

# When similarity and causality compete in category-based property generalization

BOB REHDER

*New York University, New York, New York*

Five experiments were performed to investigate the category-based generalization of *nonblank* properties, properties that were novel but that were attributed to existing category features with causal explanations. Experiments 1–3 tested how such explanations interact with the well-known effects of similarity on such generalizations. The results showed that when the causal explanations were used, standard effects of typicality (Experiment 1), diversity (Experiment 2), or similarity itself (Experiment 3) were almost completely eliminated. Experiments 4 and 5 demonstrated that category-based generalizations exhibit some of the standard properties of causal reasoning; for example, an effect (i.e., a novel category property) is judged to be more prevalent when its cause (i.e., an existing category feature) is also prevalent. These findings suggest that category-based property generalization is often an instance of causal inference.

Knowledge of categories is essential to many cognitive functions, including reasoning, explanation, and communication. Also important is the fact that knowledge of categories can help one acquire yet more knowledge. For example, upon experiencing an insect bite whose venom produces a stinging sensation, a person might conclude that all members of that insect species have venom that produces a stinging sensation. In other words, on the basis of one's knowledge of categories (a species of insect), an isolated episode (a bite that stings) gets turned into a general fact about the world. Knowledge thus begets more knowledge by enabling specific experiences to be generalized.

A substantial literature has developed documenting the conditions under which people are likely to make category-based generalizations of this kind (for reviews, see Heit, 2000; Rips, 2001). To start with, note that such generalizations consist of three components: one (or more) *base example*, the *target category*, and a *novel property* that is displayed by the base and that may be generalized to the target. Of course, whether an individual believes that the generalization is warranted will depend on the specific beliefs he or she already has relating all three of these components. For example, even young children know that some properties of an insect (e.g., "has venom that produces a stinging bite") are potentially true of the insect's species generally but that others (e.g., "killed by a pesticide on Tuesday") are not—or at least, this second property will generalize to a very different category (Gelman, 1988). Thus, in order to focus on the relationship

between the base and the target categories, early research into category-based generalizations used *blank properties*, properties about which people have no (or at least little) prior knowledge. For example, in a typical experiment, participants might be told that "sparrows have sesamoid bones," where "sesamoid bones" is the blank property. They would then be asked to judge the strength of a generalization (e.g., "all birds have sesamoid bones").

This line of research has led to a number of prominent findings regarding the conditions under which blank properties are generalized. *Typicality effects* refers to the fact that more typical base examples support stronger generalizations. For example, people will be more confident that all birds have sesamoid bones when the base category is sparrows rather than penguins, because sparrows are more typical birds than are penguins. *Diversity effects* refers to the fact that a more diverse set of base examples leads to stronger generalizations. All else being equal, people are more likely to conclude that all birds have sesamoid bones, given that the sparrows, hawks, and chickens do, as compared with sparrows, robins, and bluejays, because sparrows, hawks, and chickens are more different from one another than are sparrows, robins, and bluejays.<sup>1</sup>

Two formal models have been proposed to account for typicality, diversity, and numerous other influences on category-based generalizations. First, according to the *similarity coverage model* (SCM; Osherson, Smith, Wilkie, López, & Shafir, 1990), people generalize by retrieving from memory typical examples of the target category and then computing the average maximal similarity between those examples and the base example(s). In other words, generalizations are stronger to the extent that the features of the premises "cover" the features of the target category. In addition to coverage, the second component of the SCM—similarity—allows it to account for gener-

---

I thank Reid Hastie, Douglas L. Medin, and three anonymous reviewers for their comments on earlier versions of the manuscript. Correspondence concerning this article should be addressed to B. Rehder, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003 (e-mail: bob.rehder@nyu.edu).

alizations between categories that are not hierarchically nested.<sup>2</sup> Second, Sloman's (1993) *feature-based induction model* (FBIM) accounts for many of the same effects as the SCM (including typicality, diversity, and similarity between nonhierarchically nested categories) but does so in terms of a single similarity process (the degree of overlap between features of the base categories and the target). Although the SCM and the FBIM differ in this and other important ways, the fundamental predictions of both are derived from the similarity relations that obtain between the base and the target categories.

Although the SCM and FBIM provide comprehensive frameworks in which to understand the role of similarity in the generalization of blank properties, subsequent research has confirmed the suspicion that led to use of blank properties in the first place—namely, that generalizations will be affected by the ways in which the new property is meaningfully related to the base and target categories. For example, Heit and Rubinstein (1994) demonstrated that the projection of a property from one category to another depends on whether the causal mechanisms thought to be responsible for the property in the base are also present in the target. They found that the projection of a biological property (e.g., a liver with two chambers) from bears to whales was stronger than from tuna to whales, presumably because whales are more likely to share biological mechanisms with other mammals than with fish. This preference was reversed for a behavioral property (e.g., travels in a zigzag path), apparently because whales are thought more likely to share a survival behavior with other prey animals in the same ecology (i.e., tuna) than with a predator (i.e., bears; see also Murphy & Ross, 1999; Ross & Murphy, 1999; Smith, Shafir, & Osherson, 1993).

The importance of causal explanations shared between premises and conclusions has been demonstrated by Sloman (1994, 1997), who found that, for example, people showed greater confidence in the conclusion “many war veterans are hired as bodyguards,” given the premise “many ex-cons are hired as bodyguards,” than they were in “many war veterans are unemployed,” given “many ex-cons are unemployed,” presumably because of the presence of a shared explanation in the first argument (i.e., both war veterans and ex-cons are experienced fighters) and the absence of one in the second.

Although these studies have confirmed that category-based generalizations are influenced by causal explanations in addition to similarity, there nevertheless remain numerous questions regarding how explanations promote generalizations. What are the properties of those explanations that result in especially confident generalizations? And how do explanations interact with the better understood influences of similarity? Regarding this second question at least, research has suggested that causal explanations not only influence generalizations, but also may largely undermine, or supplant, similarity-based processes. For example, López, Atran, Coley, Medin, and Smith (1997) studied category-based generalizations among the Itzá Maya (an indigenous population in central Guatemala) and found that the Itzá failed to exhibit

standard diversity effects, frequently appealing to specific causal mechanisms instead (e.g., judging that the prevalence of a disease in a species depended on the mechanisms that might spread that disease). American undergraduates, in contrast, exhibited standard typicality and diversity effects on the same items (see also Bailenson, Shum, Atran, Medin, & Coley, 2002). Although these findings could be attributed to some cultural difference other than domain knowledge, the dominance of causal explanations also holds with culturally homogenous groups. Proffitt, Coley, and Medin (2000) found that North American tree experts, instead of exhibiting standard typicality and diversity effects, reasoned causally about how diseases might spread among trees (see also Bailenson et al., 2002; Shafto & Coley, 2003). These studies raise the possibility that, in the presence of domain knowledge, similarity-based considerations may become irrelevant to how people choose to generalize events.

This article consists of two parts. In the first part, the relative importance of the effects of similarity and causality on category-based property generalization is assessed. Although the correlational research just reviewed has been highly suggestive, there is always the possibility that the results arise due to factors other than domain knowledge. In contrast, in the present study, a purely experimental approach was used in which participants from the same population were first taught about new categories and then were presented with category members that displayed novel properties that were accompanied by a description of the causal factors that led to their presence (hereafter referred to as *nonblank properties*) or with ones without such explanations (blanks).

An example of an experimental category used in this study and its characteristic features and novel properties are presented in Table 1. For example, the participants were told about a species of ant called *Kehoe Ants* and about five features characteristic of those ants. The fact that these features were characteristic was conveyed to the participants by telling them that each of them appeared in 75% of *Kehoe Ants*. The participants were then presented with a series of trials in which a particular *Kehoe Ant* displayed a novel property and were asked to estimate what proportion of all *Kehoe Ants* had that property. On some trials, the novel property was nonblank—that is, it was accompanied by a causal explanation that tied it to one of the ant's characteristic features. For example, the participants would be presented with a *Kehoe Ant* with the novel property “has venom with a stinging bite” and then additionally would be told that the *Kehoe Ant's* first feature,  $F_1$  (“high amounts of iron sulfate in the blood”), was the cause of this new property: “The stinging sensation is caused by the high concentration of iron sulfate in the venom.” Table 1 also lists the nonblank novel properties that were described as being caused by the other four characteristic features of *Kehoe Ants* ( $F_2$ ,  $F_3$ ,  $F_4$ , and  $F_5$ ).

