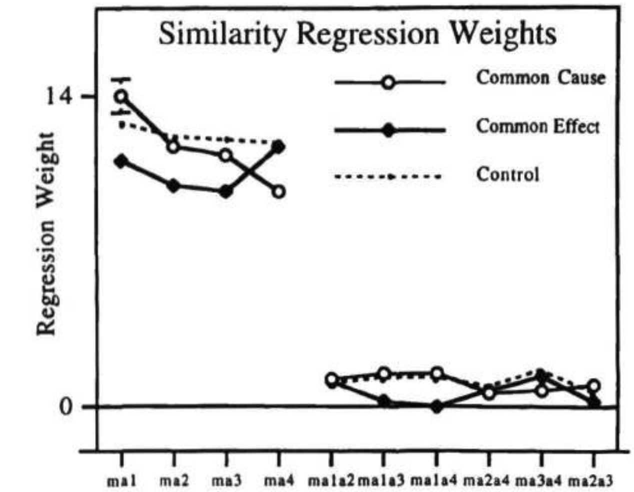

Figure 3

A Categorization/Similarity Dissociation

We have shown that categorization ratings, but not similarity ratings, are affected by whether causal relationships are confirmed or violated in presented exemplars. We illustrate this result with two examples. Figure 4a presents the categorization ratings of 0111 and the similarity of 0111 to the category prototype 1111 (1 meaning the characteristic attribute value, and 0 the uncharacteristic value) in the common-cause and control conditions. The presence of common-cause causal relationships resulted in 0111 being given a much lower categorization rating (45.2) relative to the control condition (70.9) because, presumably, in 0111 all three common-cause causal links are violated (the cause is absent but all three effects are present). However, the presence of these causal relationships had little effect on the similarity of 0111 to the prototype relative to the control condition (69.4 vs. 72.5).

Analogously, in Figure 4b the presence of common-effect causal links had a large effect on the categorization rating of exemplar 1110 relative to the control condition (53.5 vs. 74.5) because, presumably, 1110 violates all three of the causal relationships (all three causes are present but the effect is absent). But, the presence of those causal relationships had little effect on the similarity of 1110 to the

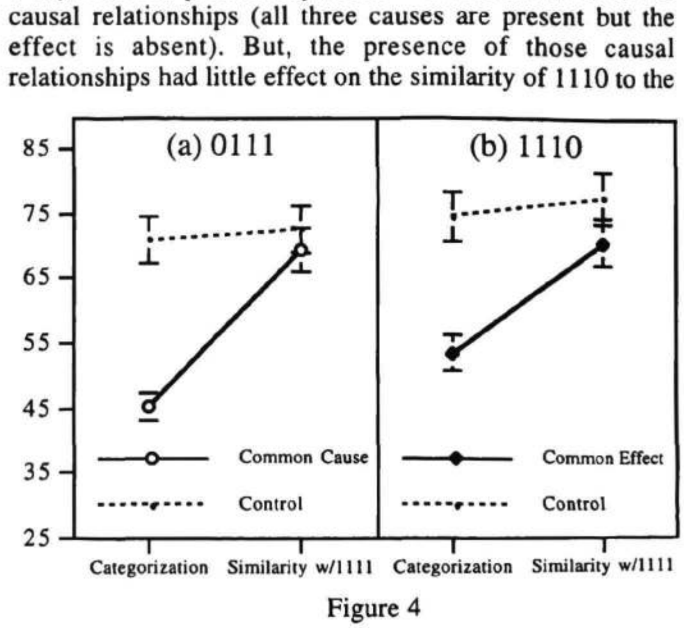

Figure 4

prototype relative to the controls (70.9 vs. 77.7).

Note that we did not directly ask our participants to rate the similarity of each exemplar *to the category*, but rather only to other category members. However, assuming that similarity to the category prototype can be taken as similarity to the category, then Figure 4 demonstrates a categorization/similarity dissociation, because causal relationships had a large effect on categorization ratings, but little on similarity ratings. In the Discussion section we return to the issue of whether similarity to the category prototype is the same as similarity to the category.

Discussion

The present study demonstrates that categorization and similarity are not influenced in the same way by the presence of causal relationships between category attributes. On the one hand, causal relationships led participants to weigh attributes similarly in the categorization and similarity rating tasks, specifically, weighing the common-cause most heavily in the common-cause condition, and the common-effect most heavily in the common-effect condition. On the other hand however, while categorization ratings were dependent on whether the causal relationships were confirmed or violated (that is, on what we have called "configural properties"), the confirmation or violation of causal relationships had no effect on the perceived similarity of two exemplars. Thus, causal knowledge about categories appears to influence categorization processes through means other than changing similarity relations.

The results of Smith and Sloman (1994) raised the possibility that categorization/similarity dissociations such as those reported by Rips (1989) might only occur in rarefied circumstances, specifically, with sparse stimuli lacking characteristic attributes, and when participants are required to think-aloud. Contrary to this view, in the present study categorization and similarity ratings were differentially influenced by causal knowledge even when exemplars had many characteristic values (e.g., three in the exemplars 0111 and 1110 of Figure 4, compared to Smith and Sloman's one), and without our participants thinking-aloud.

