Cognitive Science 33 (2009) 301–344

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01015.x

Causal-Based Property Generalization

Bob Rehder

Department of Psychology, New York University

Received 18 July 2007; received in revised form 19 November 2008; accepted 21 November 2008

Abstract

A central question in cognitive research concerns how new properties are generalized to categories. This article introduces a model of how generalizations involve a process of causal inference in which people estimate the likely presence of the new property in individual category exemplars and then the prevalence of the property among all category members. Evidence in favor of this *causal-based generalization* (CBG) view included effects of an existing feature's base rate (Experiment 1), the direction of the causal relations (Experiments 2 and 4), the number of those relations (Experiment 3), and the distribution of features among category members (Experiments 4 and 5). The results provided no support for an alternative view that generalizations are promoted by the centrality of the to-be-generalized feature. However, there was evidence that a minority of participants based their judgments on simpler associative reasoning processes.

Keywords: Causal-based induction; Generalization; Causal reasoning

1. Introduction

Induction—reasoning to uncertain conclusions—appears in many forms. In some cases one makes an uncertain inference about a specific object or event. Given a particular dog, one asks whether it is safe to pet; given a particular berry, one asks if it is safe to eat. But in other cases one makes *inductive generalizations* that are intended to characterize an entire class of situations or objects. On the basis of a finite number of medical cases one might induce a general causal law, such as that unsafe sex can cause AIDS or that mosquitoes can cause malaria. Or from a few examples one might make a generalization about a property being displayed by many or most members of a particular category, such as that koalas eat bamboo, Apple laptops have fire-prone batteries, or Madagascar fire ants have poisonous

Correspondence should be sent to Bob Rehder, Department of Psychology, 6 Washington Place, New York, NY 10003. E-mail: bob.rehder@nyu.edu

venom. This article is concerned with *category-based generalizations* such as these in which new properties are projected to an entire class of objects.

Research over the last several decades has attempted to identify those factors that promote strong category-based generalizations, and a number of important empirical regularities have been documented. People's generalizations often exhibit *typicality effects* in which an example of a novel feature will result in it being more strongly generalized to a category if the example is more typical or representative of the category. For example, people will be more confident that all fruits have a novel property given an example of an apple that has it as compared to a fig, because apples are more typical fruits than figs. People also sometimes exhibit *diversity effects* in which a more diverse set of examples leads to stronger generalizations to a category. For example, people are more sure that all birds have a novel property given that sparrows, hawks, and chickens have it, as compared to sparrows, robins, and blue jays. This is the case because the former set of categories is more different from one another than the latter. Typicality, diversity, and other related effects have led to the creation of well-known models of category-based generalization such as the *similarity-coverage model* (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990) and Sloman's (1993) *feature-based induction model*. Although there are differences between these models, their predictions are each derived from the similarity relations that obtain between the examples and the target category.

But while such models account for a wide range of laboratory phenomena, it may be that their applicability to real-world reasoning situations is limited, because people often know more about the situation than just similarity relations between examples and categories. Whereas the effects of typicality, diversity, and other phenomena have been established experimentally with features about which people supposedly have no prior knowledge (i.e., ''blank properties''), it is often the case that people have at least some inkling regarding how the to-be-generalized feature could be related to the category in question. The purpose of this article is to introduce a model of the reasoning processes invoked in the presence of the causal knowledge that relates a to-be-generalized property to the category. This view, referred to here as *causal-based generalization* (hereafter CBG), focuses on the specific causal explanations that might lead one to believe that most or all members of a particular category display a particular property. Although it may be vague or incomplete, one often has at least some general idea of the type of causal mechanisms that produce or generate a particular property, and this knowledge can be used to estimate the prevalence of that property in some population (i.e., category of object).

Prior research has been highly suggestive of the influence on generalizations of causal knowledge linking a new feature to its target category (for reviews, see Heit, 2000; Rehder, 2007; Rips, 2001). For example, Heit and Rubinstein (1994) found that a behavioral feature (e.g., travels in a zigzag path) was generalized more strongly from tunas to whales as compared to from bears to whales, whereas this preference was reversed for biological features (e.g., a liver with two chambers that acts as one). Why should this reversal arise despite the fact that the categories involved (bears, tunas, and whales) were unchanged, and thus so too were the similarity relations? One explanation is that participants thought that bears and whales share biological properties (such as two-chambered livers), because such properties are likely to arise from causal mechanisms associated with their common biological category (mammals). Tunas and whales, on the other hand, are more likely to share a survival behavior (traveling in a zigzag path) because they are both prey animals living in a common ecology (also see Gelman & Markman, 1986; Sloman, 1994, 1997; Springer & Keil, 1989).

Smith, Shafir, and Osherson (1993) similarly found that subjects engaged in a form of causal reasoning when generalizing hypothetical properties of familiar categories. For example, undergraduates were more likely to generalize the property ''can bite through barbed wire'' to German shepherds from poodles than from Dobermans. This result obtained despite the fact that German shepherds are more similar to Dobermans than to poodles. It seems that the participants were reasoning about the causal preconditions for the capacity to bite through barbed wire, and they judged that a German shepherd certainly could do so if a small dog like a poodle could, but not necessarily if the base category was another powerful dog (also see Osherson, Smith, Myers, & Stob, 1994).

Moreover, recent research suggests that causal explanations not only influence generalizations, but they may also largely supplant, or eliminate, similarity-based effects. For example, Medin, Coley, Storms, and Hayes (2003) found that a more diverse set of premise categories consisting of a mammal and a bird (e.g., cats and sparrows) supported *weaker* generalizations (e.g., of a blank property ''enzyme X'') to lizards as compared to a less diverse set of two mammals (cats and rhinos). Apparently, the fact that cats eat sparrows suggested to subjects a possible causal mechanism by which enzyme X might be shared by them but not by lizards, which are in a different food chain (also see Bailenson, Shum, Atran, Medin, & Coley, 2002; Lopez, Atran, Coley, Medin, & Smith, 1997; Proffitt, Coley, & Medin, 2000; Shafto & Coley, 2003). In a series of experiments Rehder (2006) orthogonally manipulated similarity-based factors such as typicality and diversity and the presence of causal relations linking a to-be-generalized feature to the category, and he found that similarity-based effects were almost entirely eliminated when a generalization could be based on causal relations instead (Lassaline, 1996; Wu, & Gentner, 1998).

Clearly, then, causal relations have an important and perhaps decisive role in determining how and to what degree properties are ascribed to categories. This article presents a specific model of the reasoning processes that underlie causal-based generalizations along with a series of experimental tests of that account. To provide control over the causal knowledge that serves as the basis of a generalization, each experiment instructed undergraduates about a new category and then presented a novel property that was described as causally related to the features of that category. Participants were then asked to judge how prevalent that property was in members of the category. To test CBG, in most experiments the direction of the causal link between the existing category feature and the novel one was manipulated as an experimental factor. For example, Table 1 presents one of the experimental categories used in this study, Romanian Rogos (a type of automobile). Rogos were described as possessing four characteristic features, including butane-laden fuel, loose fuel filter gaskets, and so on. After learning about Rogos, participants were presented with a series of trials in which they were told about one of the novel features presented in Table 1. For instance, some participants were presented with the property of a zinc-lined gas tank and told that

Table 1
Features and causal relationships for Romanian Rogos, an artificial category

Characteristic Feature	Novel Feature	Novel Feature as Effect	Novel Feature as Cause
Butane-laden fuel	Zinc-lined gas tank	Butane-laden fuel causes a zinc-lined gas tank. The butane interacts with the chromium in the metal of the gas tank, which results in a thin layer of zinc on the inside of the tank.	A zinc-lined gas tank causes the fuel to be butane laden. The zinc prevents corrosion of the tank, but it interacts chemically with gasoline to produce butane.
Loose fuel filter gasket	Vibrations during braking	A loose fuel filter gasket causes vibrations during braking. The fuel which leaks through the fuel filter gasket falls on one of the brake pads, causing abrasion which results in the car vibrating while braking.	Vibration during braking causes a loose fuel filter. The rattling caused by the vibrations eventually leads to the fuel filter gasket becoming loose.
Hot engine temperature	Thin engine oil	Hot engine temperature causes thin engine oil. The oil loses viscosity after it exceeds a certain temperature.	Thin engine oil causes hot engine temperature. Thin oil does not provide sufficient lubrication for the engine's moving parts, and the engine temperature goes up as a result.
High amounts of carbon monoxide in the exhaust	Inefficient turbocharger	High amounts of carbon monoxide in the exhaust causes an inefficient turbocharger. As the exhaust leaves the engine it passes through the turbocharger. The lower density of carbon monoxide in the exhaust means that the turbocharger is not sufficiently pressurized.	An inefficient turbocharger causes high amounts of carbon monoxide in the exhaust. An inefficient turbocharger fails to inject enough oxygen into the engine, and so excess carbon does not undergo combustion.

zinc-lined gas tanks are caused by one of Rogos existing features, butane-laden fuel (''Butane-laden fuel causes a zinc-lined gas tank. The butane interacts with the chromium in the metal of the gas tank, which results in a thin layer of zinc on the inside of the tank.''). Other participants were told that zinc-lined gas tanks cause butane-laden fuel (''A zinc-lined gas tank causes the fuel to be butane-laden. The zinc prevents corrosion of the tank, but interacts chemically with gasoline to produce butane.''). All participants then rated how prevalent zinc-lined gas tanks are among Rogos. Table 1 presents the four novel properties that were generalized to Rogos on the basis of their (forward or backward) causal relation with one of Rogos' four existing features.

Experiments 1–4 each used a two-factor design in which one other factor besides the direction of causality was manipulated. In Experiment 1, the existing feature to which the novel feature was causally related was itself either very common or only moderately common among Rogos. As demonstrated below, CBG predicts that generalizations should be stronger to the extent that the feature that serves as its causal basis is prevalent, and this should be the case regardless of causal direction. Experiment 2 manipulated the strength of the causal link relating existing and novel features. CGB predicts an interaction between causal direction and strength, such that generalizations should be stronger for stronger causal links when the novel feature is an effect, but this effect should be reversed when the novel feature is a cause. Experiment 3 varied the number of links relating the novel property and existing features, testing CBG's prediction that more links lead to stronger generalization regardless of direction. Finally, Experiments 4 and 5 tested two predicted interactions between the distribution of existing features among category members and causal direction (Experiment 4) and causal strength (Experiment 5).

The primary purpose of these experiments is to provide strong tests of the causal-based generalization approach, an account that is described in detail immediately below. However, a secondary purpose is to allow CBG to be distinguished from an alternative model of how causal relations influence generalizations. Hadjichristidis, Sloman, Stevenson, and Over (2004) have proposed that a feature will be generalized to the extent that it exhibits *centrality*, that is, to the extent that it has many existing category features that are its *dependents* (i.e., effects). For example, Hadjichristidis et al. found that a novel hormone was generalized more strongly from one animal to another (e.g., from a seal to a dolphin) when the hormone was described as being depended on by ''many'' of the seal's physiological functions as compared to a ''few.'' Importantly, this sole emphasis on the feature's dependents or effects distinguishes the centrality account from CBG that claims that either a feature's effects *or* its causes can serve as the basis of its generalization. For example, for the materials in Table 1 the centrality account predicts that zinc-lined gas tanks will be generalized more strongly to Romanian Rogos only when they cause butane-laden fuel (giving them one dependent among Rogos' existing features) but not when they are an effect of butane-laden fuel (and thus have no dependents).

To foreshadow the results, the following experiments will demonstrate that a feature's generalization is indeed influenced by its causes every bit as much as its effects. But although this finding and numerous others are uniquely accounted for by CBG, the results will be shown to be not in perfect accord with its predictions. Although it appears that many participants generalize on the basis of causal inference much of the time, some appear to resort to a more associationist form of reasoning in which generalizations are made on the basis of only the strength of the link relating the novel feature to an existing category feature (i.e., in a manner that is insensitive to the direction of causality between the two). Thus, although I will argue that CBG is indispensable to any account of causal-based generalizations, it needs to be augmented with simpler inferential processes that reasoners may fall back on under certain conditions.

2. A model of causal-based property generalization

According to CBG, causal-based generalizations can be characterized as occurring in two steps. In the first, the reasoner computes, for every category member that he or she can think of, the likelihood that the to-be-generalized novel property is present in that exemplar. In the second, these probabilities are summed, each weighed by the probability that the exemplar is a category member. Specifically, if N is the novel property and K is the category, then the prevalence of N in K, P(N|K), is

$$\mathrm{P(N|K)} = \sum\nolimits_{\mathrm{M_i \in K}} \mathrm{P(N|M_i)P(M_i|K)} \tag{1}$$

where $\mathrm{M_i}$ is a member of K. The probability that a given category member has N, $\mathrm{P(N|M_i)}$, is determined by how N is causally related to the features of $\mathrm{M_i}$. I first derive $\mathrm{P(N|M_i)}$ for the special case where N is causally related to only one feature of $\mathrm{M_i}$. Predictions are then obtained regarding how $\mathrm{P(N|M_i)}$ varies as a function of the prevalence of that feature and the strength and direction of the causal link, predictions that are tested in Experiments 1 and 2. Later I derive $\mathrm{P(N|M_i)}$ for the case when N is causally linked to multiple features and test the resulting predictions in Experiments 3–5.

