Property Generalization as Causal Reasoning

Bob Rehder
Department of Psychology
New York University
New York NY 10003

Send all correspondence to:
Bob Rehder
Dept. of Psychology
6 Washington Place
New York, New York 10003
Phone: (212) 992-9586
Email: bob.rehder@nyu.edu
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 its safe to pet; given a particular berry, one asks if its safe to eat (e.g., see Murphy & Ross, this volume). 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 venom. This chapter is concerned with category-based generalizations such as these in which new properties are projected to an entire class of objects.

Let’s consider a couple of examples intended to illustrate two different models of the reasoning processes which are supposed to underlie category-based generalizations. Suppose you are introduced to someone from Uzbekistan, the first person from that country you have ever met. During your otherwise pleasant conversation you realize that this person has bad breath, and are led to wonder whether bad breath is common among Uzbekistanis. If you knew nothing about Uzbekistan, you’d probably be unlikely to jump to the conclusion that bad breath was a characteristic national trait. But suppose that you actually know a little bit more about Uzbekistan. For example, imagine that you know how Uzbekistanis traditionally dress, some characteristic facial features, common religious beliefs, and so on. According to one prominent theory, you will be more likely to think that bad breath is frequent among Uzbekistanis if your new friend has many of these typical Uzbekistani traits, as compared to if he is, say, wearing a Western-style business suit, endorses a traditional western religion (e.g., Roman Catholicism), and so on.
Do you find this example compelling? Perhaps my intuitions have been ruined by reading too many papers about induction, but I can’t say that I do. I might be slightly more willing to endorse the generalization about bad breath on the basis of typical or characteristic features, but not by much, and I think the reason has to do with the second model of category-based generalization illustrated by the next example. You’re chatting with the Uzbekistani, but now, instead of knowing a few typical properties of people from that country, you know a few of the causes of bad breath. Let’s say you know that eating salted fish causes bad breath, and so too does drinking vodka. You then learn from the Uzbekistani that salted fish is one of the staples of the Uzbekistani diet and that vodka is consumed at every meal. Now suddenly the idea that large numbers of Uzbekistanis frequently have bad breath doesn’t seem so far fetched after all.

The first example suggesting that generalizations are based on an example’s typicality is the province of the well-known Similarity-Coverage Model of category-based induction (D. M. Osherson, Smith, Wilkie, Lopez, & Shafir, 1990) discussed in numerous places throughout this volume (also see Sloman, 1993, for a related model). For example, according to the Similarity-Coverage Model (hereafter SCM), people will be more confident that all birds have some new property (e.g., sesamoid bones) when given an example of a sparrow with sesamoid bones as compared to a penguin, because sparrows are more typical birds than penguins (and thus, on this account, provide more “coverage” of the features of the bird category). Likewise, the model would predict that you would generalize bad breath to Uzbekistanis more strongly if your one example was typical and thus covers the "Uzbekistani" category more completely. In addition to coverage, the second component of the SCM—similarity—allows it to account for generalizations between categories that are not hierarchically nested. For example, people will be more confident that blackbirds have sesamoid bones given the fact that crows do as compared to sparrows, because (according to the SCM) blackbirds are more similar to crows than sparrows.

In contrast, the second mode of reasoning, referred to here as generalization as causal
Property Generalization as Causal Reasoning

Reasoning (hereafter GCR), focuses on the specific causal explanations which 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 which produce or generate particular properties, and this knowledge can be used to estimate the prevalence of that property in some population (i.e., category of object). This view also explains why, despite the predictions of the SCM, the influence of typicality may be uncompelling in certain cases. For example, I usually attribute bad breath to a transient property—something consumed at a recent meal for instance—and this makes me reluctant to consider bad breath a stable property of even that individual person, to say nothing of the person’s nationality, ethnicity, or any other group he or she might belong to. It also leads me to believe that whether or not the person is displaying a typical property on some other dimension (their clothing style) is irrelevant to the question of generalizing bad breath. Of course, to be fair to the SCM, it was never intended to be a model of the mental processes that arise when one has knowledge of the causal mechanisms which might generate the to-be-generalized property. Famously, its predictions were intended to apply only to “blank properties” (like sesamoid bones) about which one has no prior knowledge. Nevertheless, the example illustrates how easy it may be for its similarity-based processes to be supplanted by explanation-based ones.

Moreover, it may often be the case that even when more typical examples do support stronger generalization, they do so not because of typicality per se, but rather because of the distinct pattern of reasoning that a typical example may support. For example, if you, unlike me, did think that a Uzbekistani dressed in traditional clothing justified a stronger generalization, I suspect you did so because you recognized that (a) bad breath can be attributed to a stable property such as the person’s diet, (b) because diets and ways of dress covary, this diet (whatever it is) may be common to many Uzbekistanis, and therefore (c) so too is bad breath. But although this makes you a deeper reasoner than me, you are reasoning nonetheless, not just basing your judgment on some unanalyzed notion of typicality.
This chapter is divided into three sections. The first will present the current empirical evidence that people in fact engage in causal reasoning when generalizing properties. This evidence will come from both studies testing the effects of "background" causal knowledge—knowledge which subjects possess before the experiment starts—and those which instruct subjects on causal knowledge as part of the experiment. The assumption is that both sorts of studies are useful. Establishing results with background knowledge helps provide some assurance that the reasoning processes involved are likely to be evoked in real-world contexts. Experimental materials, in turn, allow the controls necessary to distinguish specific hypotheses from one another. Two specific models of how causal knowledge influences category-based generalizations will be discussed.

The second section will address the relationship between causal reasoning and the similarity-based effects on property generalization. The claim here is not that similarity never influences generalizations, or that similarity-based processes are mischaracterized by the SCM, but rather that they and causal reasoning processes coexist side-by-side in the mental toolbox of processes which support induction. But when similarity and causality can each form the basis of a generalization, do people weigh (equally or unequally) both sources of information, or does one mode of reasoning override or suppress the other? I will review evidence suggesting that when causal information is available, it largely eliminates similarity-based influences.

This first two sections addresses the influence of the causal relations which directly relate existing category features and a to-be-generalized novel property. But what of the influence of causal relations not involving the novel property which relate features of the category to one another? The final section will discuss how such relations change the coherence of category, and by so doing promote the strength with which novel properties are generalized.

