Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognitive Psychology xxx (2004) xxx–xxx
Cognitive Psychology
www.elsevier.com/locate/cogpsych

Feature inference and the causal structure of categories☆

Bob Rehder[a,*], Russell C. Burnett[b]

[a] *Department of Psychology, New York University, USA*
[b] *Department of Psychology, Northwestern University, USA*

Accepted 21 September 2004

Abstract

The purpose of this article was to establish how theoretical category knowledge—specifically, knowledge of the causal relations that link the features of categories—supports the ability to infer the presence of unobserved features. Our experiments were designed to test proposals that causal knowledge is represented psychologically as Bayesian networks. In five experiments we found that Bayes' nets generally predicted participants' feature inferences quite well. However, we also observed a pervasive violation of one of the defining principles of Bayes' nets—the *causal Markov condition*—because the presence of characteristic features invariably led participants to infer yet another characteristic feature. We argue that this effect arises from a domain-general bias to assume the presence of underlying mechanisms associated with the category. Specifically, people take an exemplar to be a "well functioning" category member when it has most or all of the category's characteristic features, and thus are likely to infer a characteristic value on an unobserved dimension.

☆ We thank Sergio Chaigneau, Douglas L. Medin, Gregory L. Murphy, Lance Rips, and three anonymous reviewers for their helpful comments on an earlier draft of this manuscript.
* Corresponding author. Fax: +1 212 995 4349.
E-mail address: bob.rehder@nyu.edu (B. Rehder).

0010-0285/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2004.09.002

1. Introduction

Research of the past 15 years has shown the importance of causal knowledge—specifically, knowledge of causal relations between category-associated features—in the acquisition and use of natural concepts. Categories tend to form around clusters of causally related features (Ahn & Medin, 1992; Medin, Wattenmaker, & Hampson, 1987). Supervised category learning will depend on the causal relations that hold between a category's features (Waldmann, Holyoak, & Fratianne, 1995), and with other categories (Lien & Cheng, 2000). Interfeature causal relations also influence how items are classified (e.g., Ahn, 1998; Ahn, Kim, Lassaline, & Dennis, 2000; Rehder & Hastie, 2001; Rehder, 2003a, 2003b; Sloman, Love, & Ahn, 1998), and how novel properties are generalized to categories (Hadjichristidis, Sloman, Stevenson, & Over, 2004; Medin, Coley, Storms, & Hayes, 2003; Rehder & Hastie, 2004).

In this article, we ask whether and how this kind of causal knowledge is used to make inferences about unobservable or unobserved features of novel objects. Imagine coming across an unfamiliar bird and making an inference about whether it is likely to fly, or finding an unfamiliar plant and judging whether it is safe to eat. In each of these cases, an inference would presumably draw on (a) the observable features of the object and (b) prior, more general knowledge of a category to which the object belongs. We ask how causal relations between features—when they are available as part of the reasoner's prior knowledge of the category—are used in inference.

We begin, in the next section, with two candidate models of feature inference. The first is based on an object's typicality relative to a salient category. It does not implicate causal knowledge but serves as a useful standard for comparison. The second is based on causal relations between features and makes specific predictions about inference. To foreshadow, in Experiment 1 we find a systematic deviation from these predictions. Experiments 2–5 test alternative models designed to explain this deviation.

1.1. Approaches to feature inference

1.1.1. Feature inference by typicality

Imagine coming across an unfamiliar bird and making an inference about whether it is likely to fly. One way to make this inference is based on how typical it is of the bird category, which has flight as an associated feature. On this approach, a bird that is highly typical of the category—for example, a robin—would be judged more likely to fly than would a less typical bird—for example, an ostrich.

Although typicality seems intuitive, its rationale is not so clear. At first glance it appears to be related to the probabilistic view of concepts, on which categories form around clusters of correlated features (Rosch & Mervis, 1975). But whereas Rosch emphasized the interfeature correlations that obtain *between* categories (and which thus define clusters of features), typicality based feature inferences are licensed only if features are also correlated *within* category (such that category features are inferentially dependent on one another). Whether natural categories actually exhibit such

within-category correlations is an open question. Another possible rationale for typicality is that an object with many features characteristic of a category is more likely to, in fact, *be* a member of that category, and to possess other features characteristic of that category as a result. We address this possibility in Experiment 1. For now we wish to establish that typicality is an intuitive basis for inference but one without a clear rationale.

It has been shown that in some circumstances people do seem to make typicality based feature inferences. For example, Yamauchi and Markman (2000) taught people artificial categories and found that exemplars that possessed more features in common with training exemplars supported stronger inferences of unobserved features. This result obtained despite the absence of within-category feature correlations in the training data, and is thus suggestive of the possibility that people have a general tendency to infer features on the basis of typicality.

A brief caveat is called for here. Thus, far we have used *typicality* to mean centrality in a category: A typical category member is one with features possessed by many other category members. But typicality may be influenced by factors other than centrality. For example, exemplars are sometimes viewed as more typical to the extent that they satisfy a goal or ideal that the reasoner associates with the category (Barsalou, 1985; Burnett, Medin, Ross, & Blok, in press; Lynch, Coley, & Medin, 2000). In addition, when causal or theoretical knowledge is present, exemplars are viewed as more typical to the extent they exhibit the correlations among features that such knowledge leads one to expect (e.g., an animal that lives underwater should also have gills) (Ahn, Marsh, Luhmann, & Lee, 2002; Malt & Smith, 1984). As we shall show, consideration of some of these additional influences on typicality will bear on its potential as an explanation for people's patterns of feature inferences.

1.1.2. Feature inference by causal reasoning

A typicality based inference depends on a quality of the whole object (its total number of characteristic features), and is driven by a general expectation that better examples of the category are more likely to have any unobserved feature. For a reasoner who knows the specific causal relations that hold between features, however, there is an alternative: feature-to-feature inference, in which the presence or absence of an unobserved feature is inferred from the presence or absence of specific features to which it is related (Medin, 1983).

Consider again our unfamiliar bird and an inference about its ability to fly. An alternative to the typicality approach is to reason about the causes or enablers of flight in birds: large wings relative to body size, aerodynamic shape, and so on. Other features, like a characteristically shaped beak, may be regarded as less relevant to the inference, even though these features make the bird more typical of its category. On this approach, feature inference is a matter of causal reasoning.

To see in greater detail how this sort of reasoning might be done, it is useful to represent categories as causal models in which features appear as nodes and causal relations as directed links between nodes (Rehder, 2003a, 2003b; Waldmann et al., 1995). The structure in Fig. 1, for example, represents a category in which one feature, F_1, causes three others, F_2, F_3, and F_4.

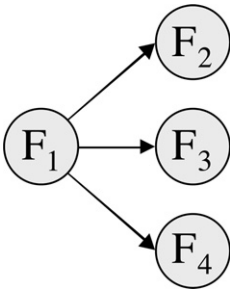


Fig. 1. Common Cause network.

Consider a category member in which F_2, F_3, and F_4 are observed (each is observed to be either present or absent) and the state of F_1 is unobserved and to be inferred. In this case the states of F_2, F_3, and F_4 are all relevant. To the extent that each of these effect features is present, it increases our confidence that its cause, F_1, is present, because from the presence of an effect one can infer the presence of its cause.

Now consider a category member in which F_1, F_2, and F_3 are observed (present or absent) and the state of F_4 is unobserved and to be inferred. In this case, F_1 is relevant. Its presence increases our confidence that any of its effects—in this case, F_4—is also present. What of F_2 and F_3? Because F_2 and F_3 are not directly causally related to F_4, the only way in which they could provide inferential support to F_4 is by providing evidence about F_1. But in this case we already know that F_1 is present (or absent). Since F_1 is observed, F_2 and F_3 become irrelevant to the inference. They are "screened off" from F_4 by F_1.

This screening-off principle is central to causal reasoning (Hausman & Woodward, 1999). It has traveled under various names—contained, for example, in Reichenbach's (1956) "principle of the common cause." In language associated with Bayesian networks, it is captured in the *causal Markov condition*, which states that a variable is independent of its nondescendents conditional on the states of its immediate parents (Pearl, 2000). In the current example, the effects of a common cause are independent of one another conditional on the observed state of the cause. We will call this method of feature inference—by causal reasoning that respects the causal Markov condition—the *straightforward causal-reasoning* method.

Note that the straightforward causal-reasoning and typicality approaches can lead to different inferences. One reason for this is that features relevant to typicality will often, due to the causal Markov condition, be irrelevant in causal reasoning. For a category with a common-cause structure (Fig. 1), the causal Markov condition states that information about the presence or absence of F_2 and F_3 is irrelevant to inferring F_4 given knowledge of F_1. In contrast, typicality predicts that inferences to F_4 will be stronger when F_2 and F_3 are present even when F_1 is observed, because such an object is more typical of its category.

To return to the central question of this paper, how are feature inferences made when causal knowledge is available? Our first goal will be to show that causal knowledge is used in feature inference by demonstrating, for example, that features are

more likely to be judged present when their cause(s) or effect(s) are present. Our second goal will be to determine more precisely how causal knowledge is used. In particular, in Experiment 1 we test whether feature inferences honor the screening off principle.