2.1. Single-link generalizations

Causal-based generalization's predictions are based on a particular representation of causal relations first introduced by Cheng (1997) and later applied to a variety of category-based tasks, including classification (Rehder, 2003a,b; Rehder & Hastie, 2001; Rehder & Kim, 2006) and feature prediction (Rehder & Burnett, 2005). On this account, the presence of some cause C enables a causal mechanism which brings about (with some probability *m* representing the strength or *power* of the cause) its effect E. However, E may also be brought about by one or more unspecified background causes. Given these and other reasonable assumptions (e.g., regarding the independence of causal mechanisms, see Cheng & Novick, 2005), the probabilities associated with each cell of the 2 × 2 contingency table involving C and E are

$$\begin{aligned}
&P(\sim\mathrm{C}\sim\mathrm{E}) = P(\sim\mathrm{C})(1-b)\\
&P(\sim\mathrm{CE}) \quad = P(\sim\mathrm{C})b\\
&P(\mathrm{C}\sim\mathrm{E}) \quad = P(\mathrm{C})(1-m)(1-b)\\
&P(\mathrm{CE}) \qquad = P(\mathrm{C})[1-(1-m)(1-b)] = P(\mathrm{C})(m+b-mb)
\end{aligned}$$

In other words, the probability that C and E will both be absent, $P(\sim\mathrm{C}\sim\mathrm{E})$, is the probability that C is absent times the probability that E is not brought about by any background cause (1 – *b*). Note that *m* is not involved in this likelihood because it is assumed that the causal mechanism relating C and E only potentially operates when C is present. The probability that C is absent but E is present, $P(\sim\mathrm{CE})$, is $P(\sim\mathrm{C})$ times the probability that E is brought about by some background cause, *b*. The probability that C is present but E absent, $P(\mathrm{C}\sim\mathrm{E})$, is $P(\mathrm{C})$ times the probability that E is not brought about the causal mechanism *and* not brought about by the background cause, (1 – *m*)(1 – *b*). Finally, the probability that C and E are both present, $P(\mathrm{CE})$, is $P(\mathrm{C})$ times the probability that E is brought about by the causal mechanism *or* brought about by the background cause (*m* + *b* – *mb*).

From these definitions, one can derive expressions for the probability of the effect E given either the presence or absence of its cause C:

$$\begin{aligned} P(E|C) &= P(CE)/P(C) \\ &= P(C)(m+b-mb)/P(C) \\ &= m+b-mb \end{aligned} \tag{2}$$

$$\begin{aligned} P(E|\sim C) &= P(\sim CE)/P(\sim C) \\ &= P(\sim C)b/P(\sim C) \\ &= b \end{aligned} \tag{3}$$

Note that Equation (2) indicates that when there are no background causes (b = 0), P(E|C) is simply the probability that E was generated by C, which is the causal strength m.

Similarly, the probabilities of C given either the presence or absence of E are

$$\begin{aligned} P(C|E) &= P(CE)/P(E) \\ &= P(CE)/[P(CE)+P(\sim CE)] \\ &= P(C)(m+b-mb)/[P(C)(m+b-mb)+P(\sim C)b] \\ &= P(C)(m+b-mb)/[P(C)m+b-P(C)mb] \end{aligned} \tag{4}$$

$$\begin{aligned} P(C|\sim E) &= P(C\sim E)/P(\sim E) \\ &= P(C\sim E)/[P(C\sim E)+P(\sim C\sim E)] \\ &= P(C)(1-m)(1-b)/[P(C)(1-m)(1-b)+P(\sim C)(1-b)] \\ &= P(C)(1-m)(1-b)/(1-P(C)m)(1-b) \\ &= P(C)[(1-m)/(1-P(C)m)] \end{aligned} \tag{5}$$

Given these definitions, I now derive $P(N|M_i)$ for when N is either a cause or an effect of a single feature of category member M_i.

2.1.1. *Novel property N as an effect*

When the novel feature N plays the role of an effect of a category feature F, then the probability of N in category members M_i that either do (M_i.F) or do not ($\sim M_i$.F) have feature F comes from Equations (2) and (3).

$$P(N|M_i.F) = m+b-mb \tag{6}$$

$$P(N|\sim M_i.F) = b \tag{7}$$

In the special case where N is related to only one category feature F, Equation (1) can be rewritten as two sums, one for those category members with F and the other for those without it.

$$\begin{aligned}
\mathrm{P(N|K)} &= \sum\nolimits_{\mathrm{i|M_i \in K \& M_i.F}} \mathrm{P(N|M_i.F)P(M_i|K)} + \sum\nolimits_{\mathrm{i|M_i \in K \& \sim M_i.F}} \mathrm{P(N|\sim M_i.F)P(M_i|K)} \\
&= \mathrm{P(N|M_i.F)} \sum\nolimits_{\mathrm{i|M_i \in K \& M_i.F}} \mathrm{P(M_i|}K) + \mathrm{P(N|\sim M_i.F)} \sum\nolimits_{\mathrm{i|M_i \in K \& \sim M_i.F}} \mathrm{P(M_i|K)} \\
&= \mathrm{P(N|M_i.F)P(M_i.F|K)} + \mathrm{P(N|\sim M_i.F)P(\sim M_i.F|K)} \qquad (8)
\end{aligned}$$

Substituting Equations (6) and (7) into Equation (8) yields

$$\mathrm{P(N|K)} = (m + b - mb)\mathrm{P(M_i.F|K)} + b\mathrm{P(\sim M_i.F|K)} \qquad (9)$$

Note that when $b = 0$, Equation (9) reduces to

$$\mathrm{P(N|K)} = m\mathrm{P(M_i.F|K)} \qquad (10)$$

that is, when N is an effect, the probability of N in K is just the probability of F in members of K times the probability that an F will produce N.

2.1.2. Novel property N as a cause

When N plays the role of a cause of a category feature F, then the probability of N in category members M_i that either do or do not have F comes from Equations (4) and (5).

$$\mathrm{P(N|M_i.F)} = \mathrm{P(N)}(m + b - mb)/[\mathrm{P(N)}m + b - \mathrm{P(N)}mb] \qquad (11)$$

$$\mathrm{P(N|\sim M_i.F)} = \mathrm{P(N)}[(1 - m)/(1 - \mathrm{P(N)}m)] \qquad (12)$$

Substituting Equations (11) and (12) into Equation (8) yields

$$\begin{aligned}
\mathrm{P(N|K)} =& [\mathrm{P(N)}(m + b - mb)/[\mathrm{P(N)}m + b - \mathrm{P(N)}mb]]\mathrm{P(M_i.F|K)} \\
&+ [\mathrm{P(N)}[(1 - m)/(1 - \mathrm{P(N)}m)]]\mathrm{P(\sim M_i.F|K)}
\end{aligned} \qquad (13)$$

When $b = 0$, Equation (13) reduces to

$$\mathrm{P(N|K)} = \mathrm{P(M_i.F|K)} + [\mathrm{P(N)}[(1 - m)/(1 - \mathrm{P(N)}m)]]\mathrm{P(\sim M_i.F|K)} \qquad (14)$$

and when $b = 0$ and $m = 1$,

$$\mathrm{P(N|K)} = \mathrm{P(M_i.F|K)}$$

that is, N is as prevalent in K as F is, because every category member with F also has N (a consequence of $b = 0$) and every one without F does not have N (a consequence of $m = 1$).

Experiments 1 and 2 now test two predictions that can be derived from the preceding equations, namely, how generalizations are affected by the prevalence of feature F in category members and the strength of the single causal link that relates N to the category.

3. Experiment 1

The purpose of Experiment 1 was to test CBG's predictions regarding how the generalization of a novel feature N is influenced by the prevalence of the category feature F to which it is causally related. As I now show, CBG's predictions correspond to the commonsense intuition that an effect is likely to be more prevalent when its cause is more prevalent and vice versa. For example, if one believes that the HIV virus has a number of potential effects (e.g., sarcoma, lymphoma, pneumonia, etc.), then one predicts those illnesses will be more likely in a population in which HIV is common as compared to one in which it is rare. Conversely, one infers that the known causes of HIV (e.g., blood transfusions, sharing of needles, and unsafe sex) are more likely to be present when HIV is common and less likely when it is rare.

Formally, proving that $P(N|K)$ increases montonically with $P(M_i.F|K)$ involves showing that $P(N|K; P(M_i.F|K) = f_0) < P(N|K; P(M_i.F|K) = f_1)$ for any $f_1 = f_0 + \Delta f$ and $\Delta f > 0$. From Equation (8) this inequality becomes

$$P(N|M_i.F)f_0 + P(N|{\sim}M_i.F)(1-f_0) < P(N|M_i.F)f_1 + P(N|{\sim}M_i.F)(1-f_1)$$

$$P(N|M_i.F)f_0 + P(N|{\sim}M_i.F)(1-f_0) < P(N|M_i.F)(f_0+\Delta f) + P(N|{\sim}M_i.F)[1-(f_0+\Delta f)]$$

which simplifies to $P(N|{\sim}M_i.F) < P(N|M_i.F)$. Thus, this inequality must be shown to hold both when N is a cause and when it is an effect.

When N is an effect, the inequality becomes [from Equations (6) and (7)], $b < m + b - mb$ which reduces to $b < 1$. That is, increasing the prevalence of feature F increases the prevalence of N as long as it is not already present in all category members. Note that the monotonically increasing relationship between $P(N|K)$ and $P(M_i.F|K)$ is transparent in Equation (10) for the special case $b = 0$.

When N is a cause, the inequality becomes [from Equations (11) and (12)],

$$[P(N)[(1-m)/(1-P(N)m)]] < [P(N)(m+b-mb)/[P(N)m+b-P(N)mb]]$$

which simplifies to $P(N) < 1$, again proving that the prevalence of N increases with F as long as N as long as it is not already present in all category members. That is, increasing the prevalence of feature F increases the prevalence of N (as long as it is not already present in all category members), regardless of whether N is a cause or an effect.

Experiment 1 tested these predictions by asking participants to generalize novel properties that were causally related to existing category features that varied in their prevalence (or *validity*, to use the term common to the categorization literature). Participants learned

Table 2
Category structure used in Experiment 1, consisting of eight Rogos (R1–8) and eight non-Rogos (NR1–8)

	D1	D2	D3	D4
Rogos				
R1	1	1	0	1
R2	1	1	0	1
R3	1	1	0	1
R4	1	1	1	0
R5	1	1	1	0
R6	1	1	1	0
R7	0	1	1	1
R8	1	0	1	1
Non-Rogos				
NR1	0	0	1	0
NR2	0	0	1	0
NR3	0	0	1	0
NR4	0	0	0	1
NR5	0	0	0	1
NR6	0	0	0	1
NR7	1	0	0	0
NR8	0	1	0	0

the validities of category features by learning to distinguish members of the target category from nonmembers (e.g., Romanian Rogos from non-Rogos) via standard classification-with-feedback training. The structure of the training exemplars is presented in Table 2. In the table a ‘‘1’’ represents a feature characteristic of the target category (e.g., a loose fuel filter gasket for Rogos), whereas a ‘‘0’’ represents an opposing value (e.g., a tight fuel filter gasket). Note that in this category structure dimensions 1 and 2 are of high validity because within each category the category's characteristic value appears in seven out of eight category members and thus are very useful in distinguishing category members from nonmembers. For example, dimension 1 has a ‘‘1’’ in seven out of eight Rogos and a ‘‘0’’ in seven out of eight non-Rogos. In contrast, dimensions 3 and 4 have lower validity because within each category the characteristic only appears in five out of eight cases.

After category learning, participants were asked to generalize four novel properties that were each causally related to one of the four existing category features. Thus, half of the novel properties were associated with high validity features (on dimensions 1 and 2), whereas the other half were associated with low validity features (on dimensions 3 and 4). Crossed with this factor was the direction of causality, as half the novel properties were described as effects, whereas the other half were described as causes. The predictions of CBG are that generalizations should be stronger for high- versus low-validity features, and this result should obtain regardless of causal direction.

These predictions can be contrasted with those of Hadjichristidis et al.'s (2004) centrality account. As mentioned, the centrality account predicts that features will be generalized more strongly as a function of their centrality, that is, the number of existing features that depend on them. Thus, for Experiment 1 centrality predicts that a novel feature will be generalized more strongly when it is a cause (and thus has one dependent) rather than an effect (and has none). While Hadjichristidis et al. have not discussed how validity affects generalizations, one might imagine that the centrality account could be extended to predict that a novel feature is more central (and thus more likely to be generalized) when its dependents are more prevalent. But given the absence of any role for causes of a to-be-generalized feature, the centrality account has no apparent basis for predicting the same when those causes are more prevalent. Experiment 1 tests the opposing predictions of CBG and the centrality account.

3.1. Method

3.1.1. Materials

Three novel categories were tested: Romanian Rogos, Lake Victoria Shrimp, and Myastars (a type of star). Each category was described as possessing four binary features. For each category there were also four novel features that participants were asked to generalize to the category, each of which was causally associated with one of the category's existing features. The novel feature could be described either as the cause or effect of the existing feature. The existing and novel features for Romanian Rogos have already been presented in Table 1. Those for the other two categories are available online at http://www.cogsci.rpi.edu/CSJarchive/Supplemental/index.html.

3.1.2. Participants

Twenty-seven New York University undergraduates received course credit for participating in this experiment. They were randomly assigned in equal numbers to one of the three categories.

3.1.3. Procedure

Experimental sessions were conducted by computer. Each session began by participants learning the category structure shown in Table 2 via standard classification-with-feedback training. The mapping of the four logical dimensions shown in Table 2 to the four category dimensions was randomized for each participant. An exemplar was presented on the computer screen with each dimension value listed on a separate line in a consistent order (dimensions 1–4). Participants classified the item either as a member of the target category (e.g., Romanian Rogo) by pressing the 1 key or as a nonmember (e.g., ''some other kind of automobile'') by pressing the 0 key, after which immediate feedback was provided. The 16 training exemplars were presented in blocks and the order of presentation of items within block was randomized. At the end of each block the participants were told the number of errors they had committed in the block. Training continued until two consecutive blocks were completed without error or a maximum of 20 blocks.