Property Generalization As Causal Reasoning

There are now numerous sources of evidence that people reason causally in order to
generalize properties. I first discuss the studies involving background causal knowledge, but, because recent reviews of this research already exist (Heit, 2000; Rips, 2001), only a few example will be covered here. I will then turn to more recent research testing causal relations which are provided as part of the experimental session.

Research With Background Causal Knowledge

One of the first studies suggesting that property generalization can depend on causal knowledge was conducted by Heit and Rubinstein (1994). Recall that, according to the SCM, a novel property will be generalized from one category to another to the extent that the categories are similar. In particular, the SCM does not allow for how such generalizations might depend on the property itself, by virtue of, say, the specific causal mechanisms it is involved in. Heit and Rubinstein showed otherwise. They found, for example, that a behavioral property (e.g., travels in a zig-zag path) was generalized more strongly from tunas to whales as compared to from bears to whales. This result may have been expected on the basis of the fact that whales are more similar to tunas than bears. But when the novel property was biological rather than behavioral (e.g., a liver with two chambers that acts as one), it was generalized more strongly from bears to whales instead. 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, that is, mammals. Tunas and whales, on the other hand, are more likely to share a survival behavior (traveling in a zig-zag path) because they are both prey animals living in a common ecology (also see Gelman & Markman, 1986, and Springer & Keil, 1989).

Sloman (1994, 1997) has provided other examples of how a property will be more strongly generalized when the causal history which leads to a property in a base category is also present in the target. For example, he found that undergraduates were more willing to project a feature like “hired as bodyguard” from war veterans to ex-convicts, presumably
because the underlying explanation (fighting experience makes for a good bodyguard) for war veterans also applies to ex-convicts. In contrast, they were less to willing to project the property “unemployed,” apparently because the reasons for unemployment of war veterans (physical injury, PTSD) does not apply to ex-convicts. Note again that these different result obtained despite the fact that similarity was held constant (in both cases war veterans were the base and ex-convicts were the target).

Finally, Smith, Shafir, and Osherson (1993) also found that subjects apparently 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 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).

Thus, there appears to be good reason to suspect that causal reasoning is intimately involved in the generalization of properties. But although suggestive, studies testing background knowledge are often correlational in nature, and thus there is always the possibility that the results arise due to factors other than causal knowledge itself. For this reason, some researchers have adopted a purely experimental approach in which subjects are presented with category members that display novel properties which are accompanied by a description of the (previously unknown) causal factors that led to their presence. The precision with which this approach allows causal knowledge to be manipulated has led to the first specific theories regarding how such knowledge affects property generalization, as I now describe.

Research With Experimentally-Provided Causal Knowledge

The theories which have been advanced thus far have been concerned with how a to-
be-generalized property is related to a category's existing features. The first of these appeals to the notion of feature centrality (Hadjichristidis, Sloman, Stevenson, & Over, 2004). Features are more central to the extent they have many features which "depend on" them, or, equivalently, to the extent they are responsible for generating or producing many other features (Sloman, Love, & Ahn, 1998). One purported role for centrality is in determining a feature's importance to an object’s category membership and its degree of typicality within a category. For example, a bird with all the characteristic features of robins except that it has unusual coloring is a more typical robin (and more likely to be a robin), than one with all the characteristic features of robin except for robin DNA. According to the centrality account, robin coloring is less central than robin DNA (and thus less important for establishing category membership) because coloring depends on DNA and not the other way around (Ahn, 1998; Ahn et al., 2000; see Rehder, 2003ab; Rehder & Kim, in press for an alternative view).

Hadjichristidis et al. (2004) have proposed that the concept of centrality is also important to induction, suggesting that novel properties which are central are more likely to be generalized from one category to another. For example, suppose one person believes that many of a seal’s physiological functions depend on a particular hormone, whereas another believes that only a few such functions depend on it. According to the centrality account, the hormone will be more central for the first person than the second. Hadjichristidis et al. suggest that the hormone will as a result not only be more important for identifying an animal as a seal for the first person, but he or she will be more likely to generalize it to another category (say, dolphins). But they point out that the strength of generalization will also depend on whether the hormone is viewed as being central in the target category, which in turn depends on whether the hormone's "dependency structure" (e.g., the physiological functions) also appears in the target category. People will sometimes know enough about the target category to make this determination directly. But when they don't, Hadjichristidis et al. suggest that reasoners use the similarity between the base and target to estimate the amount of shared dependency structure. Thus, a hormone which 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). In contrast, the dissimilarity between seals and sparrows suggest that they share many fewer physiological functions (and thus not the hormone).

Hadjichristidis et al. (2004, Experiment 2), tested this hypothesis by varying a novel property’s centrality (e.g., a fictitious hormone was depended on by either many physiological functions, few such functions, or no information about dependency was provided), and the similarity of the base and target category (high, medium, or low). As predicted, central properties (with many dependents) were projected more strongly than less central properties (with few dependents), which in turn were stronger than properties with no dependents (Fig. 1). Also as predicted, Figure 1 also shows how the strength of this effect was moderated by the similarity of the two categories: Centrality had its greatest effect when the target category was very similar to the base category.

These findings provide one important illustration of how interfeature causal relations influences property generalization. However, one can ask if characterizing these results in terms of feature centrality alone artificially limits the scope of the effect of causal knowledge on such generalizations. This leads to the second theory of property generalization which treats such generalizations as an instance of causal reasoning (Rehder, in press). According to this view, one computes whether a novel feature is likely to appear in a target category not on the basis of any one characteristic of the feature (e.g., its centrality), but rather on the basis of one’s beliefs about the causal laws which relate the feature to those of the target category.

The generalization as causal reasoning view (GCR) makes (at least) three predictions regarding how causal knowledge supports category-based generalizations. The first of these is that property generalization can be an instance of diagnostic reasoning in which one reasons from the presence of a novel property’s effects to the presence of the novel property itself. The second prediction is that generalizations can reflect prospective reasoning, in which one reasons from the presence of the causes of a novel property to infer the presence of the
property. The third prediction is that generalizations should exhibit the basic property of *extensional reasoning* in which a novel property will be more prevalent among category members to the extent its causes and/or effects are prevalent. As it turns out, each of these predictions has received empirical support.

_Evidence for diagnostic reasoning._ According to GCR, a property will be generalized to the extent that its presence can be inferred on the basis of the effects it is likely to produce. That is, generalization can be seen as that species of causal reasoning in which one "diagnoses" the presence of a property within category members on the basis of whether its "symptoms" (effects) are present in those objects.