On other trials, the novel property was blank—that is, no causal explanation regarding why the *Kehoe Ant* possessed the novel property was provided. The blank novel properties that served as controls for the nonblanks are

**Table 1**  
**Features and Nonblank and Blank Novel Properties for the Kehoe Ants Experimental Category**

	Feature	Nonblank Property	Blank Property
F <sub>1</sub>	Blood high in iron sulfate	Has a venom that gives it a stinging bite. The stinging sensation is caused by the high concentration of iron sulfate in the venom.	Has choroidal parasites attached to its eyes.
F <sub>2</sub>	Hyperactive immune system	Has an elevated body temperature. The high body temperature is caused by the hyperactive immune system.	Attacks poisonous Juniper insects.
F <sub>3</sub>	Thick blood	Is immobile in cold weather. The immobility in cold weather is caused by the thick blood.	Has an exoskeleton coated with ionized calcium.
F <sub>4</sub>	Fast nest building	Leaves behind a trail of sticky residue. The residue trail is caused by their speeded fluid production that allows them to build nests quickly.	Has antennae with three lateral splices.
F <sub>5</sub>	Slow digestive system	Has an enlarged stomach. The enlarged stomach is caused by its slow digestive system.	Becomes temporarily motionless when exposed to high ozone levels.

shown in the last column of Table 1. For example, the participants' propensity to generalize "has venom with a stinging bite" on the basis of its causal connection with feature F<sub>1</sub> (iron sulfate) was compared with their propensity to generalize "has choroidal parasites attached to their eyes," which had no causal connection with F<sub>1</sub> (or any of the other characteristic features).

Among other things, the approach of providing knowledge as part of the experimental session allowed the similarity-based effects that are the explicandum of the SCM and FBIM and the presence or absence of causal explanations to be manipulated independently. In each of the first three experiments, a two-factor design was used in which one factor was the presence or absence of an explanation and the second factor was typicality (Experiment 1), diversity (Experiment 2), or similarity itself (Experiment 3). The design of these experiments thus allowed a demonstration of the standard similarity-based effects in the absence of causal explanations and also of how those effects are moderated when such explanations are present.

To foreshadow the results, Experiments 1–3 showed that when causal explanations are used by participants, those explanations almost entirely eliminate the effects of similarity on category-based generalization. Having thus documented the dominance of causal explanations, the second part of this article will break new ground by assessing which aspects of those explanations lead to stronger category-based generalizations. For example, as the cause of the property "has venom with a stinging bite," what are the properties of "high amounts of iron sulfate in the blood" (e.g., the causal relationships it participates in with yet other features, or its prevalence in other category members) that influence the strength of the corresponding generalization? This question was taken up in Experiments 4 and 5.

## EXPERIMENT 1

The goal of the first experiment was to assess how the presence of a causal explanation for a novel property moderates the effect of typicality on category-based general-

izations. Each participant learned about two categories with four characteristic features. (The feature on the fifth dimension shown in Table 1 was used in a subsequent experiment.) For one of the categories, the novel properties were all blank, and for the other, they were all nonblank. For example, 1 participant might be presented with the nonblank properties of Kehoe Ants and the blank properties of another experimental category (e.g., Lake Victoria Shrimp), or vice versa.

In addition, each novel property was associated with category members that varied in their number of characteristic features. Of course, each of those category members possessed the characteristic feature associated with the novel property (e.g., the Kehoe Ant novel property "has venom with a stinging bite" always appeared with a Kehoe Ant that had "high amounts of iron sulfate"; see Table 1). This ensured that the purported cause of the nonblank novel property was present in each of the category members. But to vary the typicality of those exemplars, the values on the other three dimensions were systematically varied.

On the basis of the SCM and FBIM, the prediction is that the generalization of blank novel properties will strengthen with increases in the number of typical features. The critical question concerned whether typicality would also affect the generalization of novel properties with causal explanations. Note that in order to provide typicality with a greater chance to influence the generalization of nonblanks, those properties were associated with a completely different category (and, hence, were presented in a different block of trials) than were the blanks. This meant that the exemplars presented on nonblank trials differed from one another only on the basis of their typicality, a situation likely to highlight that information and, hence, promote its use.

## Method

**Materials.** Six novel categories were used: two biological kinds (Kehoe Ants and Lake Victoria Shrimp), two nonliving natural kinds (Myastars and Meteoric Sodium Carbonate), and two artifacts (Romanian Rogos [a type of automobile] and Neptune Personal Computers). Each category had four binary features, which were described

as distinctive relative to a superordinate category. The base rate for each feature was described as 75%—for example, “75% of Kehoe Ants have blood high in iron sulfate, whereas 25% have blood with normal amounts of iron sulfate.” A consequence of the 75% base rates is that the categories possessed a standard *family resemblance* structure in which no one feature defined category membership. Throughout this article, the presence of a characteristic (i.e., 75%) feature will be denoted with a “1” and its absence with a “0.”

Associated with each category were four blank and four nonblank novel properties. Whereas the blank properties had no obvious causal connection with the four characteristic features, each of the nonblank properties was accompanied by an explicit explanation of how it was causally generated by one of the characteristic features. The characteristic features and the blank and nonblank novel properties for one biological kind are presented in Table 1; those for the remaining five categories are available from the author.

**Participants.** Thirty University of Colorado undergraduates received course credit for participating in this experiment.

**Design.** There were two within-subjects factors. The first was novel property type: Each participant learned about two different categories, one with blank novel properties and the other with nonblank novel properties. The second was the typicality of the exemplar with the novel feature.

The pair of categories presented was varied as a between-subjects factor: A participant learned the two biological kinds, the two nonliving natural kinds, or the two artifacts. Which of the two categories the nonblank novel properties were presented with was counter-balanced over participants, as was whether the category with the nonblank properties was presented first or second.

**Procedure.** Experimental sessions were conducted by computer. For each of the two categories, the participants first studied several screens of information that included the category’s cover story, a description of its characteristic features, and the 75% base rates of those features. They then took a multiple-choice test that tested them on this knowledge. During the test, the participants could request help, in which case the computer re-presented the screens of information about the category. However, the participants were required to retake the test until they committed no errors and made no requests for help.

After performing the multiple-choice test, the participants were presented with a property induction task and a categorization task for that category. During the induction task, the participants were given 32 induction trials. Each trial presented a category member that possessed a novel property, and the participants were asked to estimate what proportion of all the category members possessed that property. The category member’s features were listed on the computer screen in dimension order (i.e., 1–4), followed by the sentence “this X also has Y,” where X was the category name and Y was the novel property. For nonblank properties, the causal explanation was presented in a second sentence underneath the first. Responses were entered by positioning a slider on a scale whose ends were labeled “None” and “All.” These responses were recorded as ratings on a 100-point scale, where 0 meant that no other category members possessed the novel property and 100 meant that they all did.

The 32 trials were divided into four blocks of 8 trials each. Each block presented one of the four novel properties associated with the category (either blanks or nonblanks, depending on the category). Within a block, the category member that displayed the novel property was varied. For the novel property associated with dimension  $i$ , the eight category members presented were those that possessed the feature on dimension  $i$ . For example, for Kehoe Ants, the novel features “a venom with a stinging bite” and “has choroidal parasites attached to their eyes” were always presented with the eight distinct Kehoe Ants that had blood high in iron sulfate, formed by systematically varying the values on the three remaining binary dimensions. The order of presentation of the 8 trials was randomized within a block for each participant, as was the order of presentation of blocks (i.e., novel properties).

During the categorization task, the participants rated the category membership of 32 exemplars, consisting of all possible 16 objects that could be formed from four binary attributes, each presented twice. Whether the categorization task followed the induction task or vice versa was randomized for each participant.

Note that this same categorization task was presented in all the experiments reported in this article. Its purpose was to serve as a manipulation check to determine whether the participants were sensitive to the manipulation of feature base rates. In fact, in agreement with other studies conducted with these materials (Rehder, 2003a, 2003b; Rehder & Hastie, 2001, 2004), in all the experiments the participants’ categorization ratings exhibited strong sensitivity to the base rate manipulations. For example, in this first experiment, the categorization results exhibited a strong typicality effect, since ratings increased monotonically with increases in the number of characteristic features. Because the results of the categorization task were not otherwise related to the theoretical issues in this article, they will be omitted for conciseness.

## Results and Discussion

Initial analyses of the induction results revealed that there were no effects of the between-subjects variables, such as whether the induction task preceded or followed the categorization task or whether the nonblank properties (i.e., those with causal explanations) were presented before or after the blanks. There also was no effect of whether the two categories were biological kinds, nonliving natural kinds, or artifacts. The results are therefore presented collapsed over these factors in Figure 1A as a function of whether a causal explanation was provided for the novel property and of the typicality of the exemplar in terms of the number of characteristic features it contained. As was expected, induction ratings for blank properties increased as the exemplar’s typicality (i.e., its number of characteristic features) increased, indicating that this experiment replicated the standard typicality effect for the generalization of blank properties. However, Figure 1A indicates that this effect of typicality was reduced when a causal explanation was provided.