Participants then performed the generalization task. On each trial a causal law relating one of the category's existing features to a novel feature was presented on the top of the screen (e.g., for those participants assigned the Romanian Rogo category, one of those in Table 1). Participants were asked to judge what proportion of all category members had the novel feature ''given knowledge of this causal law.'' 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 category members possessed the novel property and 100 meant that they all did.

Participants were presented with a total of four generalization trials, one for each of the category's four novel features. Thus, two of these novel features were associated with the high-validity dimensions and two with the low-validity dimensions. Within each of these pairs, one novel feature was chosen at random to be described as the cause of the existing category feature, whereas the other was described as an effect. The experimenter first previewed each of the four generalization trials with the participant, and while doing so pointed out that the direction of causality was reversed on half the trials (nothing was mentioned about the differing validities of the existing features). The participant was then left alone to answer the four questions as they were redisplayed by the computer. The order of the questions was randomized for each participant.

3.2. Results

Analysis of the training phase of the experiment revealed that participants were generally successful at acquiring the category structure. Twenty-two of 27 participants reached the learning criterion in average of 14.0 blocks and the remaining five achieved an accuracy of .90 on the final block, indicating that all participants learned aspects of the statistical properties of the category structure, presumably including which dimensions were of high versus low validity.

Ratings from the generalization test that followed classification learning for all 27 participants are presented in Fig. 1 as a function of dimension validity and the direction of

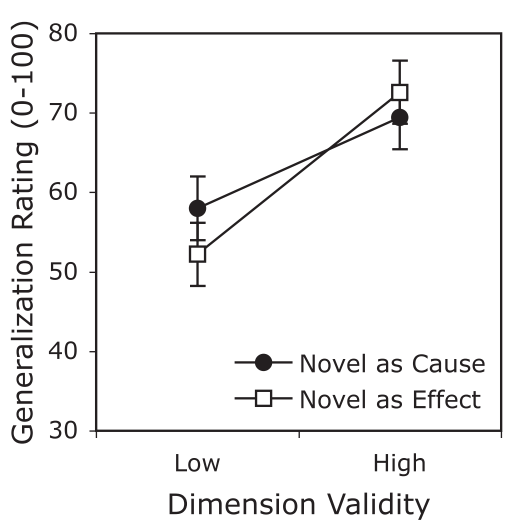


Fig. 1. Results from Experiment 1. Error bars are standard errors of the mean.

the causal relationship. The figure shows that generalization ratings were higher when the novel feature was causally related to a high-validity feature (average rating of 71) as compared to a low-validity one (57). Causal direction, in contrast, had no effect on generalizations. A 2×2 ANOVA with validity and direction as the within-subjects factors revealed a significant effect of validity, $F(1, 26) = 6.15$, $MSE = 859$, $p < .05$, no effect of causal direction, and no interaction (both Fs < 1). Note that additional analyses revealed no effect of the particular category that was learned (i.e., Rogos, Myastars, or Shrimp) on either learning or generalization.

3.3. Discussion

Experiment 1 provides a simple demonstration in favor of the causal-based generalization account: generalizations were stronger to the extent that the feature that served as the causal basis of that generalization was prevalent among category members, and this occurred despite the direction of causality.

One potential objection to this conclusion is to note that there are other ways in which the results in Fig. 1 may have obtained other than through participants' knowledge of dimension validity. For example, inspection of Table 1 indicates that to distinguish category members from nonmembers it is sufficient to learn the characteristic features on both the high-validity dimensions and on only one of the two low-validity dimensions and then apply a ''2 out of 3'' classification rule. Participants who learned the categories in this manner might then have responded with a low generalization rating for the novel feature that was causally related to the other, unlearned low-validity dimension. Although this behavior would explain the empirical results in Fig. 1 (because the two low-validity dimensions would have a lower average generalization rating than the two high-validity dimensions), it reflects not an effect on generalizations of dimension validity per se, but rather an effect of which dimensions were learned versus unlearned. To address this concern, my lab has run a version of this experiment in which participants were explicitly instructed on the validity of all four dimensions. Rather than performing a classification-with-feedback task, participants were first told that two of the characteristic features occurred in 90% of category members and that the other two occurred in 60% of category members, after which they answered the same generalization questions as in Experiment 1. The results were qualitatively the same as in Fig. 1; namely, there was a strong effect of feature validity and no effect of causal direction, as predicted by CBG (Nair, 2005).

Whereas the results of Experiment 1 and this follow-up experiment are consistent with the causal-based generalization account, they are inconsistent with the centrality account. First, that account predicted that generalizations should be stronger when a novel feature has more dependents (when it is a cause rather than an effect), but in fact there was no difference regarding whether the novel property was the cause or the effect. Second, it provides no account of the result that did occur, namely the influence of the existing feature's base rate on generalizations. Of course, given the centrality account's emphasis on a feature's effects, conceivable extensions could lead it to predict stronger generalizations when those effects are more prevalent. But its denial of any role for the feature's causes would seem to rule out an analogous extension sensitive to the prevalence of those causes.

4. Experiment 2

Whereas the purpose of Experiment 1 was to test the predictions of CBG by manipulating the prevalence of the existing feature, Experiment 2 does so by manipulating the strength of the causal relationship linking the existing and novel feature. CBG's predictions for how the prevalence of a novel feature N varies with causal strength are again best understood by separately considering the cases where it is a cause and the effect of the existing feature F. When it is an effect, the straightforward prediction is that N will be more prevalent when the causal mechanism that produces it is more likely to operate. For example, if HIV is present in a population, one more readily infers the presence of those symptoms that HIV generates with high probability than those it generates with low probability.

In contrast, when the novel property N is the cause of F, it should be *less* prevalent as the strength of the causal mechanism increases. This (perhaps less intuitive) prediction holds because a weaker causal link requires a cause to be more prevalent to account for the instances of an effect. For example, suppose you believe that the sole cause of HIV in a population is tainted blood transfusions. If you think that almost all blood transfusions transmit HIV (a strong causal link), then you predict that transfusions are about as common as cases of HIV. But if you think instead that blood transfusions transmit HIV rarely (a weak causal link), then you would have to assume a large number of transfusions to account for the observed rate of HIV.

Causal-based generalization thus predicts an interaction between causal strength and causal direction: a stronger causal link results in N being more prevalent when it is an effect but less prevalent when it is a cause. Formally, CBG's predictions regarding how generalizations are affected by causal strength are given by Equations (10) and (14) for the special case where there are no background causes (a condition that is satisfied in the upcoming experiment). When N is an effect, the fact that P(N|K) increases with m (so long as $P(M_i.F|K) > 0$) is directly apparent in Equation (10). When N is a cause, proving that the expression for P(N|K) given by Equation (14) decreases with m involves showing that $P(N|K; m = m_0) > P(N|K; m = m_1)$ for any $m_1 = m_0 + \Delta m$ and $\Delta m > 0$. From Equation (14) this inequality becomes

$$\begin{aligned}
&P(M_i.F|K) + [P(N)[(1-m_0)/(1-P(N)m_0)]]P(\sim M_i.F|K) > \\
&\qquad P(M_i.F|K) + [P(N)[(1-m_1)/(1-P(N)m_1)]]P(\sim M_i.F|K) \\
&(1-m_0)/(1-P(N)m_0) > (1-m_1)/(1-P(N)m_1) \\
&(1-m_0)/(1-P(N)m_0) > [1-(m_0+\Delta m)]/[1-P(N)(m_0+\Delta m)]
\end{aligned}$$

which reduces to $P(N) < 1$. That is, increasing causal strength m reduces P(N|K) so long as N is not already present in all category members.

To test CBG's predicted interaction between causal direction and causal strength, participants in Experiment 2 were asked to generalize a novel property with a causal link whose strength was described as either 67% or 100%. For example, for the first causal law in Table 1, participants were also told, ''Whenever an automobile has butane-laden fuel, it will cause that automobile to have a zinc-lined gas-tank X% of the time,'' where X was either 67 or 100. They were also told that there were no other causes of the effects (i.e., $b = 0$). Because generalizations are also affected by the prevalence of the feature to which N is causally related (as just demonstrated in Experiment 1), participants were also provided with information about the prevalence of the four category features. But rather than learning P(F|K) implicitly through classification learning, in this experiment participants were explicitly told that each feature occurred in 67% of category members. Fig. 2A presents CBG's quantitative predictions for P(N|K) as a function of causal strength (m = .67 or 1) and whether N is a cause or an effect, assuming (a) $b = 0$, (b) $P(M_i.F|K) = .67$, and (c) a noninformative prior on P(N) of .5.

The centrality account makes a different set of predictions for this experiment. First, as in Experiment 1, it predicts that the novel property should be generalized more strongly when it is a cause (and thus has one dependent) versus when it is an effect (and has zero). In addition, when the novel property is a cause, its centrality increases with the strength of the causal links with its dependent(s) (Sloman, Love, & Ahn, 1998), and thus so too should its generalization strength. In contrast, causal strength has no effect on the novel property's centrality when it is an effect, and thus neither should it affect its generalization. These ordinal predictions of the centrality account are depicted in Fig. 2B.

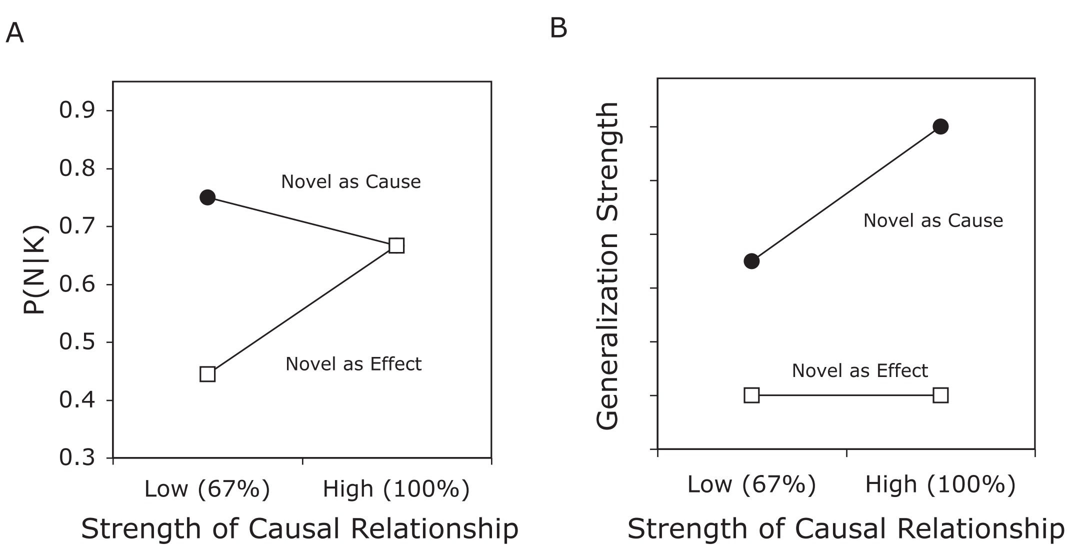


Fig. 2. Predictions for Experiment 2. (A) Predictions of CBG. (B) Predictions of the centrality account.

Whereas in Experiment 1 participants answered generalization questions about only one category, elimination of the lengthy classification training in Experiment 2 allowed each participant to answer generalization questions about all three categories (Rogos, Myastars, and Shrimp).

4.1. Method

4.1.1. Materials

The materials were the same as in Experiment 1.

4.1.2. Participants

Thirty-six New York University undergraduates received course credit for participating in this experiment. They were randomly assigned in equal numbers to one of three category presentation orders: msr, srm, or rms (r = Rogos, m = Myastars, s = Shrimp).

4.1.3. Procedure

For each of the three categories, participants first studied several screens of information that included the category's cover story, a description of its existing features, and the 67% base rates of those features. They then took a multiple-choice test that tested them on this knowledge. During the test participants could request help in which case the computer re-presented the screens of information about the category. However, participants were required to retake the test until they committed zero errors and made zero requests for help.

After performing a multiple-choice test, participants were presented with a generalization test for that category. A generalization trial followed the same general format as in Experiment 1 with the exception that the presentation of the causal law was more elaborate and thus was displayed on a screen separate from the generalization question. The causal law was described in four parts: (a) the causal relationship itself, (b) a description of the causal mechanism, (c) the strength of the cause (67% or 100%, as described above), and (d), the fact that were no other causes of the effect (e.g., ''Because there are no other causes of zinc-lined gas tanks, automobiles that don't have butane-laden fuel never have zinc-lined gas tanks.'').

The generalization question was presented on the next screen. To help participants remember the causal relation, it was rendered on the top of that screen as a diagram with an arrow linking the cause and effect features. The generalization question presented on the bottom of the screen was the same as in Experiment 1.

Participants were presented with a total of four generalization trials for each category, one for each of the category's four novel features. Half of the novel features were randomly chosen to be described as the cause of the existing feature, whereas the other half were described as the effect. Crossed with this factor was the strength of the causal mechanism: on half the trials the cause was described as producing its effect with probability 67% and with probability 100% on the other half.

4.2. Results

Initial analyses revealed no effect of the order in which the categories were presented, and thus the results are presented in Fig. 3A as a function of causal strength and causal direction. When the novel feature was an effect of an existing category feature, generalizations were stronger when causal strength was 100% (average rating of 69) versus 67% (52). However, this effect of strength was reversed when the novel feature was the cause rather than the effect (ratings of 67 and 73 for strengths of 100% and 67%, respectively). A 2×2 ANOVA with causal strength and causal direction as within-subjects factors revealed an effect of strength, $F(1, 35) = 6.68$, $MSE = 160$, $p < .005$, an effect of direction, $F(1, 35) = 28.37$, $MSE = 118$, $p < .0001$, and an interaction between the two, $F(1, 35) = 46.14$, $MSE = 99$, $p < .0001$.