The results from Hadjichristidis et al. (2004) just presented can be interpreted as just such a case of diagnostic reasoning. The fact that physiological functions "depend on" a hormone can be understood as the hormone causing or enabling those functions. Thus, the presence of one or more of the same physiological functions in another species "diagnoses" the presence of the hormone in that species. However, one complicating factor is the possibility that in the target species those physiological functions have different causes (e.g., different hormones). For this reason, the inference to the hormone is more certain when multiple of its "symptoms" (physiological functions) are present. This situation is more likely to be true to the extent that the base and target categories are similar rather than dissimilar (Figure 1).

_Evidence for prospective reasoning._ GCR not only reinterprets evidence in favor of the feature centrality view as a case of causal reasoning, it also makes new predictions which distinguish it from the feature centrality view. The second prediction of GCR is that a property will be generalized to the extent that the cause or causes which produce it are present in the target category. On this view, generalization can be seen an instance of prospective reasoning in which one reasons from causes to effects rather than from effects to causes.

To start, note that each of the demonstrations of the effect of background knowledge presented earlier implicate the role of the causal mechanisms which produce a novel property,
such as when bears and whales are likely to possess the same kind of liver on the basis of the shared biological mechanisms of mammals (which presumably generate their internal organs). The prospective reasoning view also receives support from studies which manipulate causal knowledge experimentally. For example, Rehder (in press) instructed undergraduates on an artificial, but plausible, category, such as Romanian Rogos (a type of automobile). Rogos were described as having four characteristic features, each of which was described as occurring in 75% of category members: butane-laden fuel, loose fuel-filter gasket, hot engine temperature, and large amount of carbon monoxide in the exhaust. A single example of a Rogo was then presented with a novel property, and subjects were asked whether a second Rogo was likely to have that property. On some trials no causal explanation of the novel property was provided. For example, subjects would be presented with a Rogo that had ceramic shocks, and then judged what proportion of all Rogos had ceramic shocks. Properties like ceramic shocks were chosen because they have no obvious causal connection with the other known properties of Rogos (e.g., butane-laden fuel, carbon monoxide in the exhaust, etc). But on other trials the novel property was described as being caused by one of the Rogo's characteristic features. For example, subjects would be presented with a Rogo which had a typical property (e.g., hot engine temperature) and a novel property (e.g., melted wiring), and told that the typical property caused the novel one ("The melted wiring is caused by the high-engine temperature."). A second Rogo was presented which either did or did not have high-engine temperature, and subjects were asked whether it had melted wiring.

The results, presented in Figure 2, shows that the novel property was generalized very strongly when its cause was present in the target Rogo, very weakly when it was absent, and with an intermediate rating when no information about causal mechanism was provided. The results were the same regardless of whether the category was an artifact (e.g., Rogos), a biological kind, or a nonliving natural kind. Clearly, subjects reason forward (prospectively) to infer a novel property's presence from the presence of its causes as readily as they reason backward (diagnostically) to infer a property's presence from its effects (also see Lassaline,
Evidence for extensional reasoning. A third prediction of GCR is that generalizations should exhibit extensional reasoning in which the prevalence of a novel property depends on the prevalence of its causes and/or effects. For example, the prevalence of a hormone in a target category shouldn’t just depend on the number of physiological functions it causes, but also how widespread those functions are amongst the population of target category members. Similarly, the prevalence of melted wiring caused by hot engines in a brand of automobile should depend on what proportion of those automobiles in fact have hot engines.

Maya Nair and I decided to test this prediction by explicitly manipulating the prevalence of a category feature which was the purported cause (or effect) of a novel property (Nair, 2005). The same categories as in Rehder (in press) were used (including Romanian Rogos), subjects were asked what proportion of all Rogos had a novel property on the basis of a single example of a Rogo with that property. Two factors were manipulated as within-subjects variables. The first was the base rate of the characteristic feature. Two randomly chosen features of Rogos were described as occurring in 90% of Rogos, whereas the other two were described as occurring in 60%. The second orthogonal factor was whether the novel property was described as caused by or as the cause of one of the characteristic features. The characteristic features, the novel properties, and the causal relationships between the two are presented in Table 1 for Rogos. For example, some subjects would be presented with a Rogo which had all four characteristic features and a novel property (e.g., zinc-lined gas tank) and told that one of the typical properties caused the novel one ("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 subjects would be told that the novel property was the cause of the typical property ("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 subjects would then rate what proportion of all Rogos possessed the novel property. Each subject performed four generalization trials in which the
four novel properties in Table 1 were each presented once, either as a cause and effect of a characteristic feature, and with the base rate of the characteristic feature (e.g., butane-laden fuel) described as either 90% or 60%.

Generalization ratings from this experiment are presented in Figure 3A as a function of whether the novel property was the cause or the effect and the base rate of the typical property. There was a strong effect of base rate, as subjects' generalization ratings were much higher when the novel property was causally related to a characteristic feature with a 90% versus a 60% base rate. Moreover, this result obtained regardless of whether the characteristic feature was the cause or effect of the typical property. In other words, subjects readily engage in the extensional reasoning which characterizes causal reasoning more generally.

In a follow-up experiment, we asked whether these results would obtain when the feature base rates were observed rather than given through explicit instruction. Informal post-experiment interviews in the first experiment revealed that the base rate manipulation was very powerful. Although generalization ratings were entered by positioning a slider on a computer-displayed scale which was not labeled numerically (one end was simply labeled "None," meaning that no Rogos have the novel property; the other end was labeled "All" meaning they all do), a number of subjects reported trying to position the marker at a point that they felt corresponded to 60% or 90%. That is, if a novel property was causally related to a typical property with a base rate of 90% (or 60%), many subjects felt that that property would be displayed by exactly 90% (or 60%) of Rogos. To determine if the extensional reasoning effect only obtained because the base rates were explicit, we performed a replication of the first experiment in which the feature base rates were learned implicitly through a standard classification-with-feedback task. In the first phase of the experiment subjects were asked to learn to distinguish Romanian Rogos from “some other kind of automobile.” They were presented with the exemplars shown in Table 2, consisting of eight Rogos and eight non-Rogos. In Table 2, a “1” stands for a typical Rogo feature (e.g., loose fuel filter gasket, hot engine temperature, etc.) whereas each “0” stands for the opposite value
on the same dimension (e.g., tight fuel filter gasket, cool engine temperature, etc.). The thing to note is that on two of the dimensions (Dimensions 1 and 2 in Table 1) the typical Rogo feature is very typical, because it occurs in 7 out of 8 Rogos (or 87.5%). On the other two dimensions (3 and 4), the characteristic feature is less typical, occurring in 5 out of 8 Rogos (or 62.5%). The structure of the contrast category (the non-Rogos) is the mirror image of the Rogos.