To assess these conclusions statistically, a  $2 \times 4$  repeated measures ANOVA was performed, with typicality and property type as factors. There was a main effect of typicality [ $F(3,87) = 50.16$ ,  $MS_e = 174$ ,  $p < .0001$ ], a main effect of property type [ $F(1,29) = 13.65$ ,  $MS_e = 309$ ,  $p < .001$ ], and a significant interaction between the two [ $F(3,87) = 11.95$ ,  $MS_e = 76$ ,  $p < .0001$ ], confirming that the effect of typicality was reduced when a causal explanation was provided. Nevertheless, a separate analysis of the nonblank properties revealed that typicality still had a significant impact on induction ratings even for nonblank properties [ $F(3,87) = 18.21$ ,  $MS_e = 110$ ,  $p < .0001$ ].

One important question is whether the response pattern shown in Figure 1A was manifested consistently by all 30 participants or whether it arose as a result of averaging over individuals with substantially different response profiles. In fact, two groups of 15 participants with qualitatively different responses were identified. The performance of these two groups is shown in Figures 1B and 1C. The group in Figure 1B produced higher induction ratings for the nonblank properties than for the blanks; that is, these participants were more willing to generalize a novel property when it was accompanied by a causal

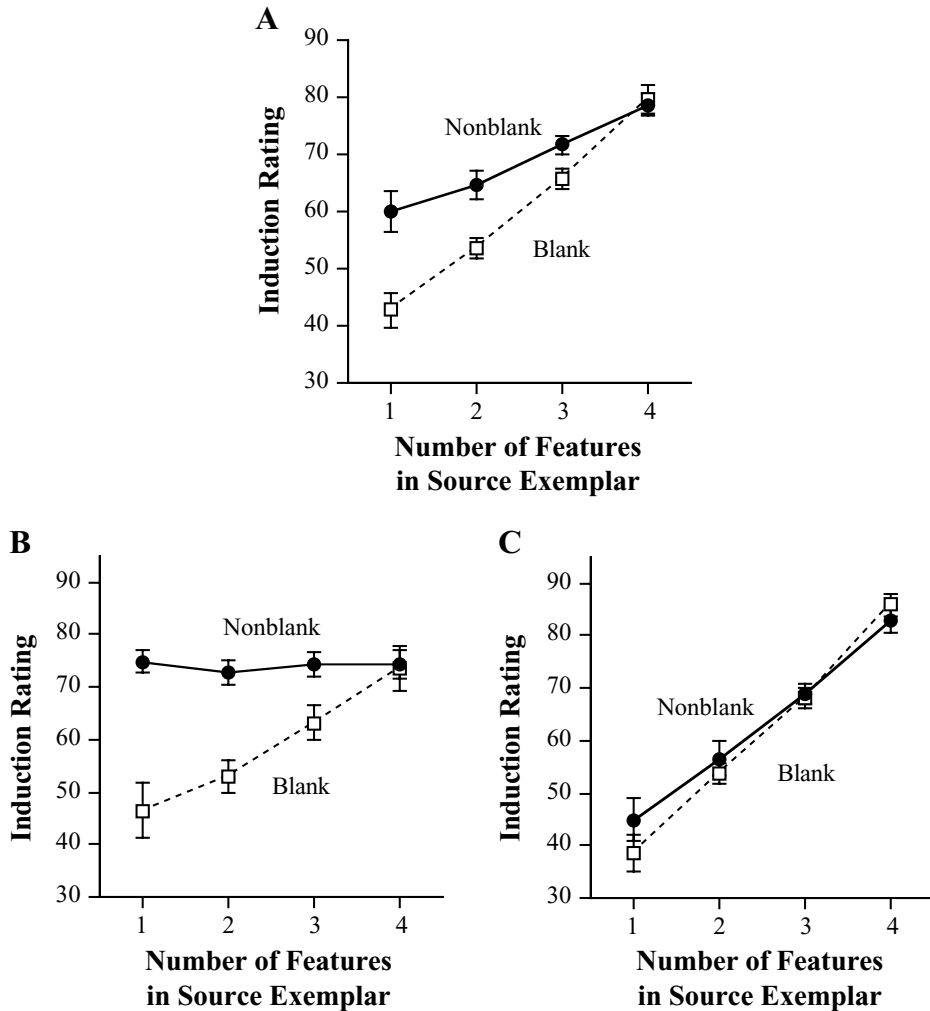


Figure 1. (A) Average induction ratings for the 30 participants of Experiment 1. (B and C) Two subgroups of 15 participants each. Error bars represent standard errors of the means.

explanation. Importantly, however, whereas ratings for blank properties were (as was predicted) sensitive to typicality, typicality had no effect on the generalization of the nonblanks. In contrast, the second group in Figure 1C did not generalize nonblanks more strongly than blanks, and both types of properties were equally sensitive to typicality. In other words, when the reasoners chose to base their responses on the causal explanations (as evidenced by the nonblank’s higher ratings in Figure 1B), there was no effect of typicality; when they chose to base their responses on typicality, there was no effect of causal explanations (Figure 1C). Apparently, the use of a causal explanation versus typicality is an all-or-none matter, with reasoners using one strategy or the other, but not both.

## EXPERIMENT 2

The SCM and FBIM predict that more typical exemplars support stronger inductive generalizations because

they “cover” more of the category. More coverage is also achieved with multiple category members that are diverse—that is, that have few characteristic features in common. Experiment 2 was performed to assess how the presence of a causal explanation for a novel property would moderate the diversity effect. Diversity was manipulated by presenting two category members with five stimulus dimensions (instead of four, as in Experiment 1; see Table 1) with the same novel property. The two category members exhibited either low diversity (i.e., they shared all five features) or high diversity (they shared only one feature). For example, the participants might be presented with two Kehoe Ants that “have venom with a stinging bite.” Both exemplars would have feature  $F_1$ , “high amounts of iron sulfate in the blood,” because iron sulfate was described as the cause of the stinging venom (Table 1). But to vary the diversity of the two exemplars, the values on the other four dimensions were manipulated. For example, in the low-diversity condition, the two ex-

emplars might be 11100 and 11100 (all features in common), whereas in the high-diversity condition they might be 11100 and 10011 (only  $F_1$  in common). Note that the two exemplars were chosen so that both always had three features, so that their typicality was held constant across low- and high-diversity trials.

As in Experiment 1, whether the novel property had a causal explanation was manipulated orthogonally. Following previous studies, the prediction was that the generalization of blank novel properties would be stronger when displayed by more diverse pairs of exemplars. The critical issue addressed in Experiment 2 concerned the effect of causal explanations on the diversity effect.

## Method

**Materials.** The materials used were the same as those in Experiment 1, with the exception that each category had five stimulus dimensions, instead of four (see Table 1).

**Participants.** Twenty-four New York University undergraduates received course credit or pay for participating in this experiment.

**Design.** There were two within-subjects factors. The first was novel property type: Each participant learned two categories, one with blank novel properties and the other with nonblank novel properties. The second was the diversity of the two exemplars with the novel feature. As in Experiment 1, there were three between-subjects factors: which pair of categories were learned, which of those categories had the nonblank properties, and whether the nonblank properties were presented first or second.

**Procedure.** The procedure had the same general format as that in Experiment 1. The participants learned about two categories and judged the generalization of nonblank properties for one and of blank properties for the other. During the induction task the participants were given 10 induction problems, in which they were presented with two category members with the same novel property, and were asked to estimate what proportion of all the category members possessed that property. Each category member was presented on the computer screen with its features listed in dimension order (i.e., 1–5). They were then followed by a sentence describing the novel property that both displayed (and for nonblank properties, the causal explanation for that property).

The 10 trials were divided into five blocks of 2 trials each. Each block presented one of the five novel properties associated with the category. Within the block that presented the novel property associated with dimension  $i$ , each category member possessed the feature on dimension  $i$ . The values on the other four dimensions were randomly chosen, subject to the constraint that three features were present and one was absent, and so that on one of the block's 2 trials the two category members were maximally diverse (had the same value on only dimension  $i$ ), whereas on the other trial they were maximally similar (had the same value on all five dimensions). The order of the 2 trials was randomized, as was the order of presentation of blocks (i.e., novel properties). The categorization task was also presented for each category, and the order of the induction and categorization tasks was randomized for each participant.

## Results and Discussion

Once again, initial analyses revealed no effect of the between-subjects variables. The results are thus collapsed over those factors in Figure 2A as a function of whether the novel property had a causal explanation and whether the exemplar pair exhibited low or high diversity. As in Experiment 1, novel properties were more strongly projected when they were accompanied by a causal explanation (75.0) than when they were not (59.1). In addition,

blank properties were more strongly projected when displayed by high-diversity exemplar pairs (63.7) than when displayed by low-diversity pairs (54.6), a finding consistent with previous research. However, when the novel property was accompanied by a causal explanation, there was no effect of diversity.