Although the results confirm the interaction predicted by CBG, the predicted decrease in generalization ratings with causal strength when the novel property is a cause was only marginal, $t(35) = 1.86$, $p = .07$. However, one important question is whether the response pattern

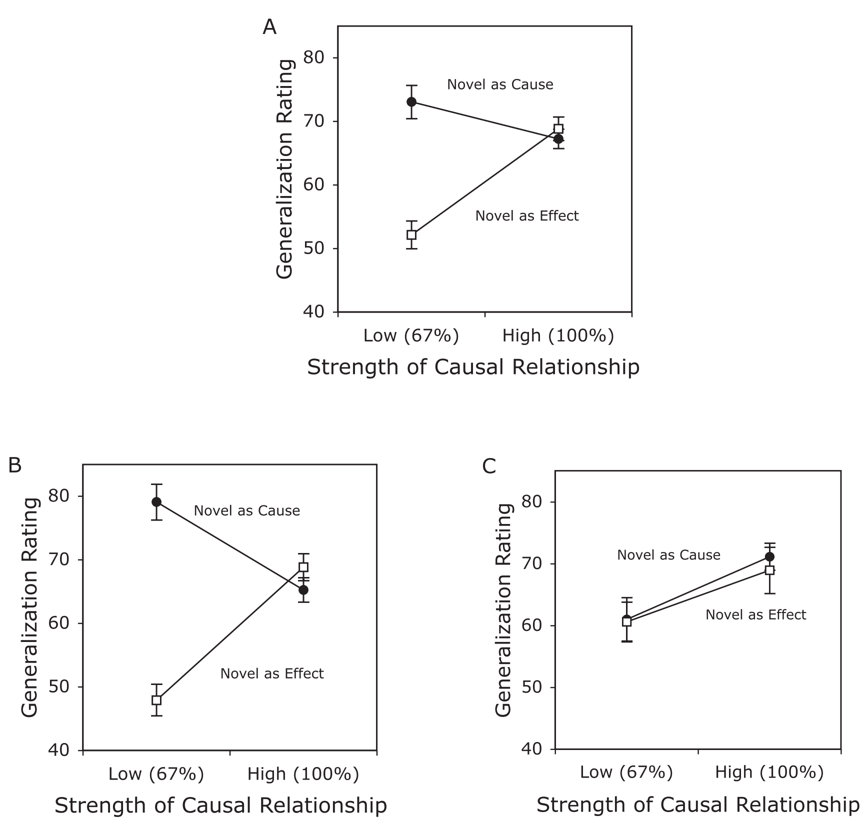


Fig. 3. Results from Experiment 2. (A) Average ratings for all 36 participants. (B and C) Two subgroups of 24 and 12 participants, respectively. Error bars are standard errors of the mean.

shown in Fig. 3A is manifested consistently by all 36 participants, or whether it arose as a result of averaging over individuals with substantially different response profiles. In fact, two groups of participants with qualitatively different responses were identified and are presented in Figs. 3B and 3C.[1] The majority group of 24 participants in Fig. 3B exhibited the predicted reversal: increasing causal strength led to higher ratings when the novel property was the effect but lower ratings when it was the cause. For this group of participants the 2×2 interaction was significant, $F(1, 23) = 224.05$, $MSE = 32$, $p < .0001$, and generalization ratings were significantly lower for causal strengths of 100% vs. 67% when the novel feature was the cause, $t(23) = 4.46, p < .001$. In contrast, the minority group of 12 participants in Fig. 3C produced higher ratings as causal strength increased. Note that despite the low statistical power associated with this small group, the effect of strength was significant, $F(1, 12) = 4.54$, $MSE = 226$, $p < .05$. In contrast to the majority group in Fig. 3B, this group exhibited no effect of causal direction and no strength-by-direction interaction, both ps $> .20$.

4.3. Discussion

The results from Experiment 2 support CBG's predicted interaction between causal strength and causal direction. Although increasing causal strength did not lead to a significant decrease in average generalization ratings when the novel feature was the cause, a more detailed analysis revealed that this pattern was in fact exhibited by the majority of participants. In contrast, for the remaining 12 participants stronger link strengths led to higher generalization ratings regardless of causal direction. This pattern of results suggests that these participants may have encoded the relationship between the existing and novel feature as a symmetrical association rather than an asymmetrical causal law, and thus were unable to appreciate the significance of the novel feature being a cause versus an effect. Additional evidence for the presence of this sort of associationist reasoning in a minority of participants will be presented in Experiments 4 and 5.

The results from Experiment 2 provided no support for the centrality hypothesis, because whereas that account predicted that causal strength should have an influence only when the novel feature was a cause (Fig. 2B), the empirical results exhibited the opposite pattern (Fig. 3A).

5. Causal-based generalizations for multiple links

The first two experiments investigated the effect on generalizations of the prevalence of the existing feature (Experiment 1) and the strength of the causal mechanism (Experiment 2). The remaining experiments test predictions derived from more complex causal networks, namely, the common cause and common effect networks shown in Fig. 4. In the common cause network, one binary variable is the cause of three others whereas in the common effect network one variable is an effect of three others.[2]

The likelihood equations for each of the 16 possible states of the four binary variables for both the common cause and common effect networks are now derived. It is assumed that for both networks the strength of the three causal links is equal, that is, $m_1 = m_2 = m_3 = m$. In addition, for the common cause network it is assumed that the strength of the alternative causes of each effect E_i is equal, $b_1 = b_2 = b_3 = b$. (Each of these assumptions is satisfied in the following experiments.) Define e_i as a variable representing whether E_i is present or absent and e as the total number of effects present. The likelihood equations for the common cause network are then given by

$$P(\sim Ce_1e_2e_3) = (1 - P(C))b^e(1 - b)^{3-e} \tag{15}$$

$$P(Ce_1e_2e_3) = P(C)(m + b - mb)^e[(1 - m)(1 - b)]^{3-e} \tag{16}$$

That is, when the common cause C is absent, the likelihood $P(\sim Ce_1e_2e_3)$ is $(1 - P(C))$ times b for each effect present and $(1 - b)$ for each effect absent. When C is present, the likelihood is P(C) times $(m + b - mb)$ for each effect present and $(1 - m)(1 - b)$ for each effect absent.

For the common effect network it is assumed that the probability of each cause is equal, that is, $P(C_1) = P(C_2) = P(C_3) = P(C)$. (An assumption also satisfied in the upcoming experiments.) Define c_i as a variable representing whether C_i is present or absent and c as the total number of causes present. The common effect likelihood equations are then

$$P(c_1c_2c_3 \sim E) = P(C)^c(1 - P(C))^{3-c}[(1 - m)^c(1 - b)] \tag{17}$$

$$P(c_1c_2c_3E) = P(C)^c(1 - P(C))^{3-c}[1 - (1 - m)^c(1 - b)] \tag{18}$$

Both likelihoods are multiplied by the probability that there are c causes present and the rest absent, $P(C)^c[1 - P(C)]^{3-c}$. When the common effect E is absent, the likelihood is further multiplied by the probability that neither any of the causes present nor the background cause produced it, $(1 - m)^c(1 - b)$. When E is present, it is further multiplied by 1 minus the probability that it is absent, $[1 - (1 - m)^c(1 - b)]$.

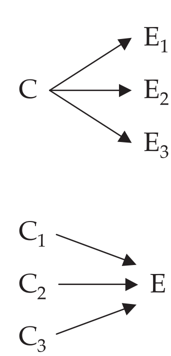


Fig. 4. Two multilink networks tested in Experiments 3–5.

Given the definitions above, it is straightforward to derive the probability of C (or E) conditioned on the state of its effects (or causes). For example, for the common cause network, if E_1 and E_2 are present but E_3 is absent, the probability of C is

$$\begin{aligned} P(C|E_1E_2{\sim}E_3) &= P(CE_1E_2{\sim}E_3)/P(E_1E_2{\sim}E_3) \\ &= P(CE_1E_2{\sim}E_3)/[P(CE_1E_2{\sim}E_3) + P({\sim}CE_1E_2{\sim}E_3)] \end{aligned}$$

which can be evaluated by substituting values from Equations (15) and (16)

$$\begin{aligned} P(C|E_1E_2{\sim}E_3) = P(C)(m+b-mb)^2(1-m)(1-b)/[P(C) \\ (m+b-mb)^2(1-m)(1-b) + (1-P(C))b^2(1-b)] \end{aligned}$$

Likewise, for the common effect network, if C_1 and C_2 are present but C_3 is absent, the probability of E is

$$\begin{aligned} P(E|C_1C_2{\sim}C_3) &= P(C_1C_2{\sim}C_3E)/P(C_1C_2{\sim}C_3) \\ &= P(C_1C_2{\sim}C_3E)/[P(C_1C_2{\sim}C_3E) + P(C_1C_2{\sim}C_3{\sim}E)] \end{aligned}$$

which can be evaluated using Equations (17) and (18).

$$P(E|C_1C_2{\sim}C_3) = [1-(1-m)^2(1-b)]/[[1-(1-m)^2(1-b)] + [(1-m)^2(1-b)]]$$

In the following experiments, a novel property N will play the role of either a common cause or a common effect of three category features. When N is a common cause of those three features, the probability that N is present in category member M_i is

$$P(N|M_i.f_1, M_i.f_2, M_i.f_3) = P(C|e_1e_2e_3)$$

where $M_i.f_i$ denotes the presence or absence of F_i in M_i. Similarly, when N is a common effect of three features, the probability of N in M_i is

$$P(N|M_i.f_1, M_i.f_2, M_i.f_3) = P(E|c_1c_2c_3)$$

Accordingly, Experiments 3–5 now test predictions that can be derived from Equations (15–18), namely, the effect on generalizations of the total number of causal links and interactions between category structure and the strength and direction of the those links.

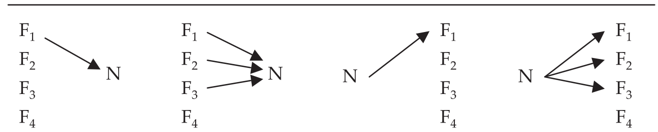
6. Experiment 3

The purpose of Experiment 3 was to test the effect of the number of causal links on generalizations. Participants were tested in the four conditions shown in the top of Fig. 5. In the figure, F_1, F_2, F_3, and F_4 represent four category features and N represents a novel feature.

Across conditions, N is either an effect of one or three category features or a cause of one or three category features. In all conditions participants were also told that each causal law had a strength of 75% (i.e., m = .75), that effects appeared in 25% of cases in which the explicit cause(s) are absent (i.e., b = .25), and that the category features F_1, F_2, F_3, and F_4 each occurred in 75% of category members.

Under these conditions, CBG's predictions regarding how generalizations should be affected by changes in the number of causal links are presented in the bottom of Fig. 5. These predictions are based on the assumption that category features are independent, that is, that there are no within-category correlations (the effect of this assumption will be assessed in Experiments 4 and 5), and a noninformative prior on N, P(N) = .5. Fig. 5 presents, for each of the potential 16 category members M_i that can be formed on four binary dimensions, the probability that the novel property N appears in M_i, P(N|M_i), as a function of causal network. For each M_i, a ''1'' means that it possesses a feature on that dimensions and a ''0'' means that the feature is absent. Due to independence, the probability that an M_i is a member of K, P(M_i|K), is $.75^f.25^{4-f}$, where f is the number of category features present in M_i. For the single-link conditions, P(N|M_i) is computed from Equations (6), (7), (11), and (12) presented earlier. For the multiple-link conditions, it is computed from Equations (15) and (16) when N is an effect and from Equations (17) and (18) when it is a cause. Finally, the figure presents the probability of N in the category, P(N|K), for each causal structure by summing over each M_i according to Equation (1).

Causal Structure



Category Structure			m = .75, b = .25	m = .75, b = .25	m = .75, b = .25	m = .75, b = .25
Uncorrelated						
	f_1 f_2 f_3 f_4	P(M_i | K)	P(N | M_i)	P(N | M_i)	P(N | M_i)	P(N | M_i)
M_1	1 1 1 1	.316	.813	.988	.765	.972
M_2	1 1 1 0	.105	.813	.988	.765	.972
M_3	1 1 0 1	.105	.813	.953	.765	.725
M_4	1 0 1 1	.105	.813	.953	.765	.725
M_5	0 1 1 1	.105	.250	.953	.200	.725
M_6	1 1 0 0	.035	.813	.953	.765	.725
M_7	1 0 1 0	.035	.813	.953	.765	.725
M_8	0 1 1 0	.035	.250	.953	.200	.725
M_9	1 0 0 1	.035	.813	.813	.765	.169
M_{10}	0 1 0 1	.035	.250	.813	.200	.169
M_{11}	0 0 1 1	.035	.250	.813	.200	.169
M_{12}	0 0 0 1	.012	.250	.250	.200	.015
M_{13}	0 0 1 0	.012	.250	.813	.200	.169
M_{14}	0 1 0 0	.012	.250	.813	.200	.169
M_{15}	1 0 0 0	.012	.813	.813	.765	.169
M_{16}	0 0 0 0	.004	.250	.250	.200	.015
P(N | K)=∑P(N | M_i)P(M_i | K)			.672	.937	.624	.738

Fig. 5. CBG's predicted values for the prevalence of a novel property N in category K, P(N|K), for the four causal structures tested in Experiment 3.