The items in Table 2 were presented in a random order in blocks, subjects classified each and received immediate feedback, and training continued until subjects performed two blocks without error or reached a maximum of 20 blocks. They then performed the same generalization task as in the previous experiment in which whether the novel property was the cause or effect was crossed with the base rate of the typical Rogo feature (87.5% or 62.5%). The results of this experiment are presented in Figure 3B. The figure indicates that subjects in this experiment, just like those in the first one, took the base rate of the characteristic feature into account when generalizing the new property. They did this even though the difference in base rates between features was manifested in observed category members rather than provided explicitly. That is, people take into account the prevalence of that characteristic feature in the population of category members when generalizing the novel property, and do so regardless of whether a novel property is the effect or cause of a characteristic feature. Note that in both experiments from Nair (2005) the results were unaffected by whether the category was a novel artifact, biological kind, or nonliving natural kind.

In summary, studies provide strong evidence that people are exhibiting some of the basic properties of causal reasoning when generalizing properties. When many of a novel property's effects are present in a category, people reason (diagnostically) to the presence of the property itself. People also reason (prospectively) to the presence of the property as a function of whether its causes are present or absent in the target. Finally, people are sensitive to how prevalent a novel property's cause (or effect) is among the population of category members.
When Causality and Similarity Compete in Property Generalization

The preceding research firmly establishes the influence of causal explanations on how properties are generalized to categories. Another important question concerns how those explanations interact with the influences of similarity which have been formalized by the Similarity-Coverage model (SCM). Do each of these sources of evidence contribute (perhaps unevenly) to the generalization of a property, or do people choose to generalize on the basis one source of evidence alone? As it turns out, research suggests that causal explanations not only influence generalizations, but they may also largely undermine, or supplant, similarity-based processes. Once again, I will first briefly review studies testing background knowledge, and then turn to the results from studies using experimentally-provided causal knowledge.

Research With Background Causal Knowledge

To investigate the role of causal knowledge in property generalization, a number of studies have compared how inductive reasoning varies across populations which differ in the nature and amount of background knowledge. For example, Lopez, Atran, Coley, Medin, and Smith (1997) studied category-based generalizations among the Itza’ Maya (an indigenous population in central Guatemala), and found that the Itza’ failed to exhibit standard diversity effects. Diversity refers to the fact that a more diverse set of base examples leads to stronger generalizations. All else being equal, people are more likely to conclude that all birds have sesamoid bones given that a dissimilar set of birds do (e.g., sparrows, hawks, and chickens) as compared to if the birds are more similar (e.g., sparrows, robins, and bluejays). Diversity effects are predicted by the SCM because less similar base categories provide better coverage of the target category. But Lopez et al. found that the Itza’ failed to exhibit diversity effects, as they frequently appealed to specific causal mechanisms instead (e.g., judging that the prevalence of a disease in a species depended on the mechanisms which might spread that disease). In contrast, American undergraduates exhibited standard diversity effects on the same items (also see Bailenson, Shum, Atran, Medin, & Coley, 2002).

Of course, these findings could be attributed to some cultural difference between
Property Generalization as Causal Reasoning

Americans and the Itza' other than domain knowledge (such as a difference in their default reasoning strategy). Contra this interpretation however, the dominance of causal explanations also holds for Americans who possess sufficient domain knowledge. Proffitt, Coley, and Medin (2000) tested three groups of North American tree experts (taxonomists, landscapers, and maintenance workers) to determine how they generalized a novel property of trees (an unspecified disease) as a function of the typicality and diversity of base categories (species of trees). Not only did two of the three groups of experts (the landscapers and maintenance workers) fail to exhibit diversity effects (just as the Itza’ failed to), none of the groups exhibited standard typicality effects. Instead, the experts frequently reasoned in a manner which reflected their causal understanding of disease processes, such as the susceptibility or resistance of certain species to diseases, the geographic distribution of a tree species (more disperse species infect more trees), and the specific mechanisms by which diseases are transmitted from one tree to another (also see Bailenson et al., 2002; Shafto & Coley, 2003).

Finally, there is ample evidence that causal reasoning in support of induction is not limited to the rarefied world of domain experts (such as the Itza’ or landscapers). Testing American undergraduates, Medin, Coley, Storms, and Hayes (2003) found, for example, that a more diverse set of premise categories consisting of a mammal and a bird (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. Similarly, the more diverse set of categories again consisting of a mammal and a bird (pigs and chickens) resulted in weaker generalization to cobras as compared to a less diverse set of two mammals (pigs and whales). It may have been that pigs and chickens were both recognized as farm animals, which suggested one or more candidate causes of the novel property (e.g., being injected with antibiotics). Inductions were then weaker because those candidate causes are absent in cobras.
These studies make a strong case that, in the presence of causal knowledge, similarity-based factors may become largely irrelevant to how people choose to generalize examples. Can we also find evidence for this conclusion when causal relations are manipulated experimentally?

*Research With Experimentally-Provided Causal Knowledge*

To answer this question Rehder (in press) conducted a series of experiments which assessed how causal explanations moderate those effects of similarity on property generalization predicted by the SCM. Each experiment used a two-factor design in which one factor was the presence or absence of a causal explanation, and the second factor was either typicality, diversity, or similarity itself. This design of these experiments allowed a demonstration of SCM’s standard similarity-based effect in the absence of causal explanations, and how that effects are moderated when such explanations are present.

*Causal explanations versus typicality.* American undergraduates were instructed on artificial categories like Romanian Rogos which possessed four characteristic features each described as occurring in 75% of category members. They were then presented with a particular Rogo that exhibited a novel property and asked to judge what proportion of all Rogos had that property. On some trials, the novel property was accompanied by a causal explanation (e.g., the Rogo was said to have melted wiring, which was caused by its hot engine temperature), whereas on other trials the novel property was “blank,” that is, no explanation was provided (e.g., ceramic shocks). The second experimental factor was the typicality of the Rogo, which had either 1, 2, 3, or 4 characteristic features. A Rogo always possessed at least the characteristic feature which was the described as the cause of a novel property (e.g., hot engine temperature in the case of melted wiring).