2. Experiment 1

Participants learned about a novel category and then made inferences about unobserved features of individual category members. In a common cause condition, participants learned (a) four features associated with the category and their likelihoods of being present in category members and (b) causal relations between these features. In particular, these participants learned that the four features were related in a common-cause structure (Fig. 1), such that one of the features caused, by independent mechanisms, the other three. For each causal relation, an underlying mechanism was provided. In a control condition, participants learned the features and their likelihoods of being present in category members, but no causal relations or mechanisms. The design thus allows us to isolate the effect of causal knowledge on feature inference.

In the common cause condition, we predicted that causal knowledge will be used powerfully in inference and that its use will be consistent with the straightforward causal-reasoning account outlined earlier. First, inferences about F_1 should be stronger as a function of the number of effects (F_2, F_3, and F_4) present. Second, consistent with the screening-off principle (the causal Markov condition), inferences about an effect feature—say, F_4—should be sensitive to whether its cause, F_1, is present, but not to whether the other effects—in this case, F_2 and F_3—are present when F_1 is observed.

In each graph the vertical axis represents the probability that the to-be-inferred feature is present (higher values = more probable). The overall number of features observed present in an object appears on the horizontal axis. Because the one to-be-inferred feature is unobserved, the number of features observed present varies from 0 to 3. The top panel shows predictions for items in which the common cause, F_1, is unobserved and to be inferred. On these items, we expect each of the effect features, when it is present, to drive inferences upward.[1] The bottom panel shows predictions for items in which one of the effects—F_2, F_3, or F_4—is to be inferred. (The overall number of features varies from 0 to 2 when the cause, F_1, is observed absent, and from 1 to 3 when it is observed present, due to the absence/presence of F_1 itself.) On these items, we expect inferences to be low when F_1 is absent and high when F_1 is present. Importantly, in keeping with the causal Markov condition, we expect inferences to be *uniformly* low when F_1 is absent and *uniformly* high when F_1 is present.

[1] Our prediction that inferences to the common cause are an increasing function of the number of effects rests on the assumption that the causal relationships are viewed as probabilitic rather than necessary. That is, although the presence of one effect will raise the probability that the common cause is present, its presence will not be certain.

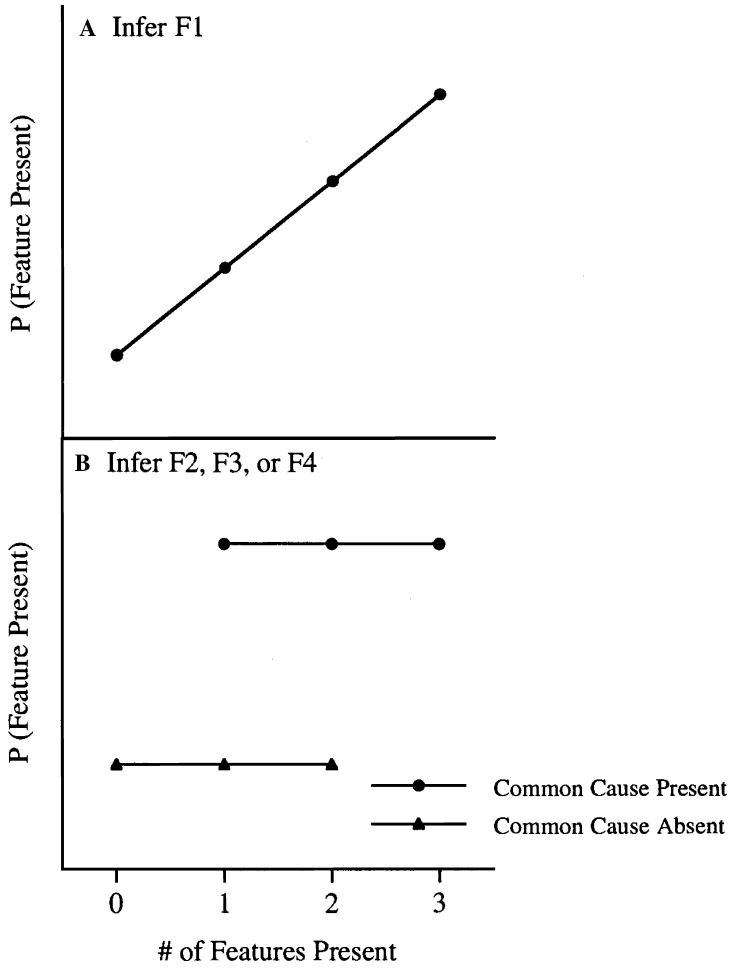


Fig. 2. Normative predictions for the Common Cause network of Fig. 1. (A) items in which F_1 was inferred. (B) items in which either F_2, F_3, or F_4 was inferred.

That is, we expect both of the lines in the bottom panel of Fig. 2 to be flat; positive slopes in these lines would indicate that, in violation of the causal Markov condition, effect features are providing inferential support for one another even when their common cause is observed.

In the control condition, of course, F_1 should have no special influence on inferences about F_2, F_3, and F_4, since these participants are given no causal knowledge and therefore no reason to distinguish F_1 from the other three features. However, if inferences reflect typicality based reasoning, then they should strengthen with the number of features observed present in an item. That is, category members that are more typical (in virtue of having more category-associated features) will be deemed more likely to have yet one more category-associated feature.

This experiment was also designed to detect any influence of an exemplar's goodness of category membership on feature inference. In a categorization phase, participants were asked to judge the likelihood of category membership of each of the exemplars for which they made inferences about unobserved features.

2.1. Method

2.1.1. Materials

Six novel categories were used: two biological kinds (Kehoe Ants, Lake Victoria Shrimp), two nonliving natural kinds (Myastars [a kind of star], Meteoric Sodium Carbonate), and two artifacts (Romanian Rogos [a kind of car], Neptune Personal Computers). Each category had four featural dimensions, and each dimension had two possible values. For example, for Lake Victoria shrimp the four dimensions were quantity of ACh neurotransmitter, speed of flight response, rate of sleep cycle, and body weight. The two values on the first dimension were "a high quantity of ACh neurotransmitter" and "a low quantity of ACh neurotransmitter"; the two values on the second dimension were "a fast flight response" and "a slow flight response"; the two values on the third dimension were "an accelerated sleep cycle" and "a decelerated sleep cycle"; and the two values on the fourth dimension were "high body weight" and "low body weight."

One value on each dimension was characteristic of the category. Henceforth we call these values "characteristic features," or just "features," and the characteristic features on the four dimensions are designated F_1, F_2, F_3, and F_4. For example, a participant learning about Lake Victoria Shrimp might learn that F_1 is "a high quantity of ACh neurotransmitter," F_2 is "a slow flight response," F_3 is "an accelerated sleep cycle," and F_4 is "low body weight." Each of these features was said to be present in 75% of category members, and the other value on each dimension was said to be present in 25% of category members. This information was summarized in a table that listed the four dimensions, their possible values, and the likelihoods of these values—for example, "quantity of ACh neurotransmitter: high (75%) or low (25%)."

For participants in the common cause condition, each category was given the three causal relationships shown in Fig. 1: $F_1 \rightarrow F_2$, $F_1 \rightarrow F_3$, and $F_1 \rightarrow F_4$. Each description of a causal link specified the cause, the effect, and the causal mechanism linking them—for example, "A high quantity of ACh neurotransmitter causes a long-lasting flight response. The duration of the electrical signal to the muscles is longer because of the excess amount of neurotransmitter." In addition, participants were shown a summary diagram much like Fig. 1 (with values substituted for variable names).

Which value was said to be characteristic on each dimension was counterbalanced over participants. This was done to allow for any preexperimental associations participants might have had between specific values (e.g., between a high quantity of a neurotransmitter and an accelerated sleep cycle) and for the possibility that they might infer associations between values based on qualities like "high" and "low" (e.g., associating a high quantity of a neurotransmitter with high body weight

because both are "high"). If we arbitrarily designate one value on each dimension as "+" and the other as "−," then the values that were described as occurring with probability 75%, in four between-subject counterbalancing conditions, were "++++," "++−−," "+−+−," and "+−−+." For example, a participant in the "+−−+" condition who learned the Lake Victoria Shrimp category was told that the 75% values were "high quantity of the ACh neurotransmitter" (the "+" value on dimension 1), "slow flight response" ("−" value on dimension 2), a "decelerated sleep cycle" ("−" value on dimension 3), and a "high body weight" ("+" value on dimension 4). As a result of this counterbalancing, the values on one dimension were combined with those on another as serving the roles of the 75% values an equal number of times across participants. Dimension values were mixed in this way so that any influence of interfeature relations that participants brought with them to the experiment would be canceled out by averaging over participants. The features and a sample of the causal relations associated with all six categories are given in Appendix A. (Appendix A also includes interfeature causal relationships that will be used in subsequent experiments.)

2.1.2. Participants

Forty-eight New York University undergraduates received course credit or pay for participating in this experiment.

2.1.3. Design

Participants were randomly assigned in equal numbers to one of the six categories, to either the common cause or the control condition, and to one of the four feature counterbalancing conditions.

2.1.4. Procedure

The experiment had three phases: learning, categorization, and inference. The learning phase came first; the categorization and inference phases were then presented in counterbalanced order. All phases of the experiment were conducted by computer, though each phase was introduced by spoken instructions.