Fig. 5 indicates that CBG's predictions correspond to the common-sense intuition that evidence in favor of a cause (effect) is stronger when it has many versus few effects (causes) present. Specifically, CBG predicts that generalization should be stronger (.937 vs. .672) when N has three causes versus one. This prediction corresponds to the intuition that a property with multiple potential causes present in a population will be more prevalent because it is likely to be generated in more members of that population. For example, one's estimate of the prevalence of HIV will be greater when one is aware of multiple possible causes (e.g., frequent blood transfusions, sharing of needles, and unsafe sex) in the population in question as compared to just one (Fischoff, Slovic, & Lichtenstein, 1978; Tversky & Koehler, 1994). Likewise, CBG predicts that generalizations should be stronger (.738 vs. .624) when N has three effects versus one, corresponding to the intuition that a property will be more prevalent if many of its effects are prevalent. For example, one's estimate of the prevalence of the HIV virus in a population increases to the extent that members of that population exhibit many of its symptoms (e.g., sarcoma, lymphoma, pneumonia, etc.) versus only a few.

Note that the conditions in Fig. 5 provide another opportunity to compare the predictions of CBG and the centrality account. On the one hand, both accounts predict stronger generalizations when N has three effects versus one. The centrality account does so because N is more central when it has more dependents. Indeed, the two conditions in Fig. 5 in which N is a cause are similar to those tested in Hadjichristidis et al. (2004) who found stronger generalizations when N had more dependents. Unlike CBG however, the centrality account predicts no effect of the number of N's causes, because a feature's number of causes has no effect on its centrality. Of course, just as in Experiments 1 and 2, the centrality account also predicts that generalizations should be stronger whenever the novel property is a cause (and thus has one or more dependents) versus an effect (and has zero).

6.1. Method

6.1.1. Materials

Each novel feature was given two additional pairs of causal relations so that it could be causally related to up to three existing category features. For example, the first novel feature in Table 1 (zinc-lined gas tanks) could be related (either as a cause or effect) to not only butane-laden fuel but also loose fuel filter gaskets and hot engine temperature. Similarly, the second Rogo novel feature ''vibrations during braking'' could be related to not only loose fuel filter gaskets but also butane-laden fuel and hot engine temperature, and so on. The three pairs of causal relations for zinc-lined gas tanks are shown in Table 3. The complete list of causal links are available online at http://www.cogsci.rpi.edu/CSJarchive/Supplemental/index.html.

6.1.2. Participants

Thirty-six New York University undergraduates received course credit for participating in this experiment. Each participant learned and then answered generalization questions for

Table 3
Causal relationships linking one novel property—a zinc-lined gas tank—to three features of Romanian Rogos (used in Experiments 3–5)

Characteristic Feature	Novel Feature	Novel Feature as Effect	Novel Feature as Cause
Butane-laden fuel	Zinc-lined gas tank	Butane-laden fuel causes a zinc-lined gas tank. The butane interacts with the chromium in the metal of the gas tank, which results in a thin layer of zinc on the inside of the tank.	A zinc-lined gas tank causes the fuel to be butane laden. The zinc prevents corrosion of the tank, but it interacts chemically with gasoline to produce butane.
Loose fuel filter gasket	Zinc-lined gas tank	A loose fuel filter gasket causes a zinc-lined gas tank. To work properly, the gas tank needs to be sufficiently pressurized. Pressure becomes low when the fuel filter gasket is loose, and the insufficiently pressured fuel begins to deposit zinc on the inside of the gas tank.	A zinc-lined gas tank causes a loose fuel filter gasket. The zinc contaminates the fuel, and then corrodes the rubber gasket as it passes through the fuel filter.
Hot engine temperature	Zinc-lined gas tank	Hot engine temperature causes a zinc-lined gas tank. The excess heat is transmitted from the engine block back along the tail pipe, which passes close to the gas tank. The metal of the gas tank becomes warm, which starts a chemical reaction in which lead-free gasoline deposits zinc on the inside of the tank.	A zinc-lined gas tank causes a hot engine temperature. The zinc enters the fuel, which causes it to burn at a higher temperature inside the engine.

two out of the three categories. Accordingly, they were randomly assigned in equal numbers to one of three category presentation orders: ms, sr, or rm (r = Rogos, m = Myastars, s = Shrimp).

6.1.3. Procedure

For each of the two categories, participants first studied several screens of information about the category as they did in Experiment 2. Each feature was described as occurring in ‘‘most’’ category members. The generalization task also followed the same format as in Experiment 2 with the exception that on half the trials participants were presented with three screens presenting three causal laws rather than just one. The strength of each law was described as 75% and participants were told that the effect occurred in 25% of cases in which the causes(s) were absent. The screen that followed the presentation of the causal law(s) presented a diagram depicting the topology of those laws on the top and the generalization question on the bottom.

Participants answered four generalization questions for each category. Half of the novel features were randomly chosen to be described as a cause, whereas the other half were described as an effect. Crossed with this factor was the number of causal relations: on half the trials participants were presented with three causal laws and just one on the other half.

6.2. Results

Initial analyses revealed no effect of the order in which the categories were presented and thus the results are presented in Fig. 6 as a function of the number and direction of causal links. The results are straightforward: generalization ratings were higher when the novel feature was causally related via three causal links (average rating of 82) versus just one (76). Causal direction, in contrast, had no effect on generalizations. A 2 × 2 ANOVA with number and direction of causal links as the two within-subjects factors confirmed a significant effect of number, $F(1, 26) = 22.8$, $MSE = 86$, $p < .0001$, no effect of direction, and no interaction (both Fs < 1).

6.3. Discussion

As predicted by CBG, the results from Experiment 3 showed that more causal links lead to stronger generalizations and do so despite the direction of causality. In contrast, the findings did not support the predictions of the centrality account. On the one hand, the results from the conditions in which N was a cause replicate those reported by Hadjichristidis et al. (2004) who also found that features were generalized more strongly when they had a greater number of effects. However, the result that properties were also more strongly generalized

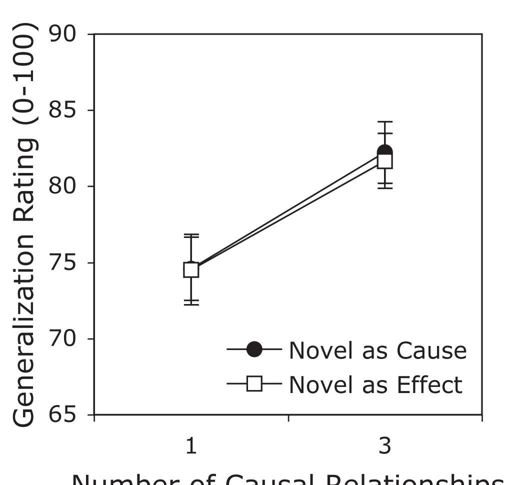


Fig. 6. Results from Experiment 3. Error bars are standard errors of the mean.

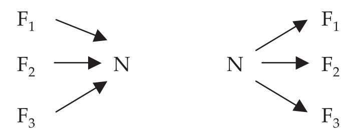
when they had a greater number of causes suggests that those previous findings were not due to centrality per se. Instead, the more parsimonious explanation is that in order to estimate the prevalence of a new feature in a population people reason causally (either forward or backwards) from its causes and effect.

Note that the results of Experiment 3 were not in perfect accord with CBG's predictions in Fig. 5, which indicates that while increasing the number of causal links strengthens generalizations when N is either a cause or an effect, that increase should be greater when it is an effect. Recall, however, that the predictions in Fig. 5 were based on the assumption that category features were independent, that is, uncorrelated. However, the presence of causal relationships connecting those features (indirectly through N) may have rendered that assumption invalid. Indeed, an intriguing observation from Experiment 3 was that a substantial number of participants requested additional information while performing the generalization task (or mentioned in postexperiment interviews that they had not been given sufficient information). Specifically, they asked whether the fact that category features occurred in 75% category members meant that most category members had all four features or whether the features were dispersed evenly through all members. In other words, they asked whether the category features were correlated or uncorrelated. Experiments 4 and 5 now demonstrate how an assumption of correlated versus uncorrelated category features affects causal-based generalizations.

7. Experiment 4

The purpose of Experiment 4 was to test how category structure (uncorrelated or correlated category features) interacts with the direction of multiple causal links. Fig. 7 presents CBG's predictions when a novel feature N is related to three category features as either a cause or an effect as a function of whether the category structure is uncorrelated or correlated. Both category structures consist of four exemplars. But whereas the features are distributed evenly through the four exemplars in the uncorrelated structure they are clustered together in the correlated structure (three exemplars have all three features and the fourth has none). Note that the feature bases rates are .75 in both category structures, because in both each feature appears in three out of the four exemplars. However, Fig. 7 indicates that these category structures should lead to different generalizations. The figure presents, for each exemplar for each category and causal structure, the likelihood that a category member will have novel property N, $P(N|M_i)$, derived from Equations (15–18), assuming a causal power of $m = 1$ and background causes of strength $b = .25$.

First consider the case in which N is an effect of three category features. For the uncorrelated category structure, N should be present in all four category members because each category member has at least one of N's causes and each of those causes is guaranteed to produce N. In contrast, for the correlated structure, three of the four category members have at least one cause but the fourth has none, and so for that category member $P(N|M_i)$ should be solely determined by other potential causes. Because of that fourth exemplar, $P(N|K)$ for the entire category should be less than 1 (.813). In other words, when N is an effect of three



Category Structure			Causal Structure: F1, F2, F3 → N, $m = 1, b = .25$	Causal Structure: N → F1, F2, F3, $m = 1, b = .25$
Uncorrelated				
	$f_1\ f_2\ f_3$	$P(M_i \mid K)$	$P(N \mid M_i)$	$P(N \mid M_i)$
M_1	1 1 1	.250	1	.985
M_2	1 1 0	.250	1	0
M_3	1 0 1	.250	1	0
M_4	0 1 1	.250	1	0
$P(N \mid K)=\sum P(N \mid M_i)P(M_i \mid K)$			1	.246
Correlated				
	$f_1\ f_2\ f_3$	$P(M_i \mid K)$	$P(N \mid M_i)$	$P(N \mid M_i)$
M_1	1 1 1	.250	1	.985
M_2	1 1 1	.250	1	.985
M_3	1 1 1	.250	1	.985
M_4	0 0 0	.250	.250	0
$P(N \mid K)=\sum P(N \mid M_i)P(M_i \mid K)$			.813	.738

Fig. 7. CBG's predicted values for the prevalence of a novel property N in category K, P(N|K), for the two causal structures and the two category structures tested in Experiment 4.

category features generalizations should be stronger for the uncorrelated structure as compared to the correlated structure.

The reverse is true when N is the cause of three category features. For the uncorrelated structure, N is highly likely to be present in the one exemplar that possesses all three features. But N must be absent in the remaining three exemplars, because each is missing one feature and N has been described as a perfectly reliable cause of all three features. In the correlated structure, N is highly likely to be present in the three exemplars in which all three features are present and absent in the fourth. That is, when N is a cause, generalizations should be stronger for the correlated structure as compared to the uncorrelated structure.

These analyses explain why several participants in Experiment 3 thought the distribution of existing category features was relevant to answering the generalization questions. They also provide an opportunity to conduct another test of CBG. Experiment 4 tested the predicted interaction between category structure and causal direction suggested by Fig. 7. On all generalization trials participants were not only presented with three causal laws (each of strength 100%) in which N was sometimes the cause and sometimes the effect, they were also shown a sample of category members that manifested either a correlated or uncorrelated structure. The samples were presented as a set of cards in which each card listed the features of one exemplar. The samples were of size 12, constructed by tripling the four exemplars in the category structures in Fig. 7. Participants were asked to estimate P(N|K) on the basis of both the causal laws and the sample. Because of the greater complexity of the generalization task, each participant learned and answered questions for just one category. Because one category can only be associated with one sample, category structure (correlated vs. uncorrelated) was manipulated as a between-subject variable. In order to simplify the reasoning required of participants, each category was described as having just three existing features (as compared to four in Experiments 1–3), as shown in Fig. 7.

Of course, Experiment 4 also provides another test the centrality model. For the causal structures in Fig. 7, that account predicts that the novel feature should be generalized more strongly when it is a cause (and has three dependents) as compared to when it is an effect (and has zero). Moreover, because its predictions are based solely on a property of the to-be-generalized feature (its number of dependents), the centrality account predicts no interaction between causal structure and the distribution of category features.

7.1. Method

7.1.1. Materials

The additional causal relations needed to instantiate the design of Experiment 4 are available at http://www.cogsci.rpi.edu/CSJarchive/Supplemental/index.html.

7.1.2. Participants

Twenty-four New York University undergraduates received course credit for participating in this experiment. They were assigned in equal numbers to the three categories and to the correlated and uncorrelated category structures.

7.1.3. Procedure

Participants learned the category's features and that they appeared in ''most'' category members. The generalization task that followed was conducted as a paper- and pencil-task. Participants were first presented with 12 cards representing 12 different category members. Each card listed the three features for that exemplar. The cards instantiated either a correlated or uncorrelated structure as shown in Fig. 7. Participants were told that the cards were a representative sample of the category.

The generalization questions that followed were each presented on two sheets of paper. The first sheet described three causal laws and each description specified the cause and effect variable, the causal mechanism, and the strength of that mechanism (100%). Participants were told that the effect occurred in 25% of cases in which the causes were absent. The second sheet included a diagram of the causal relations on the top and the generalization question (which were the same as in the previous experiments). The experimenter told the participant that their answer should be based on the different sources of information provided (which were enumerated as the feature base rates learned via the computer, the sample of 12 cards, and the three causal laws), but that it was up to them to decide which pieces of information to use. Participants responded by writing down the proportion of category members they thought would have the novel property.