Based on the SCM, the prediction of course was that the generalization of blank properties would strengthen with the number of characteristic features. The critical question concerned whether typicality would also affect the generalization of novel properties when a causal explanation was present. The results are presented in Figure 4A as a function of
whether a causal explanation was provided and the typicality of the exemplar. As expected, generalization ratings for blank properties increased as the exemplar’s typicality (i.e., its number of characteristic features) increased, indicating that this experiment replicated SCM’s standard typicality effect for blank properties. However, Figure 4A indicates that this effect of typicality was reduced when a causal explanation was provided.

One important question is whether the response pattern shown in Figure 4A is manifested consistently by all participants, or whether it arose as a result of averaging of over individuals with substantially different response profiles. In fact, two subgroups of participants with qualitatively different responses were identified, shown in Figures 4B and 4C. The subgroup in Figure 4B produced higher induction ratings for the nonblank properties as compared to the blanks, that is, they were more willing to generalize a novel property when accompanied by a causal explanation. Importantly, however, whereas ratings for blanks properties were sensitive to typicality, typicality had no effect on the generalization of the nonblanks. In contrast, the second subgroup in Figure 4C did not generalize nonblanks more strongly than blanks, and both types of properties were equally sensitive to typicality. In other words, when reasoners chose to base their responses on the causal explanations (as evidenced by the nonblank’s higher ratings in Figure 4B), there was no effect of typicality; when they chose to base their responses on typicality, there was no effect of causal explanations (Figure 4C). Apparently, the use of a causal explanation versus typicality is an all-or-none matter, with reasoners using one strategy or the other, but not both.

Causal explanations versus diversity. In a second experiment, Rehder (in press) also assessed how the presence of a causal explanation for a novel property moderates the diversity effect. Diversity was manipulated by presenting two category members with five stimulus dimensions with the same novel property. The two category members exhibited either low diversity (i.e., they shared all five features) or high diversity (they shared only one feature). (The two exemplars were chosen so that both always had three features, so that their typicality was held constant across low- and high-diversity trials.) Whether the novel property
had a causal explanation was manipulated orthogonally. On the basis of the SCM, the prediction was that the generalization of blank novel properties would be stronger when displayed by more diverse pairs of exemplars. The critical issue concerned the effect of causal explanations on the diversity effect.

In contrast to the previous experiment in which all subjects exhibited a typicality effect with blank properties, only half of the subjects in this experiment exhibited a diversity effect with blank properties. This result is in keeping with the literature demonstrating the diversity heuristic’s general lack of robustness relative to typicality (Lopez, Gelman, Gutheil, & Smith, 1992; see Heit, 2000, for a review). The important result however is that when a causal explanation for a novel property was provided, any effect of diversity disappeared entirely. That is, just as in the previous experiment, the presence of a causal explanation can eliminate the effect of one of the SCM’s similarity-based heuristic (in this case, diversity). Apparently, explanations direct attention away from features which would normally contribute to how well a category is “covered” by the exemplar(s) displaying a novel property. When this occurs, the typicality or diversity of those exemplars becomes irrelevant to the generalization.

*Causal explanations versus similarity.* As mentioned, the SCM predicts that generalizations between items that are not hierarchically nested will be influenced by their similarity. To determine how causal explanations affect the influence of similarity, Rehder (in press, Experiment 3) asked participants to generalize a property from one category member to another. One experimental factor was whether the base and target exemplars shared three of four features (high similarity condition) or only one (low similarity condition). (The base and target were chosen so that they always possessed three and two characteristic features, respectively, so that their typicality was held constant over similarity conditions.) The other factor was whether a causal explanation was provided for the novel property. For blank properties, SCM’s prediction was that the generalization will be stronger for high versus low similarity targets. The critical question concerned the influence of similarity on nonblank
The results are presented in Figure 5 as a function of the similarity between the base and target exemplar, and whether the novel property was blank or nonblank (and, if nonblank, whether the cause feature was present or absent in the target). As predicted by the SCM, blank properties were more strongly projected when the target exemplar was more similar to the base exemplar. Also as expected, nonblank properties were more strongly projected when the cause of that property appeared in the target exemplar versus when it didn’t. The important finding is that the projection of nonblank properties was much less sensitive to the similarity of the base and target exemplar. (Unlike the previous two experiments, the group-level pattern of responding, shown in Figure 5, was exhibited by virtually all participants). Once again, it appears that the effect of causal explanations is to draw attention away from features not involved in the explanation, with the result that the two exemplars’ similarity becomes largely irrelevant to how strongly a novel property is generalized from one to the other.

At first glance it may seem that these results conflict with those from Hadjichristides et al. (2004) which showed that novel properties were projected more strongly as a function of the similarity between the base and target (Figure 1). Recall however that Hadjichristides et al.’s claim is that similarity will be used to estimate dependency structure when its presence in the target category is otherwise uncertain. This situation which was manifested in their experiments by referring to a hormone’s dependents as nameless “physiological functions” making it impossible to determine whether those specific functions were present in the target category. In contrast, in the current experiment subjects were told exactly which category features were causally related to the novel feature, a situation which apparently invited them to largely ignore how the base and target might be similar on other dimensions.

In summary then, these three experiments support the claim that when a causal explanation for a novel property is available, it often supplants similarity as the basis for the generalization of that property. First, in the first experiment all participants exhibited an effect of typicality for blank properties, but half of those participants showed no sensitivity to
typicality when the novel property was accompanied with a causal explanation (nonblanks).

Second, in the second experiment half the participants exhibited sensitivity to diversity for blank properties, but that sensitivity was completely eliminated for nonblanks for all participants. Finally, in the third experiment virtually all participants exhibited sensitivity to similarity for blanks, but this effect was (almost) completely eliminated for nonblanks. Apparently, when people note the presence of a causal explanation for a novel property, it often draws attention away from the exemplars’ other features, making their similarity (or typicality or diversity) largely irrelevant to the inductive judgment.