In the learning phase, participants studied several screens of information about the category at their own pace. All participants read a cover story and a description of the characteristic and noncharacteristic values (including their likelihoods, 75 or 25%) on the four dimensions. Participants in the common cause condition also learned about the three causal relations both in verbal form and in diagrammatic form (much like Fig. 1). To ensure that all information was learned, participants had to pass a multiple-choice test. In the control condition, this test consisted of 7 questions about the four dimensions and the likelihoods of the values on each dimension. In the common cause condition, the test contained an additional 14 questions about the causal relations. While taking the test, participants were free to return to the information screens they had studied; however, doing this obligated the participant to retake the test. The only way to pass the test and proceed to subsequent phases was to take it all the way through without errors and without returning to the initial information screens for help.

In the feature inference phase, participants were shown descriptions of category members in which one feature was unobserved and judged whether it was absent or present. A description consisted of four lines of text, one for each of the four dimensions. For the three dimensions that were observed, the text indicated whether the feature was present (e.g., "a high quantity of ACh neurotransmitter") or absent (e.g., "a low quantity of ACh neurotransmitter"). For the unobserved dimension the text was simply "???". Note that category membership was certain and emphasized; the question asked was, for example, "Does this Lake Victoria Shrimp have a low body weight or a high body weight?" Responses were entered by positioning a slider on a scale whose ends were labeled, for example, "low" and "high"; these responses were recorded as a 0–100 rating, where 0 meant certainty that the feature was absent (e.g., "low") and 100 meant certainty that it was present (e.g., "high"). Inferences were made about all 32 possible category members in which one feature is unobserved and each of the other three features is observed to be either present or absent. These items were presented in a different random order for each participant.

In the categorization phase, participants were presented with the same 32 exemplars for which they inferred unobserved features. In this phase, however, the exemplars were not labeled as known category members; instead participants were asked to rate the likelihood that the exemplar was a member of the category (e.g., Lake Victoria Shrimp). Responses were entered on a scale whose ends were labeled "sure that it isn't" and "sure that it is" (recorded as 0 = sure that it isn't a category member, 100 = sure that it is). Items were presented in a different random order for each participant.

2.2. Feature inference results

Initial analyses revealed no effect of which category participants learned, the order of the two tasks, or the feature counterbalancing condition, and thus the results are collapsed over these factors.

Fig. 3 presents feature inference ratings as a function of (a) the total number of features observed present in an exemplar; (b) whether the common cause, F_1, was present or absent (if it was observed); and (c) condition (common cause or control). In fact, inferences in the control condition were consistent with the typicality approach described earlier. Inferences about any of the four features were an increasing, roughly linear function of the number of other features observed present. Participants in this condition reasoned as if they expected the four features to be correlated with one another and thus to provide inferential support to one another.

In the common cause condition, inferences about F_1 were inferences about a cause given knowledge of its effects, whereas inferences about F_2, F_3, and F_4 were inferences about an effect given knowledge of its cause and other effects of this cause. First consider inferences about F_1 (Fig. 3A). As in the control condition, these increased as a function of the number of other (now, effect) features observed present. Interestingly, the slope of this function is greater in the common cause condition than in the control condition, indicating that F_2, F_3, and F_4 were seen as more strongly predictive of F_1 when they were construed as effects of F_1 than when they

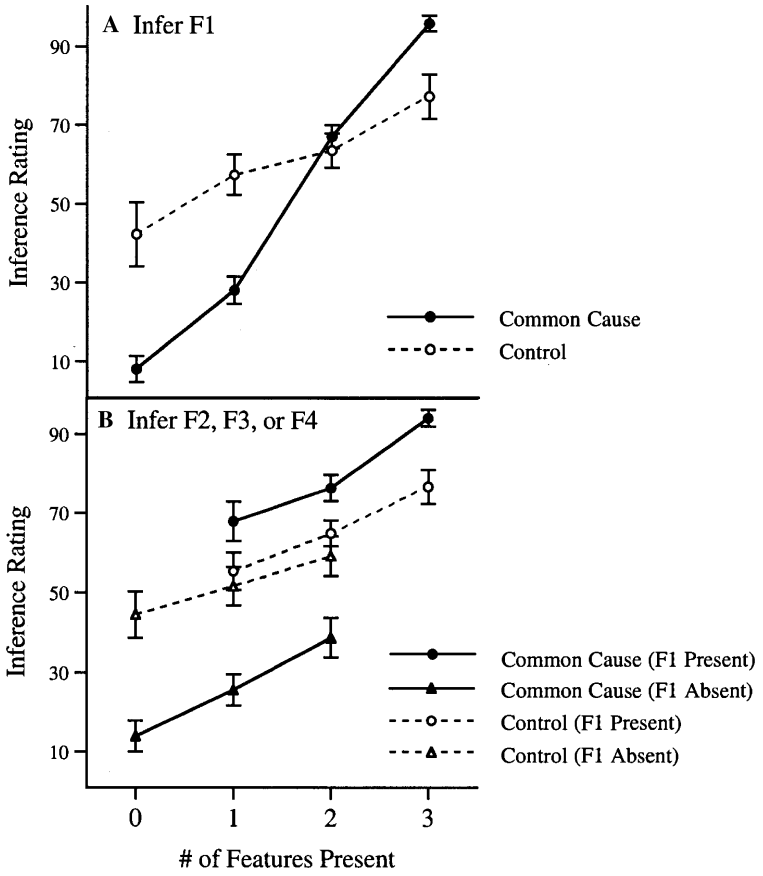


Fig. 3. Feature inference results from Experiment 1. (A) items in which F_1 was the feature to be inferred, and (B) items in which either F_2, F_3, or F_4 was to be inferred.

were construed merely as characteristic features. This influence of causal knowledge on feature inference was confirmed by submitting each participant's ratings to a regression analysis in which the sole predictor was the number of features observed present. The average weight assigned to this predictor was significantly greater in the common cause condition than in the control condition (30.2 vs. 11.1), $t(46) = 5.25$, $p < .0001$.

The influence of causal knowledge is also revealed in inferences about F_2, F_3, and F_4 (Fig. 3B). In the common cause condition these inferences were heavily influenced by the presence or absence of the common cause, F_1; other observed features had less influence. For example, items with one feature present received much higher ratings when that feature was F_1 (as in the item 1x00, which represents F_1 present, F_2 unobserved and to be inferred, F_3 absent, F_4 absent) than when it was one of the other features (e.g., 0x10 or 0x01) (means = 68.0 and 25.7, respectively). Similarly, items with two features received higher ratings when one of those features was F_1 (e.g.,

1x10 or 1x01) than when neither was F_1 (e.g., 0x11) (means = 76.4 and 38.8). A comparison of the control and common cause conditions shows that causal knowledge, when it was available, was implicated heavily in inference, such that inferences about effects were based largely on whether their cause was present or absent.

However, these inferences deviated systematically from straightforward causal reasoning, too. According to the causal Markov condition, an inference about an effect feature (when the common cause is observed) should not be influenced by the presence/absence of any other effect feature (recall the flat lines in the lower panel of Fig. 2). But, as can be seen in Fig. 3B, participants' inferences increased with the number of effect features observed present. When F_1 was observed absent, feature inference ratings were 14.0, 25.7, and 38.8 for objects possessing 0, 1, and 2 effect features, respectively; when F_1 was observed present, ratings were 68.0, 76.4, and 94.0 for objects possessing 0, 1, and 2 effect features (i.e., 1, 2, and 3 features overall), respectively. That is, effect features provided inferential support to one another, even though their common cause was observed. We will refer to this as a *nonindependence effect*, because features which, on the straightforward causal reasoning account, should be independent of one another are in fact treated as predictive of one another.

These conclusions are supported by statistical analysis. Each participant's inferences about effect features (F_2, F_3, and F_4) were predicted from a multiple-regression model in which the two predictors were (a) the number of features present and (b) a term representing the presence or absence of F_1. In evidence of the effect of causal knowledge on inference, the average regression weight associated with F_1 in the common cause condition (20.3) was both significantly greater than zero, $t(23) = 5.75$, $p < .0001$, and significantly greater than that weight in the control condition (2.5), $t(46) = 4.61$, $p < .0001$, which was itself not significantly different from zero, $t(23) = 1.58$, n.s. Moreover, the average weight associated with the number of features present was significantly greater than zero in both the common cause condition (12.7), $t(23) = 5.16$, $p < .0001$, and the control condition (9.0), $t(23) = 2.75$, $p < .05$. This sensitivity to number of features did not differ between the two conditions, $t < 1$.

2.3. Categorization results

One possible explanation of the nonindependence effect is that it is driven by likelihood of membership in the category. This explanation goes as follows. Features provide inferential support to one another because they make the item that possesses them a better or more likely member of the category; such a category member, in turn, is more likely to have a characteristic feature on an unobserved dimension. Results from the categorization phase allow us to rule out this explanation.

Because our goal is to explain the nonindependence effect, we focus here on the categorization results for the same 24 items that elicited this effect in the feature inference phase—that is, the 24 items in which the presence or absence of F_1 was observed and one of the effect dimensions was unobserved. On the one hand, in the control condition we found that participants' categorization ratings were highly correlated with their judgments regarding whether an exemplar possessed an un-

known feature, $r = .986$. In contrast, the correlation between feature inference and categorization ratings was much weaker in the common cause condition ($r = .696$). To demonstrate this weaker relationship between category membership and feature inference, Fig. 4A presents the categorization results from the common cause condition organized in the same way that the feature inference results are organized in Fig. 3B: according to whether F_1 was present or absent, and by the overall number of characteristic features present in an item.