A  $2 \times 2$  repeated measures ANOVA revealed an effect of property type [ $F(1,23) = 17.46$ ,  $MS_e = 189$ ,  $p < .0001$ ], an effect of diversity [ $F(1,23) = 6.62$ ,  $MS_e = 350$ ,  $p < .05$ ], and an interaction between the two [ $F(1,23) = 10.95$ ,  $MS_e = 210$ ,  $p < .0001$ ]. A separate analysis of the nonblank properties revealed no effect of diversity ( $F < 1$ ), confirming that the diversity effect was eliminated when a causal explanation was provided.

As in Experiment 1, I asked whether the results shown in Figure 2A were manifested by all 24 participants or whether they arose from averaging over individuals with different response patterns. Once again, two groups were detected, consisting of 11 participants each (the responses of the remaining 2 participants were idiosyncratic). The performance of these two groups is presented in Figures 2B and 2C. The group in Figure 2B produced higher induction ratings for the nonblank properties than for the blanks. However, this group showed no effect of diversity for either blanks or nonblanks. That is, this group showed no sensitivity to the diversity manipulation. The group in Figure 2C, in contrast, showed sensitivity to the diversity manipulation, because it was more willing to generalize a blank property when it was displayed by a more diverse pair of category members. When it came to nonblanks, however, not only were such properties generalized more strongly, but also the effect of diversity disappeared.

The results in Figure 2C indicate how, just as in Experiment 1, the presence of a causal explanation can eliminate the effect of a similarity-based heuristic (in this case, diversity). One account of these findings is that explanations direct attention away from features that normally would contribute to how well a category is "covered" by the exemplar(s) displaying a novel property. When this occurs, the typicality or diversity of those exemplars becomes irrelevant to the generalization.

## EXPERIMENT 3

Experiments 1 and 2 presented undergraduates with source exemplars with a novel property, and the participants were asked to generalize that property to the whole category. However, the SCM and FBIM also predict that generalizations between items that are not hierarchically nested will be determined by their similarity. To this end, in Experiment 3, the participants were asked to generalize a property from one category member to another. Two factors were manipulated. This first was the similarity of the base and the target exemplars. The categories in Experiment 3 had four stimulus dimensions (as in Experiment 1); each source exemplar possessed three features, and each target exemplar possessed two. But the target exemplar's two features were varied so that it shared either one or three features with the base. An example is

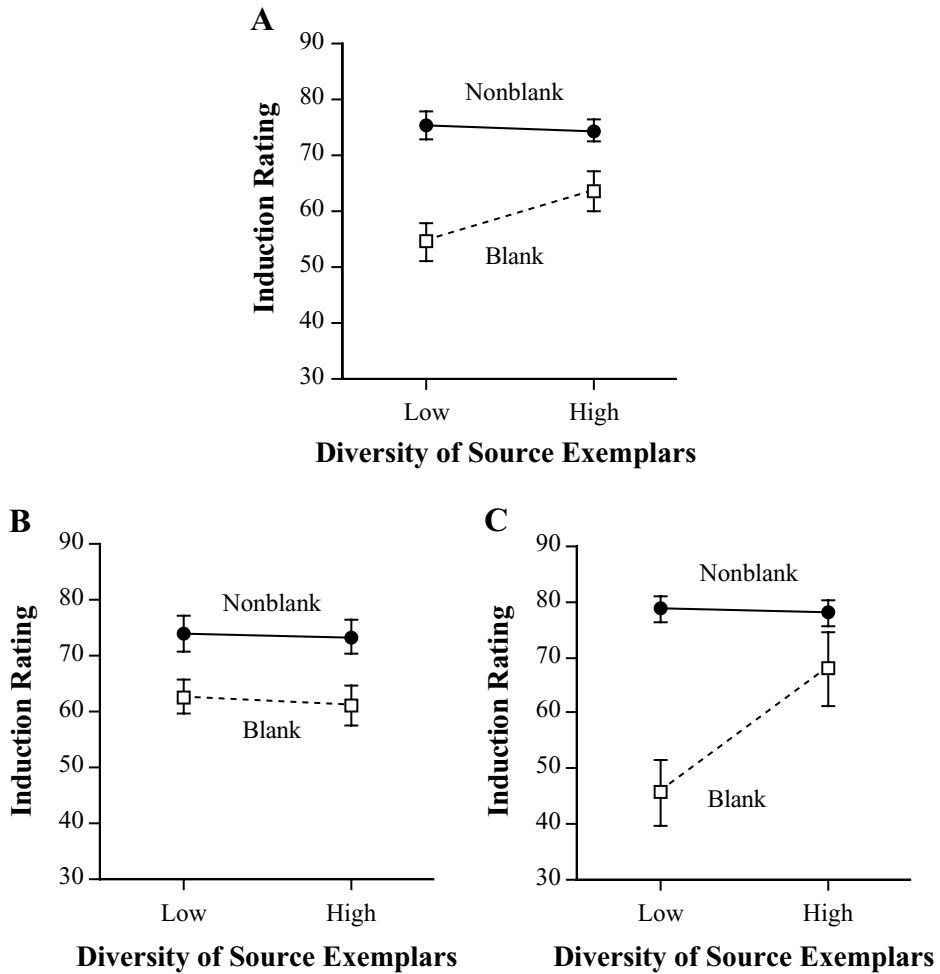


Figure 2. (A) Average induction ratings for the 24 participants of Experiment 2. (B and C) Two subgroups of 11 participants each. Error bars represent standard errors of the means.

presented in Table 2. Suppose that the source exemplar is 1110 (i.e.,  $F_1$ ,  $F_2$ , and  $F_3$  present,  $F_4$  absent). Then, 1100 is a high-similarity target exemplar, because it has the same value on three of the four dimensions, and 1001 is a low similarity target, because it has the same value on only one of the four dimensions. According to the SCM and FBIM, generalization should be stronger to the high- (e.g., 1100) than to the low-similarity target (e.g., 1001). Note that an important characteristic of Experiment 3’s design was that the typicality of both exemplars was held constant: The base and the target always had three and two features, respectively.

As in Experiments 1 and 2, the other factor manipulated was whether a causal explanation was provided for the novel property. In Experiment 3, however, an important question concerning nonblank properties was whether its purported cause appeared in the target exemplar. For example, suppose that the novel feature associated with source exemplar 1110 was described as caused by  $F_1$  (e.g., for Kehoe Ants, “a venom with a stinging bite” is caused by iron sulfate in the blood). Then, the lower part

of Table 2 presents both low- and high-similarity targets that possess the cause feature  $F_1$  (1100 and 1001) and ones that do not (0110 and 0101).

In Experiment 3, the participants learned about two categories, and explanations were provided for the novel properties for one category, but not for the other. For blank properties, the prediction was that the generalization would be stronger for high- versus low-similarity targets.

Table 2  
Example Target Exemplars From Experiment 3

Condition	Similarity to 1110	
	High	Low
Blank	1100	1001
Nonblank		
Cause in target	1100	1001
Cause not in target	0110	0101

Note—Similarity (high or low) is defined with respect to the source exemplar 1110. In this example, the novel property is assumed to be caused by  $F_1$  in the nonblank condition.

The critical question concerned the influence of similarity on nonblank properties.

### Method

**Participants.** Twenty-four New York University undergraduates received course credit or pay for participating in this experiment.

**Design.** There were two within-subjects factors. The first was the similarity of the target exemplar to the source exemplar. The second was whether the novel properties were blank or nonblank and, if nonblank, whether the cause was present in the target example or not. In addition, the same three between-subjects counterbalancing factors as those used in Experiments 1 and 2 were present in Experiment 3.

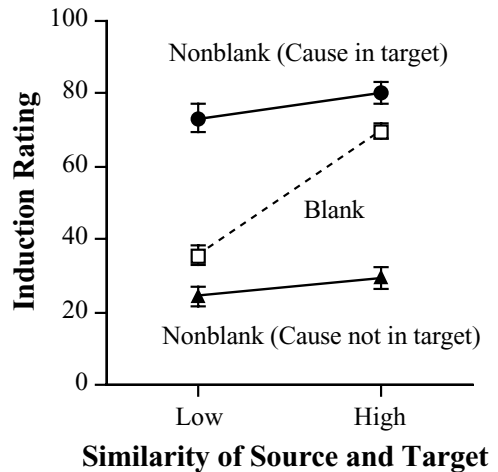
**Procedure.** The procedure had the same general format as those in the previous two experiments. The participants learned about two categories and answered questions about blank novel properties for one and about nonblanks for the other. During each category's induction task, the participants were given 16 induction problems in which they were presented with one category member (the source) that possessed a novel property (as in Experiment 1). They were asked to estimate how likely it was that a second category member (the target) also possessed that property. The target's features were presented on the computer screen underneath the source's. The participants entered their rating on a sliding scale with the left and right ends labeled "Very Unlikely" and "Very Likely," indicating the likelihood that the target also possessed the novel property.