Property Generalization and Category Coherence

The preceding studies have addressed how the causal relations in which a novel property is directly involved influence the generalization of that property. But what of causal relations not involving the property itself, such as those between the category’s existing features? For example, we not only know that birds have wings, fly, and build nests in trees, but also that they build nests in trees because they can fly and fly because they have wings. We not only know that automobiles have gas, spark plugs, and produce carbon monoxide, but also that gas and spark plugs interact to produce the carbon monoxide. How does the coherence a category gains by these causal relations affect the degree which it supports the generalization of new blank properties, such as sesamoid bones (for birds) or ceramic breaks (for automobiles)? Once again, light on this question has been shed by studies testing both background and experimentally-provided causal knowledge.

Research With Background Causal Knowledge

Numerous studies have demonstrated natural categories differ in systematic ways in the degree to which they support generalization. In one study, 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 folk theories about the coherence of natural kinds acquired by the older children. On this account, the coherence provided by emerging theories of biological kinds led the older children to expect such kinds to be more
structured and constrained, and hence more homogenous. Novel properties of biological kinds are generalized more strongly because this expectation of homogeneity extends to new properties in addition to existing ones (also see Gelman et al., 1988; Shipley, 1993).

Additional evidence comes from studies investigating which level in a taxonomic hierarchy supports the strongest inductions. Coley, Medin, & Atran (1997) presented both American undergraduates and the Itza' with a subspecies (e.g., black vultures) that exhibited an unspecified disease, and tested whether the degree to which that disease was generalized to vultures (a category at the species or folk-generic level), birds (the life form level), or all animals (the kingdom level). They found that for both groups the strength of generalizations dropped substantially when the target category was at the life form level or higher. Again, biological species appear to be especially potent targets of inductive generalizations (also see Atran et al., 1997).

One notable aspect of these latter findings is the fact that American undergraduates, like the Itza', treated the species level as inductively privileged despite the well-known result that for these individuals the basic level is normally one level higher in the hierarchy, namely, at the level of life form (e.g., tree or bird) rather than species (e.g., oak or robin) (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). One explanation that Coley et al. offer for this discrepancy is that whereas the explicit knowledge that the American students possess (which is tapped, for example, by the various tasks used by Rosch et al., such as feature listing) is only sufficient for life forms to emerge as identifiable feature clusters, their expectations are nevertheless that individual species are those biological categories with the most inductive potential. (In contrast to the Americans, for the Itza' the species level is both basic and inductively privileged, a result Coley et al. suggest reflects their greater explicit ecological knowledge as compared to the Americans.) They further suggest that these expectations reflect an essentialist bias in which species are presumed to be organized around a causally-potent essence which not only determines category membership, but also the degree to which category members are likely to share (known and to-be-discovered) features.
Can studies which experimentally manipulate the knowledge associated with a category shed any light on the details of how interfeature theoretical knowledge promotes generalizations? In particular, can we find direct evidence that a category based on an underlying cause (i.e., an essence) promotes generalizations?

**Research With Experimentally-Provided Causal Knowledge**

Rehder and Hastie (2004) addressed this question by manipulating the causal relations between known category features in order to create a laboratory analog of the essentialist knowledge associated with real-world categories. Once again, subjects were instructed on artificial but plausible categories like Romanian Rogos. In these experiments however, the Rogos' characteristic features were described as causally related to one another, rather than to a to-be-generalized property. In one condition the causal links formed a *common cause* topology in which one characteristic feature \( F_1 \) was the cause of the three remaining ones \( F_2, F_3, \) and \( F_4 \), as shown in Figure 6A. For example, for Romanian Rogos subjects would be told that butane-laden fuel \( F_1 \) caused the other three characteristic features (e.g., loose fuel-filter gasket, hot engine temperature, etc.). They were then presented with a series of Rogos which each possessed a blank novel property (e.g., ceramic breaks) and judged what proportion of all Rogos possessed that property. The particular combination of characteristic features that the Rogos possessed was varied.

The question of interest was whether the presence of a common cause causal structure, which was intended to mimic that of a category organized around a causally-potent essence (a common cause), would result in the same enhanced inductive potential exhibited by biological species. However, we felt it important to also ask whether any such enhancement required the specific common-cause causal structure. For example, one alternative hypothesis is that categories support stronger generalizations when they are organized around any central feature or theme, regardless of whether that feature is a cause or an effect. Another is that generalizations may be stronger as a result of the global coherence afforded by interfeature
relations, regardless of the specific arrangement of the causal links. To assess these possibilities, we also tested a common-effect network in which one feature ($F_4$) is caused by the other three features ($F_1$, $F_2$, and $F_3$) (Figure 6B), and a chain network in which feature $F_1$ causes $F_2$ which in turn causes $F_3$ which in turn causes $F_4$ (Figure 6C). We expected that only a common-cause network would result in stronger generalizations if they are promoted by a single underlying generative cause, both the common-cause and common-effect networks would do so if they are promoted by any central theme, and all three networks would do so if inductions are promoted by any category exhibiting overall coherence provided by interfeature causal relations.

The results are presented in Figure 7 for each causal network as a function of the particular combination of characteristic feature that the Rogo with the to-be-generalized property possessed. Each panel in Figure 7 also includes the results from a control condition which presented no causal links between characteristic features. There were two notable results. First, Figure 7A shows that, as predicted, common cause categories supported stronger inductive generalizations relative to the control condition, but only when the Rogo was typical (i.e. possessed all four characteristic features, as denoted by "1111"). When the Rogo was atypical (e.g., when it possessed the common cause $F_1$ but not the other three features, as denoted by "1000"), generalizations were weaker than in the control condition. Note that in the common cause condition this Rogo is not just atypical because it is missing three characteristic features ($F_2$, $F_3$, and $F_4$), it is incoherent in light of the category's causal laws (the common cause $F_1$, butane-laden fuel, is present but its three effects are absent).

The second important finding is that these effects were not unique to a common-cause network, because the results were essentially the same for categories with a common effect network (Figure 7B) or a chain network (Figure 7C). Each network supported stronger generalizations, so long as the category member with the novel property was typical (1111). But when the category member was incoherent, generalization were weaker than in the control condition. Exemplar 0001 is incoherent in the common effect condition because the
common effect $F_4$ is present even though its causes $F_1$, $F_2$, and $F_3$ are all absent. Exemplar 0101 is incoherent in the chain condition because $F_2$ and $F_4$ are present even though their causes $F_1$ and $F_3$ are absent (and because $F_3$ is absent even though its causes $F_2$ is present). In other words, it is not coherent categories that support stronger generalizations but rather coherent category members, and this effect holds regardless of the topology of the causal relations which make them coherent. As in my other studies reported in this chapter, these results held not just for artifacts like Rogos, but also for biological kinds and for nonliving natural kinds.