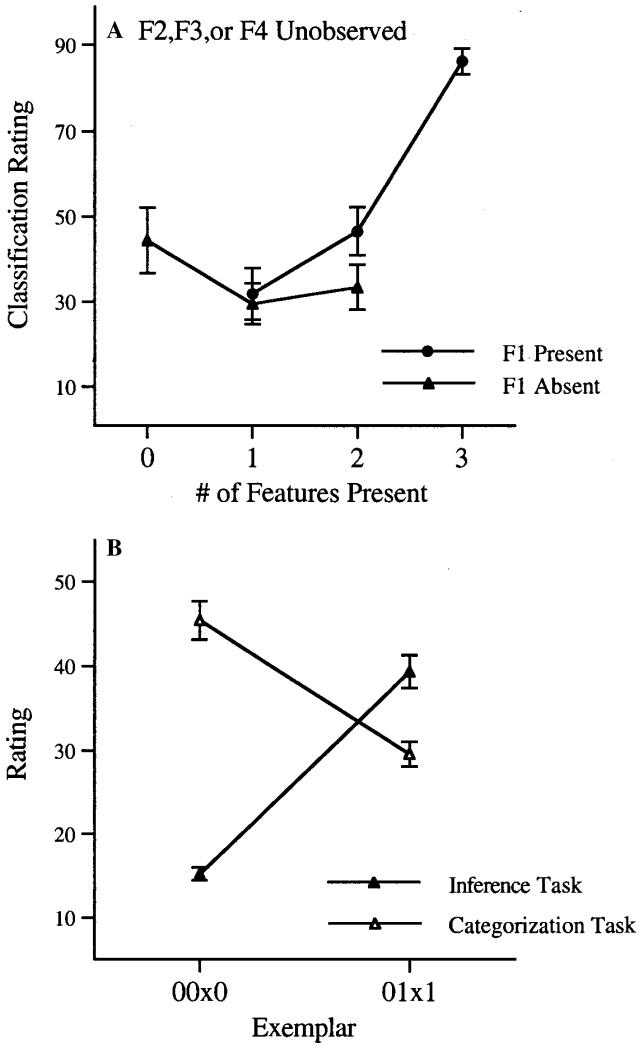


Fig. 4. (A) Categorization results from common cause condition of Experiment 1. (B) Relationship between categorization and inference results from common cause condition for two exemplars.

If the nonindependence effect were due to degree of category membership, then the slopes of the "common cause" lines in Fig. 3B should be mirrored in the categorization ratings in Fig. 4A. In fact, comparison of the figures shows that whereas feature inference ratings increased as a function of the number of effect features when the common cause was absent, the corresponding categorization ratings *decreased* as a function of the number of effect features. Apparently, when the common cause was known to be absent, participants judged exemplars to be *worse* category members to the extent they possessed many effect features because such exemplars contradict the causal relationships that they believed the category possessed: if the common cause is absent, then so too should be its effects.

An example of this dissociation between feature inference and categorization in the common cause condition is presented in Fig. 4B. Participants were more likely to infer that exemplar 0x11 possessed a characteristic feature on the second dimension than exemplar 0x00, and they made this judgment despite the fact that they considered 0x00 to be a better category member than 0x11. In other words, the nonindependence effect manifested by the different inference ratings for 0x11 and 0x00 cannot be due to 0x11 being viewed as a better or more likely category member than 0x00.

2.4. Discussion

The first purpose of this experiment was to show the use of causal knowledge in feature inference. On this, the results were clear. In the common cause condition, inferences about the common cause feature were based heavily on the presence/absence of its effects, and inferences about effect features were based heavily on the presence/absence of their cause. In this respect, inferences in the common cause condition differed sharply from analogous inferences in the control condition.

The second purpose was to begin to discover *how* causal knowledge is used in inference. We took, as an initial standard against which to evaluate participants' performance, a straightforward causal reasoning theory which includes the causal Markov condition. Participants' inferences deviated systematically from this theory, in that inferences were generally stronger as a function of the number of typical, or characteristic, features the exemplar had, even when those features were supposed to have been screened off from the feature in question. This finding of a nonindependence effect is important, because the causal Markov condition is crucial to the standard normative account of causal reasoning (Bayesian network theory) and therefore to recent psychological models based on this account (e.g., Gopnik et al., 2004).

Another important result was the effect of typicality in the control condition. This result obtained despite the fact that control participants had just learned that all four features had the same base rate (75%), a situation which might have led one to expect that the same rating would be produced on each feature inference trial. Instead, the results amounted to another form of violation of independence, as ratings indicated that they thought characteristic features were predictive of one another. Note that these findings extend those of Yamauchi and Markman (2000), who also obtained a nonindependence effect, but only when participants were first trained to classify

examples of category members. Experiment 1's participants, in contrast, never observed any category members. Apparently, a mere verbal statement of a category's characteristic features is sufficient to induce people to infer the presence of one characteristic feature given the presence of others, even in the absence of observed category members.

One rationale we gave for nonindependence was that degree-of-category-membership might be a reasonable basis for feature inference. But if this were true, then the nonindependence effect should have been mirrored in categorization ratings of the very same items. Instead, in the common cause condition there was a strong dissociation between feature inference and categorization. Analysis revealed that this dissociation arose because one factor that determined an exemplar's goodness of membership—whether observed features contradicted or corroborated the category's interfeature causal links—did not influence feature inference. This finding replicates previous research by Rehder (2003a; 2003b; Rehder and Hastie, 2001), who found that exemplars are good category members to the extent they manifest the expected pattern of correlations between causally related features.

The fact that inferences in Experiment 1 were *insensitive* to whether the exemplar manifested expected interfeature correlations suggests that the common cause participants simply accepted what we told them and treated the exemplars as category members while inferring their features. And if the common cause participants accepted the exemplars' category membership, there's no reason to believe that the control group did not as well.[2] Thus, degree-of-category membership turns out to an inadequate explanation of nonindependence in feature inference in either condition.

[2] The claim that feature inference is independent of degree of category membership is consistent with a series of studies by Murphy, Ross, and Malt, who found that inferences about items were generally not influenced by uncertainty about whether the items were indeed members of the categories that supported those inferences (Malt, Ross, & Murphy, 1995; Murphy & Ross, 1994; Ross & Murphy, 1996).

Another possible explanation of the nonindependence effect is that the exemplar with the to-be-predicted feature was perceived as more typical of the category. (Note that in this discussion it is important to distinguish our empirical effect—the fact that inferences increased in strength with the exemplar's number of characteristic features [a.k.a., the nonindependence effect]—from typicality as an explanation for that effect.) It is important to consider typicality separately from degree of category membership because the two are not always equivalent (e.g., Armstrong, Gleitman, & Gleitman, 1983; Barsalou, 1985; Burnett et al., in press). Nevertheless, we believe that the categorization results have ruled out the typicality explanation as well, on the assumption that if our participants had been asked for ratings of typicality rather than likelihood of category membership, the results would have been the same. This assumption is supported in several ways. First, the only relevant reason why typicality and likelihood of category membership diverge is that the former can be influenced by a reasoner's ideals or goals (e.g., Barsalou, 1985; Burnett et al., in press). This is not a problem in the current study, because participants did not associate ideals or goals with the categories. Second, in a recent study the second author has asked for typicality ratings of stimuli very much like the ones used here, and the results were identical. Finally, empirical research suggests that judgments of typicality are sensitive to interfeature consistency, just as judgments of category membership are. For example, Ahn et al. (2002) found that typicality in natural categories is influenced by consistency with known causal relations (see also Malt & Smith, 1984). Thus, there is good reason to believe that, if participants in our experiments had been asked for typicality ratings, those ratings would have been sensitive to causal consistency and would therefore have diverged from feature inference ratings in just the same way that likelihood-of-category-membership ratings did.

The puzzle then is: Why are features inferentially dependent on one another such that characteristic features are predictive of other characteristic features? We attempt to answer this question in the remainder of this article.

3. Experiment 2

One general approach to understanding the nonindependence effect is to consider what sources of knowledge may have influenced participants' inferences in addition to the interfeature causal relations we provided them with. In Experiment 1 we controlled for one sort of prior knowledge: By mixing the values that were characteristic on the four dimensions, we averaged over any preexperimental inter-feature associations that participants might have had. Still, we can envisage another, more general sort of prior knowledge that might have been responsible for the nonindependence effect. Though the specific categories we used were novel, they came from domains with which participants have a wealth of experience, namely, biological kinds, nonliving natural kinds, and artifacts. Participants may have augmented the categories' causal models with domain knowledge in a way that led to the nonindependence effect.

For example, participants who learned about Lake Victoria Shrimp and Kehoe Ants may have used their knowledge from the domain of biology while predicting unobserved features for those categories. It has been argued that people believe that biological kinds have underlying properties and biological mechanisms that give rise to observable features (Gelman, 2003; Medin & Ortony, 1989). As a result, participants may have assumed that the four features of Lake Victoria Shrimp and Kehoe Ants were each caused by the biological mechanisms associated with those species. If this were the case, it would explain the violations of independence found with these categories, because from the presence (absence) of one feature one can infer the presence (absence) of the underlying mechanism, and then from the underlying mechanism one can then infer the presence (absence) of an unobserved feature. Said differently, participants may have reasoned that the biological mechanisms associated with Lake Victoria Shrimp and Kehoe Ants had operated normally when the ant or shrimp possessed many characteristic features, and hence that those mechanisms were likely to have produced a characteristic value on the unobserved feature dimension as well. Conversely, when the ant or shrimp possessed many uncharacteristic features, they reasoned that something had gone awry with the operation of that species' normal mechanisms, and hence the presence of a characteristic feature was less likely.