The 16 trials were divided into four blocks of 4 trials. Each block presented one of the four novel properties associated with the category. For the novel property associated with dimension  $i$ , the source exemplar had the feature on dimension  $i$  and two other features chosen at random (giving it a total of three features). Within a block, the source exemplar was held constant, and the similarity of the target to the source was varied: On 2 of the 4 trials within a block, it exhibited high similarity (three shared features) with the source, whereas on the other 2 it exhibited low similarity (one shared feature). Whether the target exemplar possessed the cause feature was manipulated orthogonally: On 2 of the 4 trials, the cause feature was present in the target exemplar, but it was absent on the other 2. The order of presentation of these four different trial types was randomized within a block, as was the order of presentation of blocks (i.e., novel properties). The categorization task was also presented for each category, and the order of the induction and categorization tasks was randomized.

### Results and Discussion

Again, there were no effects of the between-subjects factors, and the results are therefore presented in Figure 3 as a function of the similarity between the base and the target exemplars and whether the novel property was blank or nonblank (and if nonblank, whether the cause feature was present or absent in the target). As was predicted, blank properties were more strongly projected when the target exemplar shared three features with the base exemplar than when it shared just one. Also, as was expected, nonblank properties were more strongly projected when the cause of that property appeared in the target exemplar than when it did not. The important finding is that the projection of nonblank properties was much less sensitive to the similarity of the base and the target exemplars.

A  $2 \times 3$  repeated measures ANOVA confirmed a main effect of similarity [ $F(1,23) = 41.36, MS_e = 205, p < .0001$ ], a main effect of novel property type [ $F(2,46) = 75.81, MS_e = 391, p < .0001$ ], and a significant interaction [ $F(2,46) =$



**Figure 3.** Average induction ratings from Experiment 3. Error bars represent standard errors of the means.

25.23,  $MS_e = 125, p < .0001$ ], indicating that the influence of similarity was greatly reduced for a nonblank property versus a blank one. Nevertheless, a separate  $2 \times 2$  ANOVA of the nonblank properties revealed that, in addition to a large main effect of cause present/absent [ $F(1,23) = 91.78, MS_e = 646, p < .0001$ ], there was a significant effect of similarity [ $F(1,23) = 7.20, MS_e = 121, p < .05$ ]. That is, similarity continued to exert an influence on the projection of nonblank properties. Note, however, that the magnitude of that similarity effect (an average difference of 6 points on a 100-point scale) was greatly attenuated, as compared with that for blanks (a difference of 34 points). In this  $2 \times 2$  analysis, the interaction between similarity and property type did not approach significance ( $F < 1$ ).

As in Experiments 1 and 2, I asked whether the group-level data presented in Figure 3 was manifested by all the participants. In this experiment, however, there was no evidence of subgroups with qualitatively distinct patterns of responses, since all the participants except 1 exhibited the pattern shown in Figure 3. Once again, it appears that the effect of causal explanations is to draw attention away from features not involved in the explanation, with the result that the two exemplars' similarity becomes largely irrelevant to how strongly a novel property is generalized from one to the other.

These findings regarding the importance of causal relations in generalizing properties between exemplars replicates those of Lassaline (1996, Experiment 2), who also found that the presence of a cause of the to-be-projected property in a target category led to stronger induction ratings and also a reduced effect of whether the source and the target categories were otherwise similar (see also Wu & Gentner, 1998). The present experiment demonstrates, in addition, that the explicit *absence* of that cause leads to much weaker inductions (and again, an almost complete elimination of any effect of the number of features shared by the source and the target).

## EXPERIMENT 4

Experiments 1–3 established that category-based generalizations are often dominated by causal explanations. The final two experiments were performed to determine which sorts of causal explanations maximize the strength of generalizations. One important component of an explanation, of course, is the feature that causes the novel property. For example, for the novel property “has a venom with a stinging bite,” what are the characteristics of its purported cause (“high amounts of iron sulfate in the blood”) that strengthen the conclusion that all Kehoe Ants have venom with a stinging bite?

One possibility to consider is the role that the cause feature plays in the overall conceptual structure of the category. For example, Sloman, Love, and Ahn (1998) have suggested that category features vary in their conceptual *centrality*, with features that are *depended on* by many other features occupying a more central position than features with few dependencies. On this account, the DNA of most biological organisms would be considered central, because DNA is thought to be the cause of many of the other features of the organism. More causally peripheral features (e.g., color) would be less central. Centrality appears to be highly relevant to property generalization, because novel properties may be more strongly projected when they are attributed to more central features. For example, in the present experiments, it is easy to imagine that the projection of the novel property “has a venom with a stinging bite” would have been even stronger had it been described as being directly caused by the Kehoe Ant’s DNA, instead of one of the characteristic but less central features (iron sulfate in the blood).

In Experiment 4, the participants were presented with nonblank novel properties, and they judged how likely it was that all the category members had that property (as in Experiments 1 and 2). They learned about two categories. For one, feature  $F_1$  was made more conceptually central by describing it as the cause of the other three features. For example, for Kehoe Ants, feature  $F_1$  (blood high in iron sulfate) was described as the cause of  $F_2$  (hyperactive immune system),  $F_3$  (thick blood), and  $F_4$  (faster nest building).

For the other (control) category, no causal relationships linking the four characteristic features were provided (just as in the first three experiments). The question was whether  $F_1$  would support stronger generalizations when it was described as the cause of the other features—that is, when it was more conceptually central.

### Method

**Participants.** Twenty-four New York University undergraduates received course credit or pay for participating in this experiment.

**Materials.** The materials from Experiments 1–3 were augmented with the three causal links between  $F_1$  and the features  $F_2$ ,  $F_3$ , and  $F_4$ . Each link included a description of the causal mechanism involved. For example, for Kehoe Ants, the  $F_1 \rightarrow F_3$  link was the following: “Blood high in iron sulfate causes thick blood. Iron sulfate provides the extra iron that the ant uses to produce extra red blood cells. The extra red blood cells thicken the blood.” A complete list of the causal links for all the categories is available from the author.

**Design.** One within-subjects factor was whether  $F_1$ ,  $F_2$ ,  $F_3$ , or  $F_4$  was the purported cause of the novel property. The other was whether those features were themselves causally linked. The between-subjects factors were whether the participants learned the two biological kinds, the two nonliving natural kinds, or the two artifacts, which of those two categories had the interfeature causal relationships, and whether that category was presented first or second, all of which were varied.

**Procedure.** The participants learned about two categories, using the same general procedure as that in the first three experiments. Each category had four binary features, each of which was described as occurring in 75% of the category members. In addition, for one of the categories, the participants were required to learn about the three causal relations  $F_1 \rightarrow F_2$ ,  $F_1 \rightarrow F_3$ , and  $F_1 \rightarrow F_4$ . This causal information was also presented in the form of a network diagram in which causal links were depicted as arrows between features. For the category with the causal information, the multiple-choice test consisted of 22 questions: the 8 questions used in the control condition, plus 14 questions about the causal relations and their underlying mechanisms.

During the induction task that followed the learning of each category, the participants were given four randomly ordered trials in which they were presented with a single category member with a nonblank novel property and were asked what proportion of all the category members possessed that property. On each trial, the category member displayed all four characteristic features (i.e.,  $F_1$ ,  $F_2$ ,  $F_3$ , and  $F_4$ ), but which of those features the novel property was attributed to changed from trial to trial. A categorization task was also presented for each category, and the order of the induction and categorization tasks was randomized for each participant.

### Results and Discussion

No effects of the between-subjects factors were found, and thus the induction results are presented in Figure 4 as a function of which of the four characteristic features the novel property was caused by and whether the characteristic features were related by causal relationships. The results are clear-cut: The causal links among characteristic features had no effect on the generalization of novel properties. In particular, feature  $F_1$  did not support stronger generalizations when it was described as the cause of the other fea-

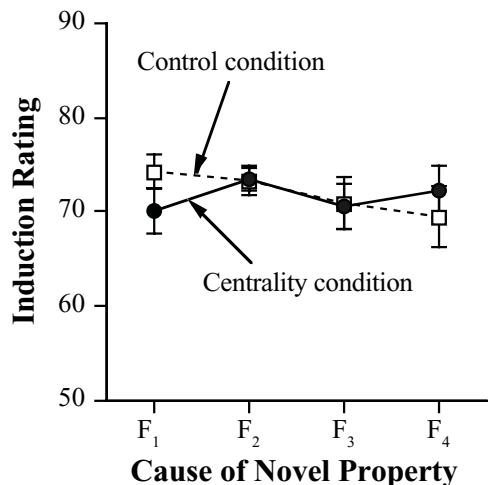


Figure 4. Average induction ratings from Experiment 4. Error bars represent standard errors of the means.

tures, despite the fact that it was more conceptually central. In a  $2 \times 4$  ANOVA of these data, neither the main effect nor the interaction reached significance (all  $ps > .10$ ).