These results have three implications for the received view that biological kinds promote inductions because of the presumed presence of an essence. First, it may not be a presumption of single common cause (an essence) which promotes inductions but rather a presumption of coherence, that is, causal relations which link category features regardless of the specific topology of those links. Second, this effect of coherence is not limited to biological kinds, but will apply equally well to other kinds (e.g., artifacts) when generalizers have reason to believe that one of those kinds is coherent. Third, this effect can be reversed when the specific category member which displays a novel feature is incoherent in light of the category's causal laws.

Of course, it should be noted that our failure to discover any special inductive potential of a common cause network as compared to the other two networks may have obtained because our experimental materials failed to fully elicit our subjects' essentialist intuitions. For example, perhaps such intuitions depend on the underlying nature of a category remaining vague (e.g., even young children have essentialist intuition, even though they know nothing of DNA, Gelman, 2003), but in our categories the common cause was an explicit feature (e.g., butane-laden fuel for Rogos). Or, perhaps those explicit features clashed with our adult subjects' own opinions about the nature of the "essence" of a brand of automobile (the intention of the automobile's designer, Bloom, 1998) or a biological species (DNA). Nevertheless, these results raise the possibility that the inductive potential of categories
thought to arise from folk essentialism about biological kinds may in fact arise from a more
general phenomenon, namely theoretical coherence, which potentially applies to kinds of
categories in addition to biological kinds.

Discussion

This chapter has reviewed three sorts of evidence related to how causal knowledge is
involved in property generalization. The first section presented evidence that people can go
beyond the similarity-based heuristics formalized by the Similarity-Coverage model and
generalize on the basis of causal reasoning. The second section addressed how causal
reasoning and the similarity-based heuristics interact, and showed that subjects which showed
an effect of causal relations exhibited virtually no effect of similarity, whereas those which
showed no effect of causal relations exhibited a sensitivity to similarity. Apparently, the use
of a causal explanation versus similarity is an all-or-none matter, with reasoners using one
strategy or the other, but not both. The third section reviewed evidence suggesting that
categories presumed to be based on an underlying cause seem to be the target of strong
generalizations, and found that stronger generalizations may not be the responsibility of a
single generative essence but rather of the presumed coherence that a category gains from the
causal knowledge which interrelates existing features.

The evidence in support of these claims came from studies testing both natural and
artificial categories, approaches which have complimentary strengths and weaknesses. For
example, a critic could dismiss the studies testing artificial categories on the grounds that the
causal knowledge was provided to participants as part of the experimental session rather than
coming from a participants’ own background knowledge, a practice which was likely to have
made that knowledge especially salient and available (bypassing the need for the reasoner to
retrieve it from their semantic memory). However, this criticism is blocked by the numerous
studies testing natural categories showing that it is common for reasoners to use causal
knowledge in the service of generalizations. The benefit of testing artificial categories, in turn,
is that it has made it possible to test specific claims regarding how causal knowledge
influences generalization. One view is that properties are more generalizable to the extent they are more central (i.e., have more dependents in the base and target categories) (Hadjichristidis et al., 2004). My own view is, in contrast, that people engage in a kind of causal reasoning in which they evaluate whether a property is likely to appear in a target category. Data were presented showing that people can reason forwards from the presence of to-be-generalized property causes to the property itself (prospective reasoning) as well as they can reason from the property's effects or dependents (diagnostic reasoning). Additional experiments also showed that people will judge a new property to be more prevalent in a population of category members to the extent that its causes and/or effects are present in that population.

Although this chapter's central claim is that properties are often generalized on the basis of causal reasoning, it is important to consider whether some of the results could be explained by a more sophisticated version of similarity. For example, recent progress has been made in extending similarity-based models to account for some results with nonblank properties, such as the GAP2 model proposed by Blok, Osherson, and Medin (this volume). In addition, one component of Medin et al.'s (2003) relevance theory of induction is that similarity emerges from a process of comparing the premise and conclusion categories. 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, this volume). However, by themselves neither the GAP2 model nor comparing premise and conclusion categories are able to account, for example, for the Heit and Rubinstein (1994) study presented earlier because similarity doesn't change as a function of the property involved. The import of the Heit and Rubinstein results, however, is just that the relative 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 which is influenced by the to-be-generalized property. Indeed, they showed that judgments regarding the generalization of anatomical and behavioral properties 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 which people can compute, 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 (in press) presented in the second section. For example, there I 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 (Figure 5). I described these results as indicating that the explanation drew attention away from the other dimensions, which of course is equivalent to saying that it determined the (single) dimension along which base-target similarity was computed.

Nevertheless, these more sophisticated views of similarity fail to explain, for instance, the diagnostic reasoning which apparently occurs during some cases of generalization. In the Hadjichristidis et al. (2004) study, subjects were asked to generalize a fictitious hormone from one category to another (e.g., from seals to dolphins). According to Heit and Rubinstein, the fact that the novel property is a hormone might induce reasoners to adopt an anatomical notion of similarity. But this 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. Clearly, to explain these results one must also make reference to the causal network in which the
hormone is involved. The view I promote is that generalizers reasoned backward (diagnostically) from the presence of the many physiological functions to the presence of the hormone.