A similar pattern of reasoning may have also occurred for our novel artifact categories. There is evidence demonstrating the importance of causal history in people's mental representation of artifacts (Bloom, 1998; Keil, 1995; Matan & Carey, 2001; Rips, 1989), and our participants may have assumed that the characteristic features of Romanian Rogos (an automobile) or Neptune Computers arose as an effect of their manufacturing process. When a particular Rogo or Neptune Computer exhibited many characteristic features they may have inferred that this process

had operated normally, and hence were more likely to infer a characteristic value on the unobserved feature dimension. Finally, even for our nonbiological natural kinds (Myastars and Meteoric Sodium Carbonate) participants may have assumed the presence of a causal chain of events that led to their formation, which in turn led to stronger inferences to characteristic features when other characteristic features were already present.

Another way we can imagine that domain knowledge may have contributed to the results from Experiment 1 is that participants could have spontaneously constructed interfeature relations among a category's characteristic features. For example, participants who were told that typical Myastars were especially hot and had a large number of planets may have used their domain knowledge to construct some reason why hot temperature caused large number of planets (or vice versa), whereas those who were told that Myastars were hot and had few planets may have constructed a reason for why hot temperature produces few planets (or vice versa). On the feature inference task, the presence of these self-generated explanations would then have led participants to infer the presence of one characteristic feature given the presence of others (producing the apparent violations of independence).

In Experiment 2 we address the possibility that people augment their causal models with domain knowledge by using a category whose domain is not identified. Participants were told that they would be learning about a new kind of object named "Dax" with four features labeled A, B, C, and D which each occurred with probability 75%. Because participants had no basis for believing that Daxes were biological, an artifact, or any other type of category, they had no reason to assume the presence of any domain-specific kind of underlying mechanism responsible for generating observable features.

3.1. Method

3.1.1. Materials

Participants were told that Daxes were some new kind of object about which they should learn. In the common cause condition they were in addition told that feature A caused features B, C, and D. During the feature inference and categorization tasks participants were presented with Daxes whose feature lists indicated whether each feature was present or absent or unknown. For example, during the feature inference task participants were told about a Dax that had "A," "no B," "C???," and "D" and asked to infer whether it had feature C or not.

3.1.2. Participants

Forty-eight New York University undergraduates received course credit or pay for participating in this experiment.

3.1.3. Design

Participants were randomly assigned in equal numbers to either the common cause or the control condition, and to the task-order counterbalancing factor (whether the classification or the inference task was performed first).

3.1.4. Procedure

The procedure was nearly the same as that in Experiment 1. In the common cause condition there were fewer questions on the multiple-choice test because no information about causal mechanisms was provided. Also, there were no unobserved features in the exemplars presented during the categorization phase.[3]

3.2. Feature inference results

The feature inference results are presented in Fig. 5. Results in the common cause condition were comparable to those in Experiment 1. Most importantly, though inferences about effect features (Fig. 5B) were based heavily on the presence/absence of the common cause F_1, they also increased with the number of other effect features present. That is, common cause participants again exhibited a substantial nonindependence effect, and did so despite the blank materials used in this experiment.

Per-participant regression analyses identical to those conducted in Experiment 1 confirmed that, for inferences about effect features (Fig. 5B), the average regression weight associated with F_1 in the common cause condition (13.3) was both significantly greater than zero, $t(23) = 2.31$, $p < .0001$, and significantly greater than the corresponding weight in the control condition (1.8), $t(46) = 4.09$, $p < .0001$. In evidence of the nonindependence effect, the average weight associated with the number of features in the common cause condition (5.7) was significantly greater than zero, $t(23) = 2.38$, $p < .05$.

The results in the control condition differed from those in Experiment 1 in that, overall, inferences did not increase with the total number of features present (i.e., lines in Fig. 5 for the control condition are essentially flat). At first glance, this seems to suggest that the nonindependence effect found in the control condition of Experiment 1 was absent here. Also unlike Experiment 1, however, there was a significant effect of task order (inference first vs. categorization first) on the influence of the number of features present. Participants who performed the categorization task first gave inference ratings that increased with the number of features present; that is, they showed the same nonindependence effect seen in Experiment 1. In contrast, participants who performed the inference task first showed a nonindependence effect in the opposite direction; their inference ratings decreased as the number of features present increased. These trends can be seen in Fig. 6 (which shows data averaged over all 32 feature inference items). Consistent with the effect of task order, a two-way mixed ANOVA of the control condition revealed a significant interaction between number of features present and task order, $F(3, 66) = 9.21$, $p < .0001$.

[3] This was the case for the classification task in Experiments 3–5 as well. Because the results from the classification tests from these experiments are thus not directly comparable to those from the feature inference task (as they were in Experiment 1), we omit reporting classification test results in the remainder of the article.

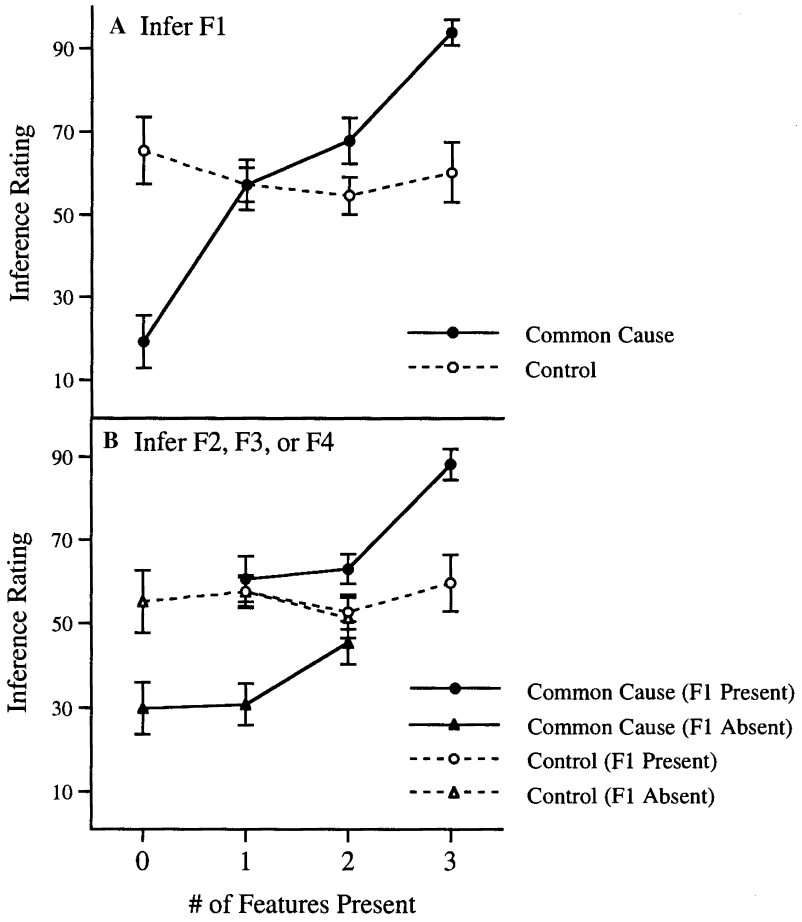


Fig. 5. Feature inference results from Experiment 2. (A) Items in which F_1 was the feature to be inferred, and (B) items in which either F_2, F_3, or F_4 was to be inferred.

3.3. Discussion

The purpose of Experiment 2 was to test whether domain knowledge about biological kinds, nonliving natural kinds, and artifacts was responsible for the nonindependence effect found in Experiment 1. The fact that this effect obtained even with the use of blank categories suggests that the nonindependence effect is not due to knowledge of underlying mechanisms in those domains. Moreover, in the absence of domain knowledge, it is hard to see on what basis participants would have spontaneously generated interfeature explanations such as "B causes C."

As in Experiment 1, we found violations of feature independence in the control condition. However, a surprising result of Experiment 2 was that the direction of this violation depended on whether the feature inference task preceded or followed the

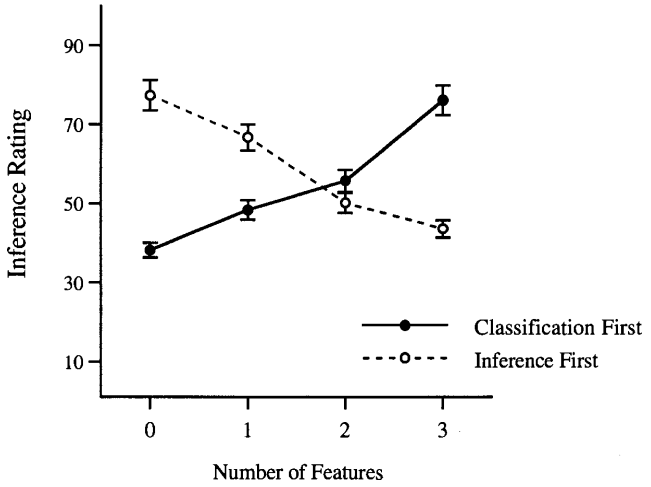


Fig. 6. Feature inference results from control condition of Experiment 2.

categorization task. Though this effect of task order is largely incidental to the thrust of this paper, it does call for a brief explanation. We suggest that the behavior of the participants who performed the inference task first and who were *less* likely to infer the presence of a characteristic feature in an item that already possessed more characteristic features can be understood as an instance of the "gambler's fallacy" (Tversky & Kahneman, 1974). These participants treated the four features as a series of events in which the occurrence of one kind of outcome (likely or unlikely) in the three observed events predicted the occurrence of the other kind of outcome (unlikely or likely) in the fourth, unobserved event. This did not happen in Experiment 1 because in that experiment it was clear that the four features were features of a category. In Experiment 2, it was clear that the four features were indeed features of a category when the categorization task was performed first; when the inference task was performed first, it was less clear. This is a vivid illustration of the power of categories to guide inference. For events of a general kind, people often take typical outcomes as predictive of atypical ones; however, when the same events are interpreted as features of a category, characteristic features predict more characteristic features instead.