It is illuminating to compare the results from Experiment 4 with those of Rehder and Hastie (2004), who found that the generalization of a *blank* novel property displayed by an exemplar with all the characteristic features was strengthened when those features were causally interrelated. They attributed this result to the exemplar's being perceived as a good category member because it manifested not just all the characteristic features, but also the category's causal mechanisms (the cause and effect of each causal relationship were both present). However, just as in Experiment 1, the importance of category membership to property generalization appears to be eliminated when the novel property can be causally attributed to a specific category feature. Again, the presence of a causal explanation seems to draw attention away from the exemplar's other features, making it irrelevant whether those other features are characteristic of the category or whether they instantiate the category's other causal laws.

## EXPERIMENT 5

Experiment 4 showed that novel properties are not necessarily projected more strongly when they are caused by more conceptually central features. Yet the intuition remains that a novel property attributed to a biological organism's DNA will be more strongly generalized than one attributed to a more superficial feature. But this may be the case not because of conceptual centrality per se, but rather because people know that all members of the species have that DNA and, thus, are likely to have the novel property that it causes as well. In comparison, a novel property that is caused by a feature that appears in many, but not all, members of the category will itself appear in many, but not all, category members. On this account, the more central feature ( $F_1$ ) failed to support stronger generalizations in Experiment 4, because it was described as having the same base rate (75%) as the rest of the features.

This possibility regarding the importance of the cause feature's base rate can be made precise by expressing it in terms of an explicit theory of causal relations. For example, I have proposed a formal representation of causal knowledge in which category features are related by *probabilistic causal mechanisms* (Rehder, 2003a, 2003b). According to this formalism, the presence of the cause feature enables the mechanism that brings about (with some probability) the effect feature. If the effect feature is taken to be some to-be-generalized novel property  $N$  and  $C$  is its purported cause, the probability of  $N$  among category members is

$$P(N) = mP(C), \quad (1)$$

where  $m$  is the *power* of the causal mechanism—that is, the probability that the mechanism linking  $C$  and  $N$  will operate (and thus bring about  $N$ ) when  $C$  is present.<sup>3</sup> Equation 1 makes explicit how the prevalence of a novel

property,  $P(N)$ , increases as the prevalence of its cause increases. It also indicates that  $P(N)$  depends on the power of the causal mechanism. For example, if the mechanism through which a novel property is caused by DNA is deterministic (i.e.,  $m = 1$ ), then  $P(N) = P(\text{DNA}) = 1$ . But if that mechanism operates probabilistically instead ( $m < 1$ ; in the way that smoking only sometimes causes lung cancer),  $N$  will be less prevalent than its cause.

Simple preliminary experiments conducted in my lab, using the same experimental categories as those in Experiments 1–4, have confirmed the importance to category-based generalization of both the base rate of the cause feature and the reliability of the causal mechanism. First, in one experiment, two of the four characteristic features were described as possessing a base rate of 90%, whereas the other two were described as possessing a base rate of 60%. (There were no causal relationships linking the four features.) The participants then performed the same induction task as that used in Experiment 4. In fact, induction ratings were much larger when the nonblank was described as being caused by a feature with a 90% base rate (89.9), as compared with one with a 60% base rate (54.0). A second experiment manipulated the power of the causal mechanism that generated the novel property, while holding the base rate of the characteristic features constant at 75%. On some trials, the causal mechanism that linked the novel property was described as highly reliable (e.g., the mechanism that linked the novel property “has a venom with a stinging bite” to the characteristic feature “blood high in iron sulfate” would include the additional sentence “in general, high amounts of iron sulfate always results in a stinging bite”). On other trials, the mechanism was described as being less reliable (e.g., “in general, high amounts of iron sulfate often, but not always, results in a stinging bite”). As was expected, more reliable mechanisms led to stronger inductive inferences (ratings of 86.0) than did less reliable ones (68.1). Together, these experiments support the claim that category-based generalizations follow some of the basic principles of causal inference (as specified in Equation 1).

The purpose of Experiment 5 was to establish that a novel property is, indeed, more strongly generalized to the extent that its cause is more conceptually central, but that this occurs because more central causes are likely to be viewed as more prevalent in category members. For example, suppose that an organism's DNA is thought to produce (with some probability) a particular enzyme and that the enzyme, in turn, is thought to produce (with some probability) the organism's characteristic color. In addition to believing that all members of the species have that DNA, one would then also be justified in believing that most (but not all) have the enzyme and that, of those that have the enzyme, most (but not all) have the typical color. That is,  $P(\text{DNA}) > P(\text{enzyme}) > P(\text{typical color})$ . On the basis of the preliminary experiments described above, one would then predict that a novel property would be most strongly generalized when attributed to DNA, would be less strongly generalized when attributed to the enzyme, and would be most weakly generalized when attributed to color.

In Experiment 5, one feature ( $F_1$ ) played the role of a defining feature (analogous to DNA) by describing it as occurring in 100% of the category members (and in members of no other category).  $F_1$  was then also described as causing a chain of other features:  $F_1 \rightarrow F_2 \rightarrow F_3 \rightarrow F_4$ . So that this causal network would influence the base rate of features  $F_2$ ,  $F_3$ , and  $F_4$ , no explicit base rate information for those features was provided. The prediction was that a nonblank property would be more strongly generalized when it was attributed to a more central feature (e.g.,  $F_1$  or  $F_2$ ) than when it was attributed to a less central one (e.g.,  $F_3$  or  $F_4$ ).

## Method

**Participants.** Twenty-four New York University undergraduates received course credit or pay for participating in this experiment.

**Materials.** The materials were the same as those in Experiment 4, except that, in order to construct a causal chain, the causal links  $F_1 \rightarrow F_3$  and  $F_1 \rightarrow F_4$  were replaced with  $F_2 \rightarrow F_3$  and  $F_3 \rightarrow F_4$ . (Again, these materials are available on request from the author).

**Design.** As in Experiment 4, the two within-subjects factors were whether  $F_1$ ,  $F_2$ ,  $F_3$ , or  $F_4$  was the purported cause of the novel property and whether those features were themselves causally linked (this time into a causal chain). The same three between-subjects counterbalancing factors as those used in Experiment 4 were present in this experiment.

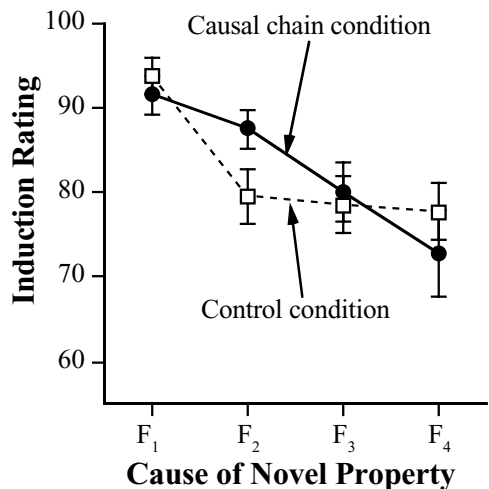
**Procedure.** Once again, the participants learned two categories. For both categories, the participants were told that the base rate of feature  $F_1$  was 100% and that no member of any other category had feature  $F_1$ . No base rate information was provided for  $F_2$ ,  $F_3$ , or  $F_4$ . For one of the categories, the participants learned about the three causal relations  $F_1 \rightarrow F_2$ ,  $F_2 \rightarrow F_3$ , and  $F_3 \rightarrow F_4$ , following the same procedure as that used in Experiment 4. After learning each category, the participants performed the same induction task as that used in Experiment 4. A categorization task either preceded or followed the induction task.

## Results and Discussion

There was again no effect of the between-subjects factors, and thus the induction results are presented in Figure 5 as a function of which of the four characteristic features the novel property was caused by and whether the characteristic features were related in a chain structure. The first thing to note is that, in both conditions, novel properties were strongly generalized (average rating of 91.5) when caused by feature  $F_1$ , which was described as having a 100% base rate. This result confirms the intuition that a property will be strongly projected when thought to be caused by a feature that is itself present in all category members (e.g., DNA in the case of a biological organism).