As mentioned, testing artificial categories has made it possible to begin to develop specific computational theories how causal knowledge influences generalization. But while a start has been made, my own work has only considered the causal knowledge which interrelates a category's features. But categories are also related by causal relations to one another, and research reviewed above documented the influence of one type of inter-category causal link—ingestion—on induction (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. Progress in providing a computational account of the influence of ingestion relations as one type of inter-category relation is reported by Tenenbaum, Kemp, and Shafto (this volume). In this research, subjects were trained to memorize a particular "food web" consisting of a number of animal species which prey on one another. For example, one food web included the facts that sand sharks eat herring which eat kelp; another that lions eat wolves which eat squirrels. Subjects were then asked to generalize a disease property "D" from one species to another in the same food web. Their judgments were well-predicted by a Bayesian model which computed the probability with which the disease was transmitted through the web from the base species to the target. Interestingly, when the property was "gene XR-23" the food web model performed poorly and it was a Bayesian model which operated on a taxonomic representation of the species that performed well instead. This result is reminiscent of Heit and Rubinstein's suggestion that the nature of the to-be-generalized property influences the mental processes responsible for induction, but suggests that this influence amounts to recruiting different complex knowledge structures (a food web versus a taxonomic hierarchy) rather just different ways of computing similarity.
More work is needed, because of course diseases can spread in ways other than ingestion (e.g., physical proximity, Proffitt et al., 2000). And the generalization of other types of properties 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). But one common thread between the causal reasoning view I have advocated here and that of Tenenbaum et al. (this volume) is the emphasis on domain-general reasoning processes. For example, one striking aspect of the results from Rehder (in press) and Rehder & Hastie (2004) testing artificial categories is the consistency of the results across biological kinds, nonliving natural kinds, or artifacts. Such results suggests that domain differences found with natural categories arises from differences in the type and amount of knowledge associated with those domains. Of course, this is not to suggest that such differences are theoretically unimportant, because one can always ask what causes them to exist. For example, it may be that young children make unique assumptions about the organization of biological kinds (e.g., that they are based on some kind of essence, or, as I prefer, that they are presumed to be coherent). Even so, it is causal reasoning which directly mediates whether or not a property is generalized to an entire class of objects.


Rehder, B. (2003b). A causal-model theory of conceptual representation and


Author Note

Bob Rehder, Department of Psychology, New York University.

Correspondence concerning this article should be addressed to Bob Rehder, Department of Psychology, 6 Washington Place, New York, NY 10003 (email: bob.rehder@nyu.edu).
Footnotes

1 No offense to the people of Uzbekistan, whose country was chosen at random from the too-long list of countries about which I know virtually nothing. The suggestion that Uzbekistani’s have bad breath (or, as alleged below, that they eat salted fish, drink vodka, or don’t wear business suits) is a figment of the author’s imagination.

2 Actually, you would be justified in thinking that its prevalence is slightly higher in Uzbekistan, because you do have a sample—even a very small one of size 1—of people from Uzbekistan with bad breath. After all, if bad breath is as uncommon in Uzbekistan as everywhere else, its pretty surprising that the one Uzbekistani you get to meet just happens to be one of the few that have bad breath. But all depends on your beliefs regarding the prior distribution over the prevalence of bad breath. If you believe that either everybody in a country has bad breath or no one does, then the single example would be enough to conclude in favor of the former.

3 Technically, according to the SCM generalizations between non-hierarchically nested categories are a function of not only their similarity but also the coverage the premise provides for the lowest-level superordinate category that includes both premise and target categories. This more precise definition is unimportant for what follows.

4 Note that Heit and Rubinstein themselves attributed these results to subjects computing similarity between the base and target differently depending on whether the novel property was anatomical or behavioral. I'll return to this issue in the Discussion.

5 Whether one finds multiple versus few symptoms more diagnostic of an underlying common cause will depend on the details of the causal beliefs involved. If one believes there is no other possible causes of the symptoms, then even just one symptom is sufficient to infer the cause with certainty. Also relevant is whether the cause is deterministic (produces its effects with probability 1) or probabilistic. A deterministic cause invariably produces all of its
effects, and thus the presence of the cause is ruled out if even one of its effects is absent. For a probabilistic representation of causality see Rehder (2003a, 2003b) and Cheng (1997). For discussion of a deterministic versus probabilistic view of causality, see Thagard (1999).

Note that in this experiment the effect of similarity was not completely eliminated (as the effects of typicality and diversity were eliminated in the first two experiments), as the generalization rating of nonblank properties was on average 6 points higher (on a 100 point scale) when the base and target were similar as compared to dissimilar, a difference which reached statistically significant. Nevertheless, note that the magnitude of this similarity effect is vastly lower than it is for blank properties (34 points). (See Lassaline, 1996, for related results.)

In a different phase of the experiment, Rehder and Hastie (2004) also asked the subjects to estimate the degree of category membership of the same exemplars which displayed the novel properties. Like the generalization ratings, category membership ratings for the most typical exemplars (i.e., 1111) were higher in the causal conditions than in the control condition, and ratings for incoherent exemplars (e.g., 1000, 0001, and 0101 in the common cause, common effect, and chain conditions, respectively) were lower than in the control condition. That is, the causal knowledge which makes exemplars appear more (or less) coherent is reflected in their perceived degree of category membership in addition to their propensity to support generalizations. Overall, the correlations between generalization and category membership ratings were .95 or higher.
### Table 1
Features and causal relationship for Romanian Rogos, an artificial category.

<table>
<thead>
<tr>
<th>Characteristic Feature</th>
<th>Novel Feature</th>
<th>Characteristic Feature → Novel Feature</th>
<th>Novel Feature → Characteristic Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Butane-laden fuel</td>
<td>Zinc-lined gas tank</td>
<td>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.</td>
<td>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.</td>
</tr>
<tr>
<td>Loose fuel filter gasket</td>
<td>Vibrations during breaking</td>
<td>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.</td>
<td>Vibration during breaking causes a loose fuel filter. The rattling caused by the vibrations eventually leads to the fuel filter gasket becoming loose.</td>
</tr>
<tr>
<td>Hot engine temperature</td>
<td>Thin engine oil</td>
<td>Hot engine temperature causes thin engine oil. The oil loses viscosity after it exceeds a certain temperature.</td>
<td>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.</td>
</tr>
<tr>
<td>High amounts of carbon monoxide in the exhaust</td>
<td>Inefficient turbocharger</td>
<td>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.</td>
<td>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.</td>
</tr>
</tbody>
</table>
Table 2
Category structure from Nair (2005, Experiment 2), consisting of eight Rogos (R1-8) and eight non-Rogos (NR1-8).

<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rogos</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R8</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Non-Rogos</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NR1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>NR2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>NR3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>NR4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NR5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NR6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NR7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NR8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 1
Results from Hadjichristidis et al. (2004), Experiment 2.
Figure 2
Results from Rehder (in press), Experiment 3.
Figure 3

A

Explicit Base Rate

B

Observed Base Rate
Figure 4
Results from Rehder (in press), Experiment 1.
Figure 5
Results from Rehder (in press), Experiment 3.
Figure 6
Network topologies tested in Rehder and Hastie (2004, Experiment 2).

A. Common Cause Network

B. Common Effect Network

C. Chain Network
Figure 7
Generalization ratings from Rehder and Hastie (2004, Experiment 1).