3.4. Individual response patterns

A primary finding in Experiments 1 and 2 was that the inferences of common cause participants were sensitive to both the specific causal relations that were provided, and the presence or absence of characteristic features. However, one possibility we have not yet considered is that this pattern of results might have arisen as a result of averaging over participants. That is, some common cause participants may have reasoned normatively (i.e., honored the causal Markov condition) whereas

others may have ignored the causal relations and responded only on the basis of typicality.

To examine this possibility, Fig. 7 presents the two regression weights for each participant in Experiments 1 and 2: the weight associated with the presence/absence of F1, and the weight associated with the overall number of features observed present in an item. In Fig. 7, participants are located in regression weight space, so that the relative weights assigned to these two factors by each participant can be readily seen.

In the common cause conditions of Experiments 1 and 2, there was variation in strategy. Informally, the response patterns given by these participants fell into a few different classes. Some were roughly consistent with the causal Markov condition in assigning high weight to the presence/absence of the common cause and little or no weight to the presence/absence of other features. These participants gave uniformly low ratings when F_1 was absent, and uniformly high ratings when F_1 was present. Others were consistent with typicality (as were most of the responses given in the control conditions); these participants gave little or no weight to F_1 and instead reasoned from the overall number of features observed present in an item. Most importantly, a large number of participants were "compromisers" who assigned moderate weight to both factors. That is, whereas one might have supposed that the trends reported in Experiments 1 and 2 were artifacts of averaging across "causal Markov" and "typicality" participants, it is in fact the case that a large number of participants showed just the trends that we have reported. These participants made powerful use of causal knowledge, but their inferences also showed the nonindependence effect.

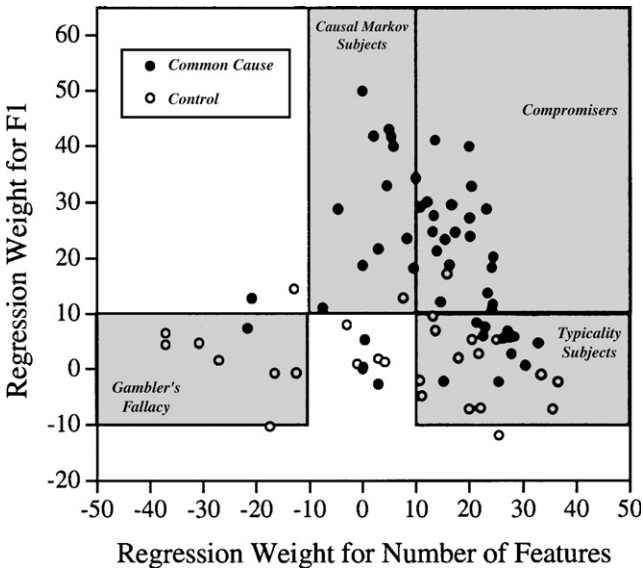


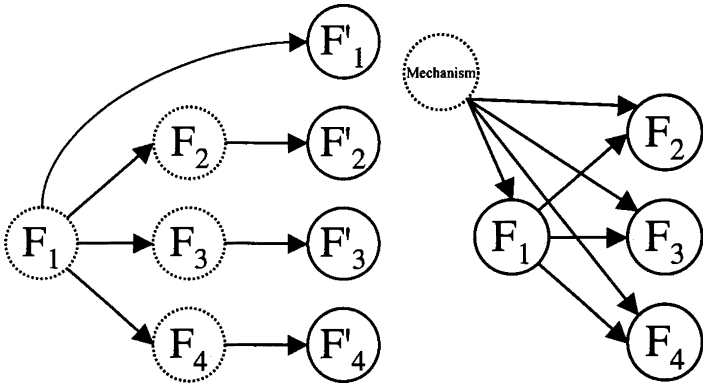
Fig. 7. Individual differences in Experiments 1 and 2.

4. Experiment 3

To summarize our findings so far, we have demonstrated a robust nonindependence effect, and also provided evidence against a number of explanations for that effect. First, the dissociation between inference and classification (Experiment 1) indicates that feature inference is not mediated by the exemplar's perceived likelihood of category membership. Second, an analysis of individual differences indicates that the nonindependence effect is not an artifact of averaging over participants. Finally, this effect was not due to domain knowledge that supports the spontaneous construction of interfeature relations, or which informs participants about the presence of underlying causal mechanisms (Experiment 2).

In Experiment 3 we continue to consider the possibility that the nonindependence effect arises not because people are suboptimal causal reasoners, but rather because they were reasoning with knowledge in addition to that which we provided. Although Experiment 2 ruled out the possibility that this knowledge was domain specific, in Fig. 8 we present two domain-general ways in which our participants may have extended the categories' causal models so as to produce a nonindependence effect. According to the first possibility, the *Feature Uncertainty Model* (Fig. 8A), participants have doubts about whether an exemplar's observed features are veridical. For example, when they were presented with a category member that had F_1 but not F_2 and F_3 and were asked to infer F_4, the absence of F_2 and F_3 may have led them to doubt that F_1 was really present (because F_1 should have produced F_2 and F_3). Uncertainty about F_1's presence meant that the inference to F_4 was weakened.

The causal model in Fig. 8A represents this situation by encoding whether features are present (nodes F_1–F_4 with dotted lines) separately from the *evidence* that those features are present (nodes F'_1–F'_4 with solid lines). The fact that the evidence



(A) Feature Uncertainity Model for Common Cause network.

(B) Underlying Mechanism Model for Common Cause network.

Fig. 8. Alternative causal models for the common cause network of Fig. 1.

nodes are informative, but not infallible, is represented by the probabilistic causal linkages from F_i to F_i' reflecting that the presence (absence) of a feature tends to produce evidence that the feature is present (absent). According to this model, (apparent) violations of the causal Markov condition arise because even when evidence for the presence of the common cause F_1 is available (F_1'), additional evidence that one or more of its effects are present allows one to be even more certain that F_1 is truly present. Greater certainty regarding the common cause allows one to make more confident inferences to an unknown effect.

Our second proposal is the *Underlying Mechanism Model* (Fig. 8B). According to this model, people have a bias to view categories as possessing underlying mechanisms, and this bias leads them to reason from observed features to the presence of those mechanisms, and then to the presence of an unobserved feature. Note that this model is structurally identical to the possibility—considered in Experiment 2—that participants augment their causal models with knowledge of the underlying mechanisms associated with biological kinds, artifacts, and nonliving natural kinds. The Underlying Mechanism Model assumes, however, that the bias to view categories as possessing underlying mechanisms is domain general. As a result, the knowledge about underlying mechanism is schematic, or skeletal, in that no understanding of how the causal mechanism operates is assumed to be present. Despite the abstract nature of this knowledge, however, it is assumed to be sufficient to influence feature inference.

The purpose of Experiments 3–5 was to test these two alternative causal models. We start in Experiment 3 by testing some of the predictions these models make for more complex inferences than those tested thus far. In Experiments 1 and 2 we found that participants made use of their knowledge of causal relations to make inferences between features that are directly connected by causal links (i.e., an effect is more likely when its immediate cause is present, and vice versa). However, Bayesian networks also make predictions for indirect inferences in which features are separated by more than one causal link. For example, although in a common cause network it would be invalid to infer the state of one effect from another when the state of the common cause was known (the causal Markov condition), this inference would be licensed when the state of the common cause was unknown. This is the case because from the state of the known effect one could infer the likely state of the common cause, and then the likely state of the unknown effect.

In Experiment 3 we assess the manner in which participants make indirect inferences by presenting them with exemplars with two unobserved features and asking them to infer one of them. Fig. 9 presents the predictions for such inferences for both the simple common cause model (Figs. 9A and B) and for the Feature Uncertainty and Underlying Mechanism Models (Figs. 9C and D). (In Fig. 9, the number of features present ranges from 0 to 2 because each exemplar had two observed features.) For the simple common cause model, the predictions for direct inferences are the same as those in Fig. 2: Inferences to the common cause become stronger as the number of effects present increases (Fig. 9A), and inferences to an effect are independent of other effects when the common cause is observed (Fig. 9B). The new prediction in Fig. 9B is for indirect inferences: When the common cause is unknown, inferences to an effect should be stronger as a function of the number of other effects present.