The second important result concerns the generalization of properties attributed to features  $F_2$ ,  $F_3$ , and  $F_4$  as a function of whether or not those features were linked in a causal chain. As was predicted, in the chain condition, the strength of inductions decreased as a function of the “distance” of the cause feature from  $F_1$ : Ratings for  $F_2$  were higher than those for  $F_3$ , which, in turn, were higher than those for  $F_4$ . In contrast, the ratings for  $F_2$ ,  $F_3$ , and  $F_4$  did not differ from one another in the control condition.

These results were supported by statistical analysis. A  $2 \times 4$  ANOVA revealed both a main effect of which feature the nonblank was caused by [ $F(3,69) = 14.44$ ,  $MS_e =$



**Figure 5.** Average induction ratings from Experiment 5. Error bars represent standard errors of the means.

184,  $p < .0001$ ) and a reliable interaction between feature and causal condition [chain vs. control;  $F(2,46) = 3.65$ ,  $MS_e = 113$ ,  $p < .05$ ], reflecting the different pattern of generalization in the two conditions. In each condition, separate analyses were conducted on the contrasts between each adjacent feature pair. In the control condition, the rating associated with feature  $F_1$  was significantly greater than that for  $F_2$  ( $p < .001$ ), but there were no significant differences between  $F_2$ ,  $F_3$ , and  $F_4$  ( $ps > .20$ ). In comparison, in the chain condition, those contrasts were the following:  $F_1$  versus  $F_2$ ,  $p < .10$ ;  $F_2$  versus  $F_3$ ,  $p < .05$ ; and  $F_3$  versus  $F_4$ ,  $p < .05$ .

These results complement those of Rehder (2003b, Experiment 3), who found that when features were arranged in a causal chain (and the root feature defined category membership, as in this experiment), features that were more central (i.e., those that occurred earlier in the chain; e.g.,  $F_2$ ) were more diagnostic of category membership than were those that occurred later (e.g.,  $F_4$ ). In that work, I suggested that this effect arose because more central features were more likely to be causally produced, or *generated*, by the underlying defining feature. The present experiment shows that this effect carries over to category-based generalizations: Novel properties are more likely to be generalized to the category as a whole when they are caused by features that are themselves more likely to be generated by the category.

## GENERAL DISCUSSION

### When Similarity and Causality Compete

The first part of this article was concerned with the interaction between causal explanations and similarity on category-based generalizations. The results of Experiments 1–3 supported the claim that when a causal explanation for a novel property is available, it often supplants similarity as the basis for the generalization of that prop-

erty. First, in Experiment 1, all the participants exhibited an effect of typicality for blank properties, but half of those participants showed no sensitivity to typicality when the novel property was accompanied with a causal explanation (nonblanks). Second, in Experiment 2, half the participants exhibited sensitivity to diversity for blank properties, but that sensitivity was completely eliminated for nonblanks for all the participants. Finally, in Experiment 3, virtually all the participants exhibited sensitivity to similarity for blanks, but this effect was (almost) completely eliminated for nonblanks. Apparently, when people note the presence of a causal explanation for a novel property, it often draws attention away from the exemplars' other features, making their similarity (or typicality or diversity) largely irrelevant to the inductive judgment.

Experiments 1–3 also provide some insight into the relative strength of the three similarity-based heuristics. The diversity heuristic appeared to be the weakest, since only about half of the participants in Experiment 2 exhibited a diversity effect, even with blank properties (and that effect of diversity was eliminated entirely when a causal explanation was provided). This finding is consistent with the literature demonstrating the diversity heuristic's lack of robustness (López et al., 1997; López, Gelman, Gutheil, & Smith, 1992; see Heit, 2000, for a review). A similarity effect with blank properties was exhibited by virtually all the participants in Experiment 3, but like diversity, that effect was almost completely eliminated with nonblanks. Finally, all the participants in Experiment 1 exhibited a typicality effect with blank properties, but that effect was eliminated for only half of those participants for nonblanks. That is, typicality proved to be the most tenacious of the three similarity-based heuristics in its ability to compete with causal explanations. Nevertheless, even the effect of typicality was completely eliminated for those Experiment 1 participants who chose to generalize on the basis of a causal explanation. In other words, basing generalizations on causal versus typicality information appears largely to be a matter of either/or, but not both.

One caveat regarding these findings concerns this study's use of experimentally created categories, rather than natural ones. Artificial categories were used in order to provide experimental control over the causal explanation of novel properties, but of course, this approach comes at the cost of some degree of ecological validity. For example, because the categories were unfamiliar, it was necessary to instantiate the typicality, diversity, and similarity manipulations in terms of explicit lists of features. For this reason, it is important to note that when the properties were blank, Experiments 1–3 succeeded in replicating the standard similarity-based effects, suggesting that, despite the use of artificial categories and feature lists, the same processes as those involved with natural categories were involved here as well. Indeed, given that feature lists were likely to make exemplars' similarity (or typicality or diversity) particularly salient, it is especially notable that the similarity-based effects were almost entirely eliminated when a causal explanation for the novel property was provided.

Of course, any claim about the effect of explanations must be tempered by the fact that those explanations were provided as part of the experimental session, rather than coming from the participants' own world knowledge. On the one hand, because the participants may have found the experimental explanations to be less plausible than those that come from their own background knowledge, Experiments 1–3 may, in fact, be *underestimating* the strength of explanations on category-based generalizations. On the other hand, in those experiments, the explanations were presented to the participants as part of the nonblank induction trials, a fact that was likely to have made them especially salient and that bypassed the need for the reasoners to retrieve the explanations from their own semantic memory. In contrast, real-world acts of generalization require that the context be sufficient for retrieval of an explanation from memory.

Nevertheless, the research reviewed earlier indicates that it is not uncommon for reasoners with substantial domain knowledge to use that knowledge in support of category-based generalization, including the Itzá Maya (experts in ecological knowledge; Bailenson et al., 2002; López et al., 1997), experts in trees (Proffitt et al., 2000), and experts in fish (Shafto & Coley, 2003). Medin, Coley, Storms, and Hayes (2003) demonstrated that causal reasoning is even common among nonexperts (i.e., college students). For example, they found that more diverse premises, such as “sparrows and seeds have enzyme X,” resulted in *weaker* generalization of X to all living things, as compared with less diverse premises, such as “sparrows and dogs have enzyme X.” Medin et al. attributed this result to the students' knowledge of a possible causal mechanism by which sparrows might acquire enzyme X from seeds (ingestion), which led them to conclude that all living things were unlikely to have enzyme X (because not all living things eat seeds). In other words, causal-based generalizations are not limited to the rarefied world of experts but, rather, may well be a common way by which the average person turns an everyday experience into general category knowledge.

Note that I have presented the similarity-based heuristics and causal reasoning as two distinct classes of mental processes that underlie the generalization of novel properties. However, it may be that the similarity-based heuristics are themselves based on a form of causal reasoning. Numerous theorists have suggested that categories possess an underlying causal structure that relates existing category features (Bloom, 2000; Keil, 1989; Medin & Ortony, 1989; Rehder & Burnett, 2005) and influences category-based generalizations (Gelman, 2003; Gelman & Markman, 1986; Rehder & Hastie, 2004). For example, Rehder and Burnett found that undergraduates were more likely to infer the presence of a feature in a given category member when it already possessed many features characteristic of the category; that is, they exhibited a typicality effect in feature inference (as opposed to category-based generalization). We attributed this finding to the fact that a more typical category member more perfectly instantiates the category's causal laws, which, in turn, makes it more

likely to possess additional category features. Similarly, people may exhibit a typicality effect in category-based generalization because a novel property will be attributed to a category's "typical" causal mechanisms when it is displayed by a typical category member. But remember that even if similarity-based effects themselves reflect a kind of causal reasoning, the present results document how those effects are largely eliminated when reasoners have knowledge of the specific causal mechanism responsible for a novel property.

### Causal Reasoning and Property Generalization

The second part of this article was concerned with establishing which properties of causal explanations determine the strength of generalizations. One finding was that causal-based generalizations obey the same basic principles one would expect of any account of causal reasoning. For example, one generally expects that an effect will become more prevalent in some population to the extent that its cause is also prevalent in that population. In addition, the effect's prevalence should vary as a function of the reliability of the mechanistic processes by which it is produced. In fact, experiments were described which showed that the strength of category-based generalizations was strongly influenced by the base rate of the feature described as the cause of the novel property and also by the reliability of that cause. These results indicate that the effect of providing causal explanations in a category-based generalization task is not merely to shift reasoners' attention toward the constituent features of those explanations, but also to evoke the basic principles of causal reasoning.