There are two ways in which the categorization/similarity dissociation presented in Figure 4 might be explained in terms of similarity comparisons. First, dissociations that are intended to show that tasks employ different component processes are not decisive, because dissociations may also arise when one task is more resource demanding than the other (Dunn & Kirsner, 1988). In this study, configural properties may have not been used in the similarity task because of resource limitations. The similarity task involved comparing two exemplars, and hence 6 configural properties needed to be computed, whereas only one exemplar was presented in the categorization task and thus only 3 configural properties needed to be computed. Assuming a similarity model that accounts for both similarity ratings between category exemplars and categorization judgments (by computing the similarity between an exemplar and some category representation, Lassaline & Murphy, 1998), the dissociation shown in Figure 4 could be accounted for if (a) the causal relations between attributes contributed to perceived similarity, but (b) there was a limit to the number of relations that could contribute due to resource limitations (of working memory, for example). Similarity models such as SME (Markman & Gentner, 1993; 1996) or SIAM (Goldstone, 1994; Goldstone & Medin, 1994) may be promising starting points for such an account.

Second, the dissociation presented in Figure 4 is between categorization ratings and similarity with the category prototype. Conceivably, similarity to the category prototype may not be the same thing as similarity to the category, and similarity with the category may track (and hence possibly explain) categorization ratings. For example, in the context model of categorization (Medin & Schaffer, 1978), similarity to a category is the sum of similarities to category exemplars stored in memory rather than similarity to the prototype. On the face it, the context model appears inapplicable to the current experiment because our participants observed no exemplars of the category. However, the context model could be adapted to the current situation by assuming that our participants encoded the conceptual information about the categories (i.e., the attribute base rates and causal relationships) by mentally constructing what they thought was a set of representative exemplars. (See Heit, 1994, for a similar strategy for representing background knowledge by the presence of prior exemplars.) It is a virtual certainty (because the number of free parameters would exceed the number of data points) that some set of constructed exemplars (combined with a set of context model similarity parameters) would fit the current categorization data. Although this would provide an account of categorization in terms of similarity comparisons, the explanatory burden would then fall on *why* participants chose to represent the category with that particular set of constructed exemplars, and part of the answer would surely involve the presence of causal relationships.

The categorization results presented here replicate those of Rehder and Hastie (1997). That study was identical to the present one except that participants were exposed (in the guise of a classification-with-feedback training task) to exemplars of the category that manifested the 75%/25% base rates but *not* the correlations among attributes implied by the common-cause or common-effect causal patterns. That study, together with the current one in which participants observed *no* exemplars of the category, demonstrate that causal relationships need not manifest themselves as correlations that inhere in observed exemplars to have an effect on categorization decisions (for related results see Malt & Smith, 1984; Murphy & Wisniewski, 1989).

Ahn and Lassaline (1996) have proposed that causes have more influence than effects on categorization. The current results from the common-effect condition in which the common effect was more heavily weighted than the three cause attributes do not support this hypothesis. Perhaps more importantly, the current results emphasize that causal knowledge affects not just the weight given to *attributes*, but also the weight given to *configurations of attributes* that either confirm or violate that knowledge. There is still much we don't know about which configurations people consider to be consistent with prior causal knowledge. Our use of two-way interactions in the regression analyses presupposes that the causal relationships were viewed as "necessary and sufficient" by our participants, that is, an attribute value pair violates the relationship either when the cause is absent and the effect present (a violation of necessity), or when the cause is present and the effect absent (a violation of sufficiency). In fact, people may consider an attribute pair to be inconsistent with a causal relationship only when necessity is violated, or only when sufficiency is violated. Further research will be required to determine which notions of confirmation and violation people typically employ.

Finally, although our results highlight the differences between categorization and similarity, we are nonetheless in agreement with those who argue that similarity almost always plays at least some role in categorization (Allen & Brooks, 1991; Smith & Sloman, 1994). Of those participants in this study who learned causal relationships between category attributes, 45% made no use of those relationships and generated classification ratings on the basis of simple family-resemblance. And, when the causal knowledge was used, it was usually in addition to, rather than instead of, the attribute base rate information. Thus, a full account of human categorization needs to include processes sensitive to both similarity and causal relations.

References

Ahn, W., & Lassaline, M. E. (1996). *Causal structure in categorization: A test of the causal status hypothesis (Part I)*. Unpublished manuscript.

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of the explicit rule. *Journal of Experimental Psychology: General, 120*, 3-19.

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

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*, 211-227.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review,95*, 91-101.

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

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

Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 3-28.

Goldstone, R. L., & Medin, D. L. (1994). Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 29-50.

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: The MIT Press.

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

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1264-1282.

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

Lassaline, M. E., & Murphy, G. L. (1998). Alignment and category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 144-160.

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

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*, 431-467.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition, 24*, 235-249.

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

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

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberchien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications*. Chichester, England: Ellis Horwood.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*, 53-79.

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 416-432.

Rehder, B., & Hastie (1997). The roles of causes and effects in categorization. In M. G. Shafto & P. Langely (Eds.), *The Nineteenth Annual Conference of the Cognitive Science Society* (pp. 650-655). Stanford, CA.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. New York: Cambridge University Press.

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General, 122*, 468-486.

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

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22*, 377-386.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*, 222-236.

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General, 124*, 181-206.