Participants answered a total of four generalization questions. Half of the novel features were randomly chosen to be described as a cause of the three category features, whereas the other half were described as their effect. To simplify presentation of the four questions, for half the participants the novel property was the cause in the first two questions and the effect in the second two, and this order was reversed for the other half. (Presentation order was counterbalanced over participants.) The experimenter pointed out the change in causal direction that occurred between the first two questions and the last two.

7.2. Results

Initial analyses again revealed no effect of which category was learned, and thus the results are presented in Fig. 8A as a function of category structure and causal direction. This figure confirms the presence of an interaction between structure and direction: when the novel feature was an effect of the existing category features, generalizations were stronger

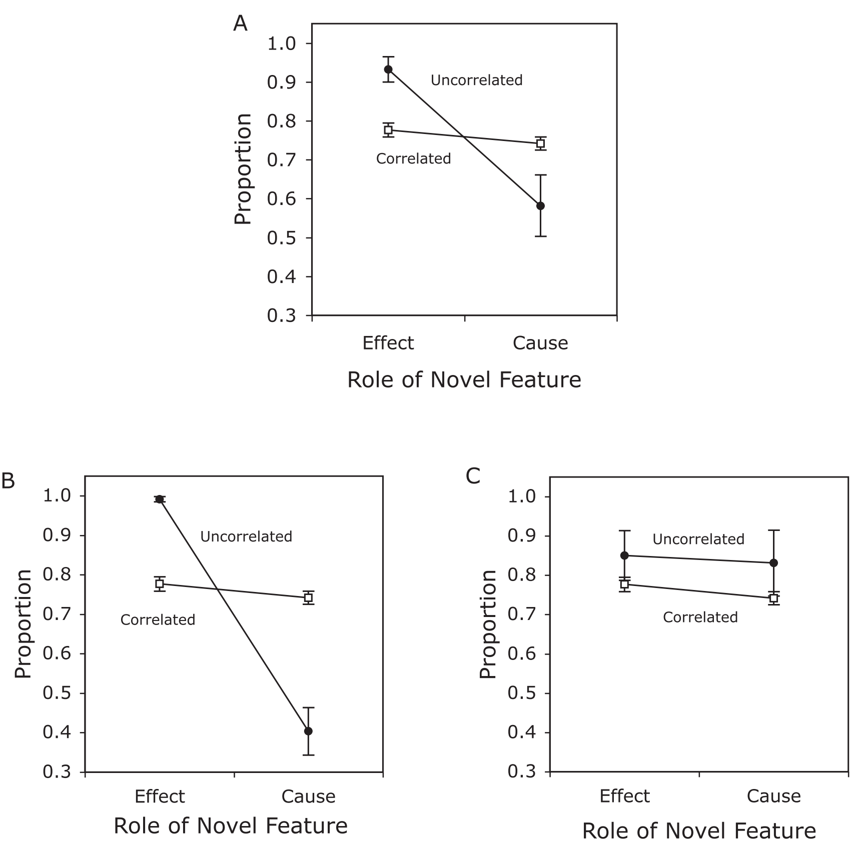


Fig. 8. Results from Experiment 4. (A) Average ratings for all participants. (B and C) Two subgroups of seven and five uncorrelated participants, respectively. Error bars are standard errors of the mean.

for the uncorrelated (.93) versus the correlated category structure (.78). This effect of category structure was reversed when the novel feature was the cause (.58 vs. .74). A 2×2 ANOVA with category structure as a between-subject factor and causal direction as a within-subject factor revealed a main effect of direction, $F(1, 22) = 15.7$, $MSE = .028$, $p < .001$ and an interaction between direction and category structure, $F(1, 22) = 10.5$, $MSE = .028$, $p < .005$.

Although the results confirm the interaction predicted by CBG, one surprising result was the strength of generalizations for the uncorrelated category structure when the novel property was the cause. In this condition, CBG predicted P(N|K) should be .246 (Fig. 7), whereas participants produced an average proportion of .58. One possibility is that responses were very different from one another in that condition, a conjecture supported by its unusually large standard error. In fact, two groups of uncorrelated participants with qualitatively different responses were identified and are presented in Figs. 8B and 8C.[3] (For purposes of comparison, Figs. 8B and 8C also present the results of the correlated group, which exhibited no systematic individual differences.) For the group of seven uncorrelated participants in Fig. 8B, their predicted proportion of .40 when the novel property was the cause was closer to the theoretically predicted value of .246. Despite low statistical power, this group exhibited a large effect of causal direction, $t(6) = 10.0$, $p < .0001$. In contrast, the group of five participants in Fig. 8C showed no sign of any effect of causal direction ($t < 1$), suggesting that they treated the causal links as a symmetrical relation. These findings are similar to those in Experiment 2 that also found that a minority of participants treated the causal relations symmetrically.

7.3. Discussion

Experiment 3 demonstrated how generalizations become stronger as a function of a novel feature's number of causal links regardless of causal direction. However, according to CBG (and a number of Experiment 3's participants), the prevalence of a novel property with multiple causal links depends on the distribution of category features that are related to that property. In fact, the results of Experiment 4 exhibited the predicted interaction between category structure and the causal role of the novel property.

This interaction between causal structure and category structure can also be understood as follows. Because one generally expects the effects of a single common cause to be highly correlated, the correlated structure is more in line with what one would expect when N is the cause of three category features. For example, if HIV always produces each of its symptoms, then one expects an HIV patient to display all those symptoms. But even if the causes of HIV are 100% reliable, one does not expect those causes to be correlated, and thus the uncorrelated structure is more in line with what one would expect. The results of Experiment 4 indicate that when evaluating a potential generalization reasoners not only take into account the properties of individual causal links such as the base rate of the cause or effect (Experiment 1) and link's causal strength (Experiment 2), they also consider the overall pattern of evidence exhibited by multiple category features to evaluate the prevalence of a novel property in light of the entire causal network relating the property and the category.

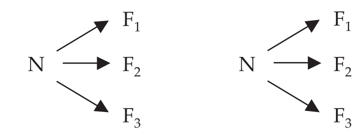
However, not all uncorrelated participants exhibited the predicted decrease in generalization strength when the novel property was the cause as compared to an effect. Instead, as was the case in Experiment 2, a substantial minority of participants appeared to treat the relation as symmetrical association. I will return to this issue in the General Discussion.

The centrality account is unable to account for the main results of this experiment. For the category structures in Fig. 7, that account predicts that the novel feature should have been generalized more strongly when it was a cause (when it had three dependents) as compared to when it was an effect (when it had zero), but in fact the main effect of causal direction was in the opposite direction. The centrality account also failed to predict the interaction between causal structure and the distribution of category features. This limitation means the centrality account will also be unable to account for the results of the next experiment that presents another sort of interaction between these factors.

8. Experiment 5

Experiment 5 presents one final test of CBG by testing a second kind of interaction between a novel feature's causal laws and category structure. All participants were taught a causal structure in which a novel feature N was a cause of three category features. However, rather than the strength of the causal links being fixed at 100% as in Experiment 4, in Experiment 5 causal strength was varied between 67% and 100%. CBG's predicted interaction between causal strength and category structure is presented in Fig. 9. The two conditions with causal strengths of 100% (right side of Fig. 9) have already been tested in Experiment 4: it was predicted (and found) that generalizations should be stronger for the correlated versus the uncorrelated category structure. But this predicted difference is reduced (indeed, is slightly reversed) for causal strengths of 67%. For the uncorrelated structure, because the absence of one feature in an exemplar now does not count as decisive evidence against the presence of N, N is likely to be present in all four exemplars (with probability .964 for exemplar 111 and .750 for the other three exemplars). In the correlated structure, N is almost certainly present in three of four exemplars but is almost certainly absent in the fourth. Thus, generalizations should be about the same or slightly stronger for the uncorrelated versus the correlated category structure for causal strengths of 67%.

This predicted interaction between causal strength and category structure can also be understood as follows. When causal links are deterministic, one expects the effects of a common cause (N) to be highly correlated, and thus a correlated structure is more in line with what one would expect if N was present. But when causal links are probabilistic, one expects the effects to be more weakly correlated, and thus the uncorrelated structure becomes more consistent with N. For example, if one knows that the HIV virus always produces its symptoms, then the presence of HIV is likely in an individual who exhibits all of those symptoms but unlikely if even one is absent. But if the



Category Structure			Causal Structure: $m = .67$, $b = .25$	Causal Structure: $m = 1$, $b = .25$
Uncorrelated				
	$f_1\, f_2\, f_3$	$P(M_i \mid K)$	$P(N \mid M_i)$	$P(N \mid M_i)$
M_1	1 1 1	.250	.964	.985
M_2	1 1 0	.250	.750	0
M_3	1 0 1	.250	.750	0
M_4	0 1 1	.250	.750	0
$P(N \mid K)=\sum P(N \mid M_i)P(M_i \mid K)$			.804	.246
Correlated				
	$f_1\, f_2\, f_3$	$P(M_i \mid K)$	$P(N \mid M_i)$	$P(N \mid M_i)$
M_1	1 1 1	.250	.964	.985
M_2	1 1 1	.250	.964	.985
M_3	1 1 1	.250	.964	.985
M_4	0 0 0	.250	.036	0
$P(N \mid K)=\sum P(N \mid M_i)P(M_i \mid K)$			.732	.738

Fig. 9. CBG's predicted values for the prevalence of a novel property N in category K, P(N|K), for the two causal structures and the two category structures tested in Experiment 5.

symptoms are produced probabilistically, the presence of HIV is not ruled out when just one is missing.

The centrality account again makes a different set of predictions. First, it predicts that generalizations should be stronger when causal strengths are 100% vs. 67%, because the novel feature will be more central in the former condition versus the latter. But, as mentioned, the centrality account is unable to predict any interaction between category structure and causal structure. These contrasting predictions of CBG and the centrality account are tested in Experiment 5 in which category structure was manipulated as a between-subject factor and causal strength as a within-subject factor.

8.1. Method

8.1.1. Materials

The materials were the same as those in Experiment 4, except that half of the causal links were described as having a strength of 67% rather than 100%.

8.1.2. Participants

Forty-eight New York University undergraduates received course credit for participating in this experiment. They were assigned in equal numbers to the three categories and to the correlated and uncorrelated category structures.

8.1.3. Procedure

The procedure was generally the same as in Experiment 4. For each participant, half of the novel features were randomly chosen to have a causal link strengths of 100%, whereas the other half had strengths of 67%. As in Experiments 3 and 4, participants were told that the effect occurred in 25% of cases in which all three causes were absent. Half the participants were presented the strong novel properties in the first two generalization questions followed by two weak ones; this order was reversed for the other half of participants. The experimenter pointed out the change in causal strength that occurred between the first two questions and the last two.

8.2. Results

Initial analyses again revealed no effect of which category was learned, and thus the results are presented in Fig. 10 as a function of category structure and causal strength. When

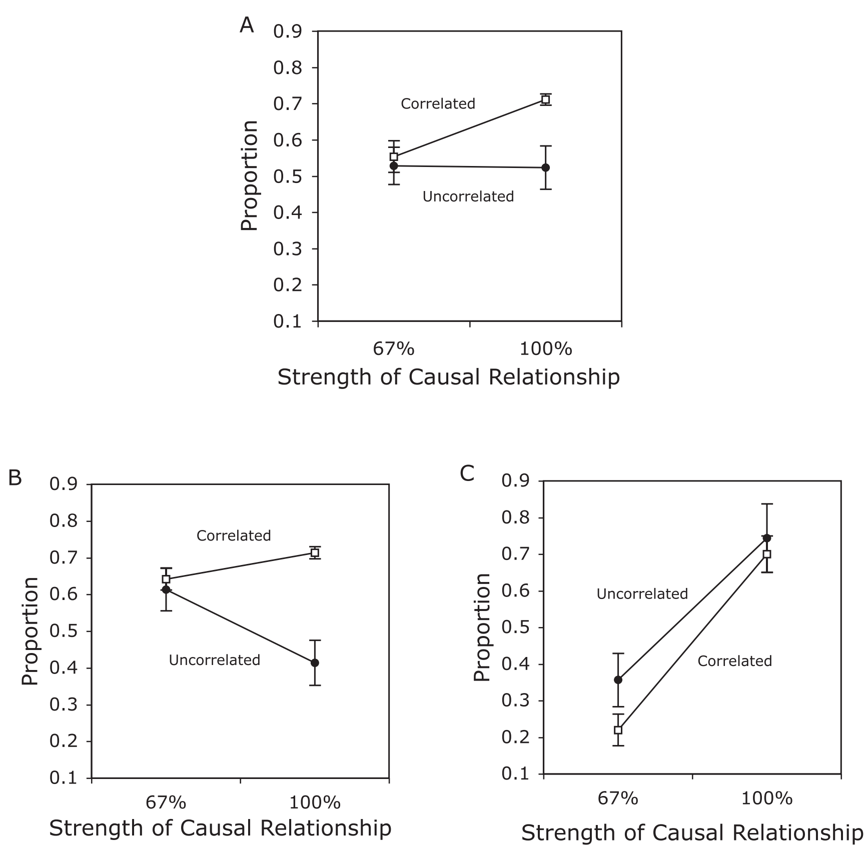


Fig. 10. Results from Experiment 5. (A) Average ratings for all participants. (B) Subgroups of 19 and 16 correlated and uncorrelated subjects. (C) Subgroups of five and eight correlated and uncorrelated subjects. Error bars are standard errors of the mean.

causal strength was 100%, generalizations were stronger for the correlated versus the uncorrelated category structures (.71 vs. .54), replicating the results from Experiment 4. But this effect was eliminated for causal links of strength 67%, confirming the predicted interaction between causal strength and category structure. A 2 × 2 ANOVA with category structure as a between-subject factor and causal strength as a within-subject factor revealed a marginal effect of causal strength, $F(1, 46) = 3.81$, $MSE = .037$, $p = .06$, a main effect of category structure, $F(1, 46) = 4.29$, $MSE = .063$, $p < .05$, and the predicted interaction between strength and structure, $F(1, 46) = 4.23$, $MSE = .037$, $p < .05$.