The experiments reported in this article tested the effect of causal explanations with a very specific form, one in which the novel property was caused by a single existing category feature. But, of course, category-based generalizations can involve explanations with other forms. For example, a novel property may be thought to be the *cause of* (rather than being *caused by*) one or more category features. In fact, it is straightforward to apply the causal reasoning perspective to this case: The prevalence of a novel property  $N$  that causes some feature  $E$  will increase as the prevalence of  $E$  increases and will decrease as it decreases (see Equation 1). In fact, an experiment reported in Hadjichristidis, Sloman, Stevenson, and Over (2004) is suggestive of the fact that property generalization can be based on reasoning from effects to causes, as well as from causes to effects. Participants were told of two novel properties that were "dependent on" by either many or few other category features (i.e., they had either many or few effect features). Hadjichristidis et al. found that the participants projected the property with many dependents to a new category more often than the one with few dependents. Interestingly, the size of this effect was larger when the base and the target were similar, a result which Hadjichristidis et al. attributed to two principles: (1) More central properties (those with more dependents) are more projectable, and (2) the effect of centrality depends on the extent to which the base and the target categories share the central property's *dependency structure* (i.e., its effects).

In the absence of explicit knowledge about the property's actual dependency structure, it can be estimated from the similarity between the base and the target (e.g., lower similarity implies fewer shared effect features).

The causal-reasoning perspective advocated here provides a similar explanation of these findings, although in somewhat different terms: The prevalence of a novel property among category members,  $P(N)$ , will be high to the extent that its presence can be inferred from  $N$ 's effect(s). An advantage of the causal-reasoning account, however, is that it provides a single explanation of both the experimental results presented here and those in Hadjichristidis et al. (2004). Specifically, a novel property can be strongly projected to a category either when its causes are common and numerous in number or when its effects are common and numerous in number (or both). In other words, the concept of *centrality* and *dependency structure* can be subsumed into a broader framework in which category-based generalizations can be seen as a species of causal inference more generally.

### Centrality and Property Generalization

Another purpose of the present study was to test whether generalizations were influenced by the cause feature's conceptual centrality. It was predicted on intuitive grounds that any feature that is the cause of many other features (e.g., DNA in the case of a biological organism) will provide a strong basis for the projection of a novel property to all category members. In fact, however, Experiment 4 revealed no influence of centrality: Generalizations were as strong when the novel property was caused by a feature with many other effects as when it was caused by a feature with none. But in that experiment, both central and less central features were described as possessing the same base rate. An alternative explanation for why DNA serves as a strong basis for generalization in biological organisms is that its presence in all category members means that the novel property it generates is also likely to be in all (or at least many) category members. In fact, Experiment 5 showed that a central feature that was described as occurring in 100% of the category members supported the strongest generalizations in any of the five experiments (91.5 on a scale of 100).

Nevertheless, it would be a mistake to conclude that conceptual centrality is always irrelevant to how people generalize new properties. In Experiment 5, features were arranged in a causal chain in which no base rate information was provided other than for the defining root feature,  $F_1$ . As was predicted, features that were more central, or "deeper," in the causal chain supported stronger generalizations. But remember, this effect apparently arose not because of centrality or depth per se but, rather, because deeper features are more likely to be generated (i.e., to have a greater base rate; Rehder, 2003b) and thus, in turn, are more likely to generate the novel property.

This research has focused on how within-category causal relations influence category-based generalizations. However, it is important to note that generalizations will also be influenced by the causal relations that hold

between categories. An example presented above demonstrated the importance of ingestion relations among categories (a property displayed by both sparrows and seeds will be attributed to the fact that sparrows eat seeds; Medin et al., 2003; see also Kemp & Tenenbaum, 2003). Another common finding is that the generalization of a disease depends on how it might spread among species (ingestion, physical proximity, and so on; López et al., 1997; Proffitt et al., 2000; Shafto & Coley, 2003). Properties might be shared between substances on the basis of their common origin (e.g., between sand and glass; Medin et al., 2003), and so on.

Nevertheless, there are many cases in which reasoners will have knowledge of the specific causal mechanisms that lead a category member to have a novel property, and this study has established three general findings about how such properties are generalized. First, when the participants used causal explanations, standard effects of similarity were almost completely eliminated. Second, novel properties were generalized more strongly when their cause was prevalent among category members, suggesting that such generalizations can be seen as instances of general causal reasoning. Third, more central features supported stronger generalization but did so because central features are just those that tend to be most prevalent.

## REFERENCES

- BAILENSON, J. N., SHUM, M. S., ATRAN, S., MEDIN, D. L., & COLEY, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition*, **84**, 1-53.
- BLOOM, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- GELMAN, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, **20**, 65-95.
- GELMAN, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.
- GELMAN, S. A., & MARKMAN, E. M. (1986). Categories and induction in young children. *Cognition*, **23**, 183-208.
- HADJICHRISTIDIS, C., SLOMAN, S. A., STEVENSON, R., & OVER, D. (2004). Feature centrality and property induction. *Cognitive Science*, **28**, 45-74.
- HEIT, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, **7**, 569-592.
- HEIT, E., & RUBINSTEIN, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 411-422.
- KEIL, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- KEMP, C., & TENENBAUM, J. B. (2003). Theory-based induction. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 203-208). Boston: Cognitive Science Society.
- LASSALINE, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 754-770.
- LÓPEZ, A., ATRAN, S., COLEY, J. D., MEDIN, D. L., & SMITH, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology*, **32**, 251-295.
- LÓPEZ, A., GELMAN, S. A., GUTHEIL, G., & SMITH, E. E. (1992). The development of category-based induction. *Child Development*, **63**, 1070-1090.
- MEDIN, D. L., COLEY, J. D., STORMS, G., & HAYES, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, **10**, 517-532.
- MEDIN, D. L., & ORTONY, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-196). Cambridge: Cambridge University Press.
- MURPHY, G. L., & ROSS, B. H. (1999). Induction with cross-classified categories. *Memory & Cognition*, **27**, 1024-1041.
- OSHERSON, D. N., SMITH, E. E., WILKIE, O., LÓPEZ, A., & SHAFIR, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.
- PROFFITT, J. B., COLEY, J. D., & MEDIN, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 811-828.
- REHDER, B. (2003a). Categorization as causal reasoning. *Cognitive Science*, **27**, 709-748.
- REHDER, B. (2003b). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1141-1159.
- REHDER, B., & BURNETT, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, **50**, 264-314.
- REHDER, B., & HASTIE, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, **130**, 323-360.
- REHDER, B., & HASTIE, R. (2004). Category coherence and category-based property induction. *Cognition*, **91**, 113-153.
- RIPS, L. J. (2001). Necessity and natural categories. *Psychological Bulletin*, **127**, 827-852.
- ROSS, B. H., & MURPHY, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, **38**, 495-553.
- SHAFTO, P., & COLEY, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 641-649.
- SLOMAN, S. A. (1993). Feature-based induction. *Cognitive Psychology*, **25**, 231-280.
- SLOMAN, S. A. (1994). When explanations compete: The role of explanatory coherence on judgments of likelihood. *Cognition*, **52**, 1-21.
- SLOMAN, S. A. (1997). Explanatory coherence and the induction of properties. *Thinking & Reasoning*, **3**, 81-110.
- SLOMAN, S. A., LOVE, B. C., & AHN, W.-K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, **22**, 189-228.
- SMITH, E. E., SHAFIR, E., & OSHERSON, D. (1993). Similarity, plausibility, and judgments of probability. *Cognition*, **49**, 67-96.
- WU, M. L., & GENTNER, D. (1998). Structure in category-based induction. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 1154-1158). Mahwah, NJ: Erlbaum.

## NOTES

1. The hedge "all else being equal" refers to the fact that the effect of diversity holds when typicality is controlled for. In the example, the two sets of base categories (i.e., {sparrows, hawks, chickens} and {sparrows, robins, bluejays}) are assumed to have the same average typicality.

2. According to the SCM, such generalizations are a function of (1) the coverage the premises provide for the lowest level superordinate category that includes both the base and the target categories and (2) the similarity between the base and the target. The SCM is thus able to account for the fact that the base-target similarity has an effect on inductions above and beyond any effect of coverage per se. For example, people will be more confident that blackbirds have sesamoid bones given the fact that crows do, as compared with sparrows, because (according to the SCM) blackbirds are more similar to crows than to sparrows. This is the case even if crows do not cover the superordinate category (e.g., *bird*) as well as do sparrows.

3. Rehder (2003b) specified that the probability of a feature  $F$  is a function of whether it is brought about by its direct cause  $C$  or by some unspecified background cause—that is,

$$P(F) = mP(C) + b - mbP(C),$$

where  $b$  is the probability that the background cause operates. For simplicity, in Equation 1, I assume that  $b = 0$ .