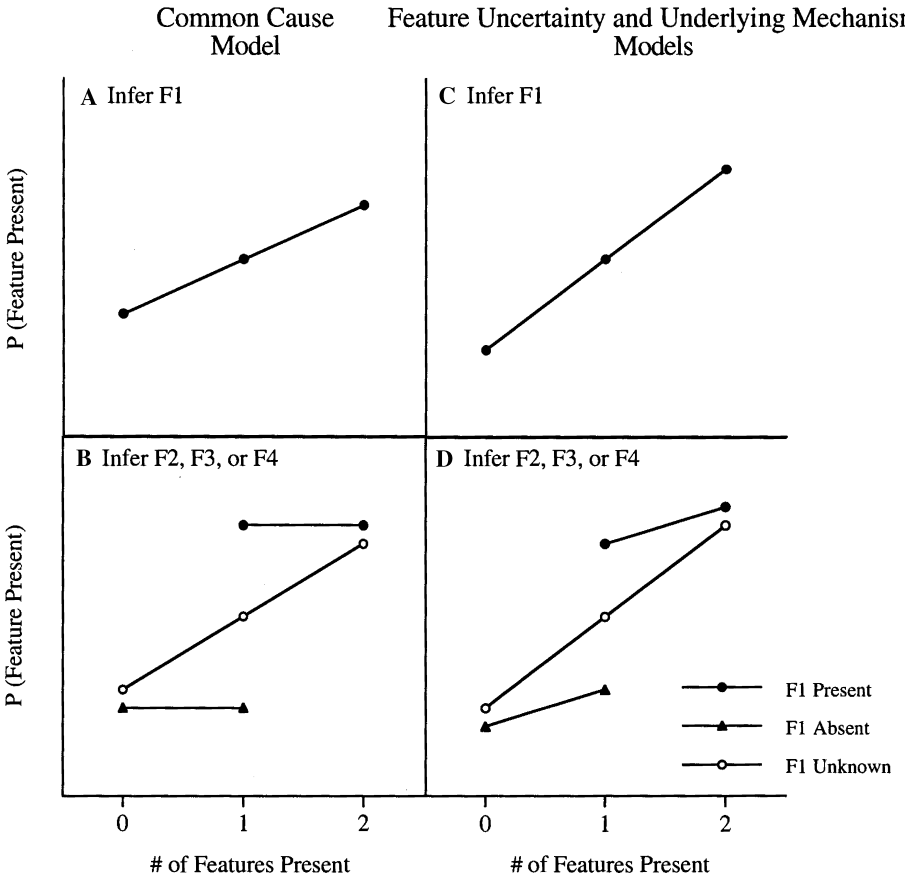


Fig. 9. Normative predictions for Experiment 3.

The predictions of the Feature Uncertainty and Underlying Mechanism Models for Experiment 3 are presented Figs. 9C and D. In contrast to the simple common cause model, these models predict that inferences to an effect should be sensitive to the number of other effects when the common cause is observed—that is, they predict the nonindependence effect found in Experiments 1 and 2. Nevertheless, Fig. 9D also shows that both models predict that indirect inferences (i.e., cases when the common cause is unobserved) should exhibit greater sensitivity to the number of other effects (i.e., the slope should be more positive) as compared to direct inferences.[4]

[4] The ordinal predictions in Fig. 9 hold within condition for a wide range of parameterizations of the Bayes' nets shown in Fig. 8 when the interfeature causal links are relations of probabilistic rather than deterministic necessity and sufficiency. In particular, they hold when (a) causal links are of equal strength, and (b) each of the features of the causal model are equally probable (as stipulated in the description of the categories presented to participants).

A finding that both direct and indirect inferences follow the overall pattern shown in Figs. 9C and D will provide additional support for the two alternative causal models we have proposed as explanations for the nonindependence effect. Whereas these models make the same predictions for Experiment 3, starting in Experiment 4 we will begin to conduct tests that discriminate between them. We will show that although the models make the same predictions for a common cause network, they make different predictions for other network topologies.

4.1. Method

4.1.1. Materials

In Experiment 3 we returned to the novel categories first used in Experiment 1 (Lake Victoria Shrimp, Kehoe Ants, etc.). However, to provide a bit more generality to our findings, we made a minor modification to the description of the feature values. Instead of the characteristic and uncharacteristic features being polar opposites (e.g., "fast flight response" vs. "slow flight response"), the uncharacteristic feature was described as normal relative to a superordinate category (e.g., "75% of Lake Victoria Shrimp have a fast flight response, whereas 25% have a normal response").

4.1.2. Participants

Forty-eight Northwestern undergraduates received course credit or pay for participating in this experiment.

4.1.3. Design

Participants were randomly assigned in equal numbers to either the common cause or the control condition, and to one of the six categories. The order of the classification and inference tasks was randomized for each participant.

4.1.4. Procedure

The procedure was identical to that in Experiment 1, with the exception that the exemplars presented on the feature inference task possessed two unobserved features, and participants were asked to infer one of them. During this task 48 distinct feature inference problems were presented.

4.2. Feature inference results

Initial analyses revealed no effect of which category participants learned or the order of the two tasks, and thus the results are collapsed over these factors. The results are presented in Fig. 10. When inferring the presence of feature F_1, both the common cause (Fig. 10A) and control (Fig. 10C) participants increased their ratings as a function of the total number of features already present. Ratings were again more sensitive to the total number of features for the common cause group, which suggests that, as in Experiments 1 and 2, they reasoned backwards from the effect features to the common cause feature. Per-participant regression analyses with number of features as the predictor confirmed that the effect of the number of features was sig-

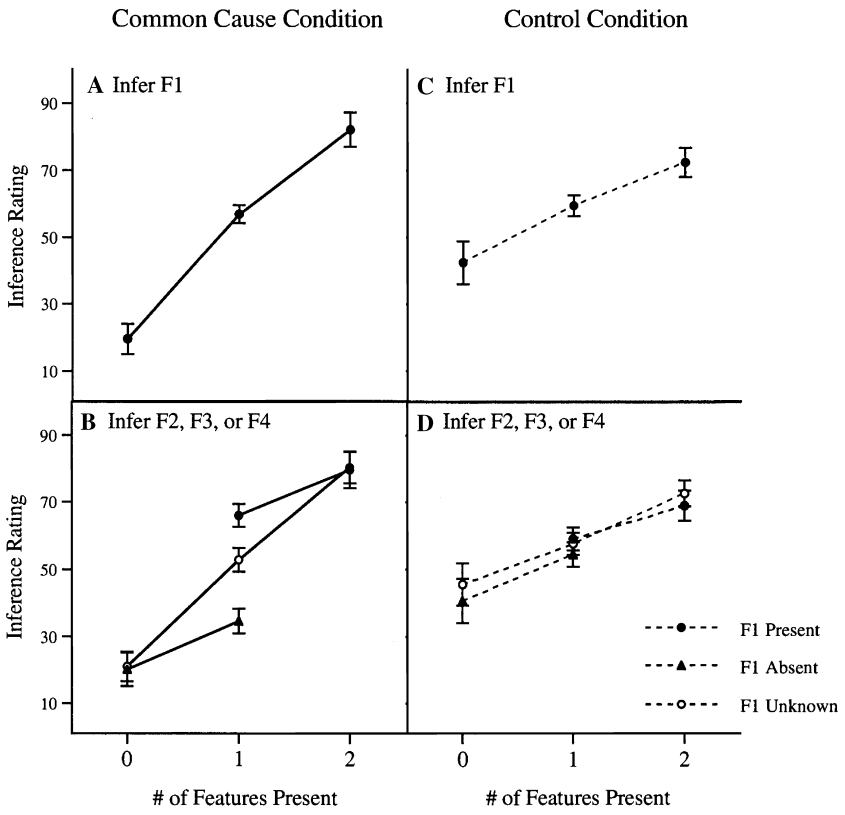


Fig. 10. Feature inference results from Experiment 3.

nificant in both conditions (both p's $< .0001$) but that the regression weight was greater in the common cause (29.6) than in the control condition (13.6), $t(46) = 2.66$, $p < .05$.

The results of interest are those in which an effect feature is inferred as a function of whether the common cause is present, absent, or unobserved and the number of effect features present (Fig. 10B). The results confirm the predictions of the Feature Uncertainty and Underlying Mechanism Models. As in Experiments 1 and 2, when the common cause feature is observed ratings increase as the number of effects increases. However, this sensitivity to number of effects is stronger when the common cause is unobserved, as predicted by our two alternative causal models (Fig. 9D).

Per-participant regressions were conducted with five predictors: a contrast code representing whether the common cause was observed or unobserved; a contrast code representing whether, if it was observed, the common cause was present or absent; the number of effect features present in the exemplar; and two predictors representing the interactions between the contrast codes and the number of effects. As expected, there was an overall effect of whether the common cause was present or absent [regression weight of 15.7, significantly different from zero, $t(23) = 4.82$,

$p < .0001$]. In addition, there was a significant effect of the number of effect features observed present [weight of 20.7, $t(23) = 4.43$, $p < .0001$]. Importantly, the influence of the number of effect features was moderated by whether the common cause was observed or unobserved, $t(23) = 4.98$, $p < .0001$, confirming that the positive slope in Fig. 10B was significantly greater when the common cause was unobserved than when it was observed.

As expected, results in the control condition (Fig. 10D) were a simple function of the number of effect features present (weight of 12.9), $t(23) = 3.00$, $p < .01$.