Nevertheless, although the predicted interaction between strength and structure obtained, its form was different than predicted. Whereas CBG predicted that generalizations should decrease (from .803 to .246) as causal strengths increase for the uncorrelated structure, Fig. 10 indicates that ratings were in fact unaffected by strength in that condition. Once again, however, examination of individual responses revealed considerable clustering of the data.[4] Within the correlated group, two subgroups of 19 and 5 participants with qualitatively different responses were identified. Within the uncorrelated group there were two subgroups of 16 and 8 participants. These subgroups are displayed in Figs. 10B and 10C.

The correlated and uncorrelated subgroups in Fig. 10B ($n = 19$ and 16, respectively) exhibit the predictions of CBG in Fig. 9. Not only is there a significant interaction between strength and structure, $F(1, 33) = 26.14$, $MSE = .0123$, $p < .0001$, ratings decreased with strength in the uncorrelated condition, $t(18) = 4.50$, $p < .0001$, and increased with strength in the correlated condition, $t(18) = 2.30$, $p < .05$. In contrast, the subgroups in Fig. 10C ($n = 5$ and 8) manifest an effect of causal strength only. Despite low statistical power, these subjects exhibited a large effect of strength, $F(1, 11) = 295.15$, $MSE = .0039$, $p < .0001$. (The hint of an interaction between strength and structure in Fig. 10C did not reach significance, $F(1, 11) = 3.95$, $MSE = .0039$, $p = .09$.) That is, as was the case in Experiments 2 and 4, it appears that a substantial minority of participants treated the causal links as a symmetrical relation in which the strength of that relation was the only factor relevant to the generalization. As a consequence, they failed to appreciate the implications that causal structures have for the distribution of features across category members.

8.3. Discussion

Experiment 5 provides another demonstration of how reasoners consider the pattern of evidence exhibited by multiple features when computing generalizations. One expects features with a common cause to be highly correlated when causal links are strong and less so when they are weak. As predicted by CBG, the correlated structure led to stronger generalizations than the uncorrelated structure when causal strengths were 100% and this difference was eliminated for causal strengths of 67%.

Although the centrality account correctly predicted the main effect of causal strength in this experiment, it failed to predict the interaction between causal structure and category structure. Interestingly, however, the presence of a minority group of 13 subjects that exhibited an effect of causal strength but no interaction with category structure is suggestive of the possibility that at least some people generalize in a manner consistent with the centrality model. But working against this interpretation is that fact that that model did not account for the minority groups in either Experiment 2 (compare Fig. 3C with Fig. 2B) or Experiment 4 (the centrality account predicts stronger generalization when N is a cause, an effect that is absent in Fig. 8C). Instead, a more parsimonious explanation of the minority groups in Experiments 2, 4, and 5 is that they were reasoning as if the causal relationship was a bidirectional associative link, always producing stronger generalizations for stronger links. I will return to this point in the General Discussion.

9. General discussion

The primary purpose of this article is to present evidence for the role of causal reasoning in the generalization of new properties. In the first section, I review the evidence in favor of this claim and the support it provides for a model of the reasoning processes that underlie causal-based generalizations. In the following section I discuss the evidence that was also found for a simpler, more associative reasoning account in a minority of participants. I then discuss the implications this work has for the alternative centrality account and also to what extent the current findings can be accommodated by more sophisticated versions of similarity. The article closes with a discussion of other forms of causal-based generalizations.

9.1. Causal-based property generalization

Experiments 1–5 were each designed to test a prediction of a causal-based approach to generalization. In Experiment 1, it was predicted that generalizations should be stronger as a function of the prevalence of the existing feature to which the to-be generalized property was causally related, regardless of whether that property was a cause or an effect. This result obtained both when prevalence was conveyed to participants as an explicit base rate and implicitly through supervised category learning. In Experiment 2, it was predicted that an interaction should obtain between causal direction and causal strength, a prediction confirmed by the experimental results. In Experiment 3 a greater number of causal links led to stronger generalizations, and they did so regardless of direction. Experiment 4 elaborated on this result, showing that the effect of the number of causal links depends on the distribution of features among category members. Finally, in Experiment 5 it was predicted (and found) that the distribution of features among category members should also interact with the strength of the novel property's causal laws. In sum, all five experiments provided support for the role of causal reasoning in generalization.

Two properties of CBG led to its success. One is the assumption that generalizers will infer whether individual category members are likely to possess a new property on the basis of the causal links that relate them. To model this reasoning, CBG assumes a representation of causal relations that has been successfully applied to causal learning (Cheng, 1997; Buehner, Cheng & Clifford, 2003), classification (Rehder, 2003a,b; Rehder & Kim, 2006), and prediction (Rehder & Burnett, 2005). A key property of this representation is that it captures the fundamentally asymmetric nature of causal relations, and the interactions involving causal direction in Experiments 2 and 4 indicate that participants were indeed sensitive to the direction of the causal arrow. At a minimum, these results sharply rule out any account that represents causality as a mere symmetric relation.

Of course, the present experiments were not specifically designed to test CBG's model of causal reasoning, so the present results might be consistent with other accounts that incorporate these sorts of reasoning asymmetries. Other research, however, provides data on how CBG fares as a model of human causal reasoning. According to CBG, a category's network of interfeature causal links are instances of *Bayesian networks* or *causal graphical models*, a formalism with roots in artificial intelligence that constrains causal inferences in a manner that is considered normative (Glymour, 1998; Glymour & Cheng, 1998; Pearl, 1988, 2000). Rehder and Burnett (2005) tested whether people's causal inferences of missing category features was consistent with this account. Although they found that inferences were generally in accord with the normative model, they also observed systematic deviations known as *screening off errors* in which variables that should be treated as conditionally independent were treated as dependent instead (also see Mayrhofer, Goodman, Waldmann, & Tenenbaum, 2008). CBG would also fail to predict screening off errors and thus would mispredict people's generalizations in some conditions. To account for their results, Rehder and Burnett proposed that reasoners were in fact reasoning normatively but with additional knowledge structures that led to the (apparent) screening of errors. Thus, CBG might need to be augmented with this assumption or some other to correctly predict generalizations in situations susceptible to such errors.

One potential concern with the present research is that the causal knowledge was provided as part of the experimental session. A critic might argue that this practice made that knowledge especially salient and available, bypassing the need for the reasoner to retrieve it from their semantic memory. Further, many of the experiments provided very concrete information about both the base rate of category features and the strengths of causal relations, and it is reasonable to ask whether people consider such information during most everyday generalizations. However, many of the studies reviewed earlier showed an effect of real-world causal knowledge on inductions involving natural categories (Bailenson et al., 2002; Lopez et al., 1997; Medin et al., 2003; Proffitt et al., 2000; Shafto & Coley, 2003), suggesting that recruiting causal knowledge in the service of generalizations may be the norm rather than the exception. It is also likely that reasoners recognize that the strengths of the causal relationships are relevant for making well-founded generalizations. For example, when information about the strengths of causal links and background causes was omitted in an earlier version of Experiment 3, several subjects spontaneously requested this information, indicating that they knew such information was important for the judgment they were being asked to make.

The second key assumption of CBG is that reasoners will generalize by computing the likely presence of the novel property in individual category members and then compute a weighted average over all category members. This assumption enabled CBG to account for the interactions between category structure and causal direction in Experiment 4 and causal strength in Experiment 5. Are human reasoners likely to think about multiple category members when generalizing? In Experiments 4 and 5 at least, considering individual category members was likely promoted by presenting participants with cards displaying each exemplar's features. In real-world settings, people might exhibit less sensitivity to feature distribution information because it requires retrieving examples of category members from memory. On the other hand, the reasoning literature is rife with examples of how the *availability* of relevant examples is a key factor determining people's judgments in a variety of domains (e.g., Kahneman & Tversky, 1973). And, as mentioned, when feature distribution information was omitted (in Experiment 3), several subjects spontaneously asked for that information (or mentioned at the end of the experiment that they noticed its absence), suggesting people are aware of the relevance of this factor. Future research may establish how those category members that are available due to the fact they are typical (Rosch & Mervis, 1975) or ideal (Barsalou, 1985; Lynch, Coley, & Medin, 2000; Rehder & Hastie, 2004) are those that have the greatest impact on causal-based generalizations.

Finally, it is worth noting that another important assumption of CBG is that the causal reasoning that underlies generalizations is domain general. This claim received support from the fact that in all five experiments there was no effect of whether the tested category was an artifact or a natural kind (biological or nonliving). It is not my claim that such kinds do not differ from one another in systematic ways of course. There is ample evidence that people often view biological kinds as possessing internal mechanisms (Gelman, 2003; Medin & Ortony, 1989), whereas the causal structure of artifacts, in comparison, is more ambiguous (e.g., although complex artifacts like cars and computers possess considerable internal mechanisms, simple artifacts like pencils and wastepaper baskets do not; cf. Bloom, 1998; Chaigneau, Barsalou, & Sloman, 2004; Malt, 1994; Malt & Johnson, 1992; 1998; Matan & Carey, 2001; Rips, 1989). But, according to the causal-based generalization view, what varies from domain to domain, indeed from category to category, is the content of the causal knowledge. What stays the same are the causal-reasoning processes that compute a generalization on the basis of that knowledge (Keil, Smith, Simons, & Leven, 1998).

9.2. Associative reasoning in causal-based generalizations

Another key finding from the present research is that a substantial minority of subjects appeared to *not* engage in the process of causal inference specified by CBG. Instead, they performed as if they treated the causal links as a simple symmetric relation. Specifically, whereas these individuals showed sensitivity to the strength of the causal relations, they were insensitive to their direction in Experiments 2 and 4 and to feature distribution information in Experiments 4 and 5. Of course, it impossible to reach strong conclusions regarding these groups due to the small numbers of individuals involved: a more thorough examination of their reasoning (requiring testing more subjects in order to yield minority groups of larger size) might have revealed a sensitivity to additional factors. But that all three subgroups appeared to generalize solely on the basis of causal strength is suggestive of the possibility that these individuals were applying a simple ''associative'' strategy in which stronger associations produced stronger generalizations.

This finding that participants can be grouped according to whether they generalize on the basis of causal reasoning or some other sort of more basic strategy has also been reported by Rehder (2006). In that study, participants either based their generalizations on a causal link or on one of the similarity-based effects described earlier (e.g., typicality or diversity) but not both. Clearly, results such as these can be indicative of the presence of alternative reasoning systems. For example, one venerable tradition in cognitive theory has been the distinction between *associative* (or *holistic*) reasoning and *analytical* reasoning. Associative reasoning is generally thought to be fast, nondeliberative, operates ''in parallel,'' is similarity based, and consumes relatively few cognitive resources, whereas the analytical system is slower, conscious, operates sequentially, is rule based, and is effortful (Sloman, 1996). Applied to both the current work and Rehder (2006), this distinction suggests that many participants may have applied the slower, more deliberate ''rules'' of causal inference in computing generalizations, whereas some invoked less effortful associative reasoning processes that are sensitive to similarity and that treat associative links as undirected. Future research will be needed to determine whether alternative generalization strategies consistently map onto the analytic/holistic distinction (e.g., by showing that use of analytic strategies becomes less prevalent under response time pressure) and what conditions promote use of one strategy versus another. Nevertheless, it important to recognize that the current experiments show that people at least have the underlying competence to base their generalizations on the relatively sophisticated processes of causal inference when they are willing to allocate the significant cognitive resources that such processes presumably require.

9.3. Causal-based generalization versus centrality

A second goal of this research was to contrast the causal-based generalization approach with the alternative centrality model. In fact, the centrality model was not supported by any of the preceding experiments. First, for Experiments 1–4 the centrality account predicted that the novel property should have been generalized more strongly when it was a cause (and had dependents) as compared to when it was an effect (and had no dependents), but this result failed to obtain in all four experiments. Second, for Experiment 3 it predicted stronger generalizations only when a novel property had three effects versus one, when in fact they were also stronger when it had three causes versus one. Third, the centrality account was unable to account for the interactions involving causal direction (in Experiments 2 and 4). Fourth, the centrality account was unable to account for the interactions involving the distribution of category features (in Experiments 4 and 5).

But despite these failed predictions, it is important to recognize that the factor that determines centrality—a property's number of dependents—indeed influences how that property is generalized. For example, Experiment 3 found stronger generalizations when a novel property had three effects as compared to one. But rather than attributing this result to centrality per se, CBG claims that participants were engaged in a form of causal reasoning that resulted in additional factors also influencing generalizations, namely, the novel property's number of cause features, the distribution of features in the category, and so forth. In other words, CBG subsumes the effect of a property's number of dependents into a more general framework in which people reason causally (either forward or backward) from existing features to novel ones in order to compute generalizations.

Although CBG thus rejects the centrality approach in favor of a more general framework, it is important to note that centrality was only one of two components of the Hadjichristidis et al. (2004) model of induction. These researchers argue that a feature's projectibility will also be influenced by whether the feature is central in the target category, which in turn depends on whether its dependency structure also appears in that category. Although reasoners will sometimes know enough about the target category to make this determination, when they do not they will use the similarity between the example and target to estimate the amount of shared dependency structure. Thus, a hormone that is responsible for a seal's physiological functions is more likely to be present in dolphins than sparrows, because the similarity between seals and dolphins suggests they have many physiological functions in common (and thus also the hormone). The dissimilarity between seals and sparrows suggest that they share many fewer physiological functions (and thus not the hormone). In fact, Hadjichristidis et al. found that although properties with ''many'' dependents were generally projected more strongly than those with a ''few,'' the magnitude of this effect decreased as the target category became less similar to the example.

This finding suggests an important way in which in causal reasoning and similarity can work together during an inferential task. In the present experiments, participants were instructed on causal knowledge that was very specific and detailed (as compared to the more nebulous ''many physiological functions''), a situation likely to minimize the use of this sort of similarity-based information. But when reasoners possess knowledge that is vague or incomplete (as they often do, Rozenblit & Keil, 2002), similarity may be used as a stand-in for that knowledge. Of course, according to CBG similarity may get invoked in this manner when a reasoner's vague and incomplete knowledge pertains to either the novel feature's effects *or* its causes. For example, if a behavior of a biological kind has ''many'' versus ''few'' (unspecified) causes, its probability of being generalized to a second kind will likely be a function of its similarity to the first, because those causes are more likely to be present in the target when the two kinds are similar.

9.4. Causal-based generalization versus sophisticated similarity

Although this article claims that properties are often generalized on the basis of causal reasoning, it is important to consider whether the results could be explained by a more sophisticated version of similarity. For example, one component of Medin et al.'s (2003) *relevance theory* is that similarity emerges from a process of comparing the premise and conclusion categories (also see Blok, Medin, & Osherson, 2007; Blok, Osherson, & Medin, 2007). For example, comparing skunks and zebras would yield stripedness as a highly relevant feature, whereas comparing skunks and onions would yield strong odor (or, equivalently, stripedness becomes available in the context of skunks and zebras, whereas strong odor is available in the context of skunks and onions; Shafto, Coley, & Vitkin, 2007). However, because according to these accounts similarity does not change as a function of the property involved, they are unable to account for the Heit and Rubinstein (1994) results in which the strength of generalization (between, say, bears and whales as compared to tunas and whales) varied as a function of whether the property was anatomical or behavioral.

Heit and Rubinstein themselves proposed that their results could be explained in terms of similarity, but that similarity was computed flexibly, that is, in a manner that is influenced by the to-be-generalized property. Indeed, they showed that generalization judgments were well predicted by the rated similarity of the base and target categories when those ratings were made ''with respect to'' anatomy or behavior, respectively. Heit and Rubinstein suggested that anatomy and behavior may be two of a small number of fixed similarity measures or that people might use the theoretical knowledge invoked by the to-be-generalized property to dynamically select which dimensions along which similarity between base and target should be computed. The latter account can also explain some of the results from Rehder (2006) that showed that the presence of a causal explanation linking the to-be-generalized property to an existing feature in the base resulted in reasoners being very sensitive to whether that feature was present in the target and insensitive to whether the base and target were similar on other dimensions.

Nevertheless, these more sophisticated views of similarity fail to explain those results that are uniquely characteristic of causal inference. For example, a symmetric relation like similarity is unable to account for the causal asymmetries manifested by the interactions involving causal direction in Experiments 2 and 4. It is also unable to account for the diagnostic reasoning that apparently occurred in Experiment 3 and the Hadjichristidis et al. (2004) study in which participants were asked to generalize a fictitious hormone from one category to another (e.g., from seals to dolphins). Merely adopting an anatomical notion of similarity fails to explain why the hormone was projected more strongly from seals to dolphins when it was responsible for many versus few physiological functions, because in both cases the base category (seals), the target category (dolphins), and the property (the hormone) are the same, and thus so too is (anatomical) similarity. Finally, similarity has no means to account for the effects of a correlated versus uncorrelated category structure documented in Experiments 4 and 5. Clearly, to explain these results one must also make reference to the causal network in which the hormone is involved, and the view I promote is that generalizers reason both forward (prospectively) and backward (diagnostically) to the presence of a to-be-generalized property in category members.

9.5. Other forms of causal-based generalization

The preceding experiments have addressed how the causal relations involving a novel property influence the generalization of that property. But what of the causal relations not involving the property itself, such as those between the category's existing features? For example, we know that birds build nests in trees because they can fly and fly because they have wings. Research has shown that the *coherence* a category gains by causal relations between existing features affects the degree to which it supports the generalizations. For example, Gelman (1988) found that second graders but not preschoolers were more likely to generalize new properties to natural kinds rather than to artifacts. She attributed this result to the coherence provided by the folk theories about biological kinds possessed by the older children which led them to expect such kinds to be more homogenous (and this expectation of homogeneity extended to new properties, resulting in stronger generalizations; also see Coley, Medin, & Atran, 1997; Gelman & O'Reilly, 1988; Lopez et al., 1997; Shipley, 1993). Consistent with this view, Patalano and Ross (2007) found that properties were generalized more strongly to categories that were rated as coherent on a variety of measures (e.g., skydivers) as compared to less coherent categories (e.g., joggers; Patalano et al., 2006). Finally, Rehder and Hastie (2004) manipulated coherence by introducing causal relations between existing category features and found that coherent category members supported stronger generalizations of blank properties.

Generalizations are also affected by relations between categories. Research mentioned earlier documented the influence of one type of inter-category causal link—ingestion—on generalizations (e.g., a property displayed by both cats and sparrows will be attributed to the fact that cats eat sparrows). Indeed, in addition to comparison-based similarity, a second component of Medin et al.'s (2003) relevance theory is the assumption that causal relations between premise and conclusion categories will often form the basis of induction. To this end, Tenenbaum, Griffiths, and Kemp (2006) have proposed a Bayesian model that computes the probability with which a property is transmitted through a ''food web'' (a network of ingestion relations) from an example to a target (also see Kemp & Tenenbaum, 2009; Shafto, Kemp, Bohawitz, Coley, & Tenebaum, 2008). More work is needed, because of course diseases can spread in ways other than ingestion (e.g., physical proximity, Proffitt et al., 2000). And generalizations might depend on yet other types of inter-category relations (properties might be shared between substances based on their common origin, e.g., between sand and glass; Medin et al., 2003).

9.6. Summary

This article has introduced a model of how generalizations are made on the basis of the causal laws that relate a to-be-generalized property with existing category features. The results of five experiments provided support for the causal-based generalization view and were inconsistent with both the centrality view and more sophisticated versions of similarity purported to account for some cases of theory-based generalizations. There was, however, evidence that a substantial minority of reasoners engaged in a simpler form of associative reasoning rather causal inference when computing generalizations.

Acknowledgments

I thank Brett Hayes, ShinWoo Kim, and three anonymous reviewers for their comments on earlier drafts of this article. I also thank Samanta Shaw for her tireless efforts at running experimental sessions. Preliminary results of Experiment 2 were submitted to New York University as part of an unpublished Honors Thesis conducted by Adam Nabatian. Experiment 1's follow-up experiment was based on an Honors Thesis conducted by Maya Nair.

Notes

1. The presence of subgroups in the results was confirmed by a cluster analysis (using SAS PROC CLUSTER METHOD = EML). For each subject, a variable representing the strength of the interaction was computed by subtracting the difference in ratings in the strong (m = 1) versus weak (m = .67) conditions when N was an effect from that difference when N was a cause. The analysis revealed that log-likelihood was minimized assuming three clusters with means of 45.8 (n = 6), 27.7 (n = 16), and −1.8 (n = 12). That is, whereas the first two clusters exhibited the predicted interaction (Fig. 3B), the third did not (Fig. 3C). Because the first two clusters exhibited a qualitatively identical pattern of responding, they are collapsed in Fig. 3B for simplicity of exposition.
2. There are other ways that a novel feature can be related to multiple category features of course. For example, the online supplement available at http://www.cogsci.rpi.edu/CSJarchive/Supplemental/index.html presents CBG's predictions for the case where N is a cause of one category feature and an effect of another.
3. The presence of subgroups was confirmed by submitting the difference in ratings in the uncorrelated condition when N was a cause versus an effect to a cluster analysis. The analysis revealed that log-likelihood was minimized with two clusters with means of −.59 (n = 7) and .02 (n = 5). That is, whereas the first cluster exhibited a much larger difference in ratings when the N was an effect versus a cause (Fig. 8B), there was no such difference in the second cluster (Fig. 8C).
4. The presence of subgroups was confirmed by submitting the difference in ratings when N was a cause versus an effect to a cluster analysis in both the correlated and uncorrelated conditions. In the correlated condition, log-likelihood was minimized with two clusters with means of .07 (n = 19) and .48 (n = 5). In the uncorrelated condition, it was minimized with three clusters with means of −.44 (n = 5), −.08 (n = 11), and .39 (n = 8). Because the first two clusters exhibited a qualitatively identical pattern of responding, they are collapsed in Fig. 10B for simplicity.

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.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experiment Psychology: Learning, Memory, and Cognition*, *11*, 629–654.

Blok, S. V., Medin, D. L., & Osherson, D. N. (2007). Induction as conditional probability judgement. *Memory & Cognition*, *35*, 1353–1364.

Blok, S., Osherson, D., & Medin, D. (2007). From similarity to chance. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches* (pp. 137–166). New York: Cambridge University Press.

Bloom, P. (1998). Theories of artifact categorization. *Cognition*, *66*, 87–93.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1119–1140.

Chaigneau, S. E., Barsalou, L. W., & Sloman, S. A. (2004). Assessing the causal structure of function. *Journal of Experimental Psychology. General*, *133*, 601–625.

Cheng, P. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–405.

Cheng, P. W., & Novick, L. R. (2005). Constraints and nonconstraints in causal learning: Reply to White (2005) and to Luhmann and Ahn (2005). *Psychology Review*, *112*, 694–707.

Coley, J. D., Medin, D. L., & Atran, S. (1997). Does rank have its privilege? Inductive inference within folkbiological taxonomies. *Cognition*, *64*, 73–112.

Fischoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 330–344.

Gelman, S. A. (1988). The development of induction within natural kinds and artifact categories. *Cognitive Psychology*, *20*, 65–95.

Gelman, S. A. (2003). *The essential child: The 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.

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

Glymour, C. (1998). Learning causes: Psychological explanations of causal explanation. *Minds and Machines*, *8*, 39–60.

Glymour, C., & Cheng, P. W. (1998). Causal mechanism and probability: A normative approach. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 296–313). Oxford, England: Oxford University Press.

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, and Cognition*, *20*, 411–422.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Keil, F. C., Smith, W. C., Simons, D. J., & Leven, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition*, *65*, 103–135.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*, 20–58.

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 754–770.

Lopez, 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.

Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, *28*, 41–50.

Malt, B. C. (1994). Water is not H_2O. *Cognitive Psychology*, *27*, 41–70.

Malt, B. C., & Johnson, E. C. (1992). Do artifacts have cores? *Journal of Memory and Language*, *31*, 195–217.

Malt, B. C., & Johnson, E. C. (1998). Artifact category membership and the intentional-historical theory. *Cognition*, *66*, 79–85.

Matan, A., & Carey, S. (2001). Developmental changes within the core of artifact concepts. *Cognition*, *78*, 1–26.

Mayrhofer, R., Goodman, N. D., Waldmann, M. R., & Tenenbaum, J. B. (2008). Structured correlation from the causal background. In V. Sloutsky, B. Love & K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 303–308). Mahwah, NJ: Erlbaum.

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, MA: Cambridge University Press.

Nair, M. (2005). *The role of causal reasoning in category-based generalizations*. Unpublished Honors Thesis. New York: New York University.

Osherson, D., Smith, E. E., Myers, T. S., & Stob, M. (1994). Extrapolating human probability judgment. *Theory and Decision*, *36*, 103–126.

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

Patalano, A., Chin-Parker, S., & Ross, B. H. (2006). The importance of being coherent: Category coherence, cross-classification, and reasoning. *Journal of Memory & Language*, *54*, 407–424.

Patalano, A. L., & Ross, B. H. (2007). The role of category coherence in experience-based prediction. *Psychonomic Bulletin & Review*, *14*, 629–634.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufman.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, England: Cambridge University Press.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and 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, and Cognition*, *29*, 1141–1159.

Rehder, B. (2006). When similarity and causality compete in category-based property induction. *Memory & Cognition*, *34*, 3–16.

Rehder, B. (2007). Property generalization as causal reasoning. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches* (pp. 81–113). New York: Cambridge University Press.

Rehder, B., & Burnett, R. (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.

Rehder, B., & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 659–683.

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

Rips, L. J. (2001). Necessity and natural categories. *Psychological Bulletin*, *127*, 827–852.

Rosch, E. H., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Rozenblit, L. R., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521–562.

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, and Cognition*, *29*, 641–649.

Shafto, P., Coley, J. D., & Vitkin, A. (2007). Availability in category-based induction. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches* (pp. 114–136). New York: Cambridge University Press.

Shafto, P., Kemp, C., Bonawitz, E. B., Coley, J. D., & Tenebaum, J. B. (2008). Inductive reasoning about causally transmitted properties. *Cognition*, *109*, 175–192.

Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 265–301). New York: Academic Press.

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. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–23.
Sloman, S. A. (1997). Explanatory coherence and the induction of properties. *Thinking and Reasoning*, *3*, 81–110.
Sloman, S. A., Love, B. C., & Ahn, W. (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.
Springer, K., & Keil, F. C. (1989). On the development of biologically specific beliefs: The case of inheritance. *Child Development*, *60*, 637–648.
Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*, 309–318.
Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *10*, 547–567.
Wu, M. L., & Gentner, D. (1998). Structure in category-based induction. Paper presented at the 20th Annual Conference of the Cognitive Science Society, Madison, WI.