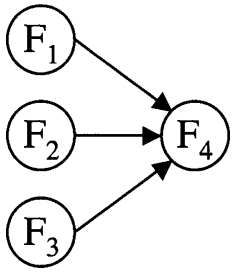
4.3. Discussion

The purpose of Experiment 3 was to test the patterns of inference predicted by the Feature Uncertainty and Underlying Mechanism Models. There were two notable results. First, as in Experiments 1 and 2, in the common cause condition inferences to an effect feature were stronger to the extent that other effect features were present when the common cause was observed, an apparent violation of the causal Markov condition. Second, when inferring an effect feature, participants' ratings were even more sensitive to the number of other effects when the common cause was unobserved. This pattern of results was predicted by both Feature Uncertainty and Underlying Mechanism Models. In particular, these models explain the apparent violation of the causal Markov condition in terms of participants' reasoning with a more complex model than the one with which we provided them, one which assumes that observed features are not veridical, or that features are related by unstated causal mechanisms.

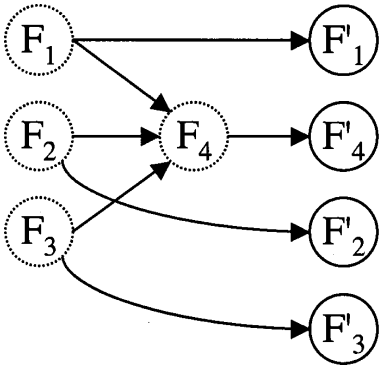
5. Experiment 4

The goal of Experiment 4 is to discriminate between the Feature Uncertainty and Underlying Mechanism Models. To accomplish this, instead of the common cause network used in Experiments 1–3 we used the common effect structure shown in Fig. 11A. In the common effect network, one category feature (F_4) is independently caused by each of the other features (F_1, F_2, and F_3). To understand the unique predictions that the two models make for a common effect structure, we first consider the predictions made by the common effect model itself for the same direct and indirect inference problems used in Experiment 3 (ones in which two features are observed and the task is to infer one of the two unobserved features). These predictions are shown in Figs. 12A and B. As expected, when inferring the common effect feature F_4, inferences should be stronger to the extent that more causes are present (Fig. 12A). Also as expected, when inferring one of the cause features (Fig. 12B), inferences are stronger when the common effect is present, weaker when it is absent, and intermediate when it is unknown.

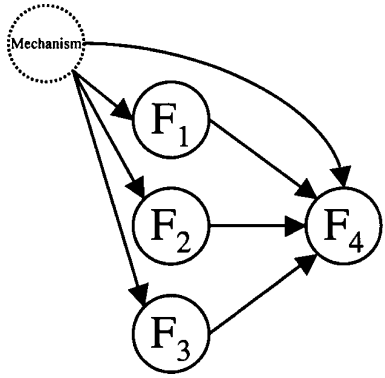
Of special interest is the sensitivity of inferences to a cause feature as a function of the number of other causes present, and how that sensitivity depends on whether the common effect is present, absent, or unobserved (Fig. 12B). When the common effect



(A) Common-Effect Model.



(B) Feature Uncertainty Model for Common Effect network.



(C) Underlying Mechanism Model for Common Effect network.

Fig. 11. Alternative Common Effect Models.

is absent or unobserved, inferences to a cause should be independent of the other causes. This is because causes are (unconditionally) independent of one another in a common effect network. In contrast, when the common effect is present, inferences to a cause should be *weaker* when another cause is present, because that other cause provides a potential explanation of the common effect. The weakened inference to a cause in the presence of another cause is an example of *discounting* in the case of multiple sufficient causation during causal attribution (Morris & Larrick, 1995). The distinct pattern of predictions associated with a common cause versus a common effect network has been the focus of considerable investigation in both the philosophical and psychological literatures (Rehder & Hastie, 2001; Rehder, 2003a; Reichenbach, 1956; Salmon, 1984; Waldmann & Holyoak, 1992; Waldmann et al., 1995).

In Figs. 11B and C we present Feature Uncertainty and Underlying Mechanism versions of a common effect network, and the predictions of those models are presented in Figs. 12C–F. The two models differ regarding their predictions when a cause feature is being inferred. On the one hand, the Underlying Mechanism Model predicts that the slope of each line will be shifted to the positive (Fig. 12F) relative to the basic common effect model. This is the case because from any cause one can infer

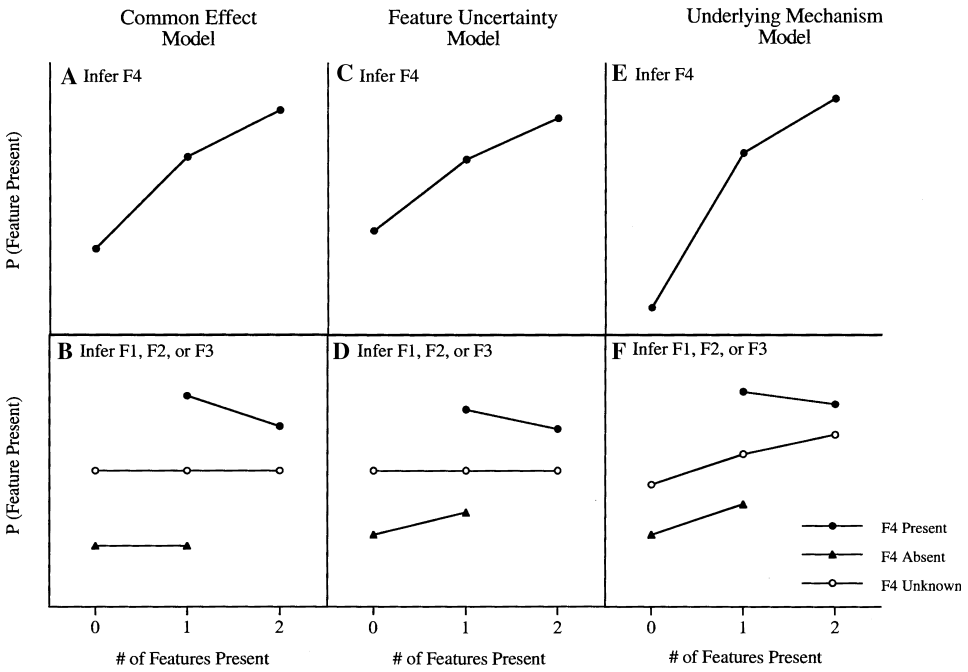


Fig. 12. Normative predictions for Experiment 4.

another cause, because those causes are themselves linked by an underlying mechanism (Fig. 11C).[5] In contrast, the Feature Uncertainty model (Fig. 11B) predicts that a cause feature does not imply another cause when the common effect is unobserved (Fig. 12D). This is the case because causes are independent of one another, and hence knowledge of one cause yields no information about another. In Experiment 4 we test whether indirect inferences follow the pattern predicted by the Feature Uncertainty Model or the Underlying Mechanism Model.

5.1. Method

5.1.1. Materials

The materials used in Experiment 4 were identical to those in Experiment 3, except that common effect participants were taught the three causal links that make up a common effect network: $F_1 \rightarrow F_4$, $F_2 \rightarrow F_4$, and $F_3 \rightarrow F_4$ (see Appendix A).

[5] Once again, the ordinal predictions in Fig. 12 hold within condition assuming that (a) the causal relations are probabilistic and of equal strength, and (b) each feature is equally probable. An exception is the predictions for the Underlying Mechanism model when the common effect is observed (Fig. 12F). In that case, reasoning through the underlying mechanism could potentially outweigh the discounting effect such that even the F_4 present line in Fig. 12F would have a positive shift in slope.

5.1.2. Participants

Forty-eight New York University undergraduates received course credit or pay for their participation.

5.1.3. Design

Participants were randomly assigned in equal numbers to either the common effect or the control condition, and to one of the six categories. The order of the classification and inference tasks was randomized for each participant.

5.1.4. Procedure

The procedure was identical to that in Experiment 3.

5.2. Feature inference results

Analogous to Experiment 3, the 48 feature inference trials were grouped according to the number of features observed present in the exemplar and whether the to-be-inferred feature was the common effect or one of the cause features (and the latter results were grouped according to whether the common effect was present, absent, or unobserved). There was no effect of which category participants learned or the order of the two tasks, and thus the results are collapsed over these factors. The results are presented in Fig. 13. When inferring the presence of feature F_4, participants in both the common effect (Fig. 13A) and control (Fig. 13C) conditions increased their ratings as a function of the total number of features already present. Ratings were more sensitive to the total number of features for the common effect group, which suggests that they reasoned forward from the presence/absence of the cause features to the presence/absence of the common effect. Per-participant regressions with number of features as the predictor confirmed that this effect was significant in both conditions (both p's $< .0001$) but that the regression weight was significantly greater in the common effect condition (35.2) than in the control condition (24.7), $t(46) = 2.68$, $p < .05$.

Fig. 13B presents the results when a cause feature is inferred as a function of whether the common effect is present, absent, or unobserved, and the number of other cause features present. The results confirm the predictions of the Underlying Mechanism Model and contradict those of the Feature Uncertainty Model. First, when the common effect is present, ratings *decrease* as the number of cause features increases. Second, when the common effect is absent, ratings *increase* with the number of causes. Thus far, these results are consistent with the predictions of both alternative models. However, when the common effect feature is unobserved, ratings increase with the number of causes present, as predicted by the Underlying Mechanism Model but not the Feature Uncertainty Model (Figs. 12D vs. F).

Each participant's inference ratings for items in which a cause feature was to be inferred were predicted with a regression equation with five predictors: a contrast code representing whether the common effect was present versus unobserved or absent, a contrast code representing whether it was unobserved or absent, the number of cause features present, and the two interactions between the contrasts and the number of causes. There was a significant effect of the number of cause features present (regression

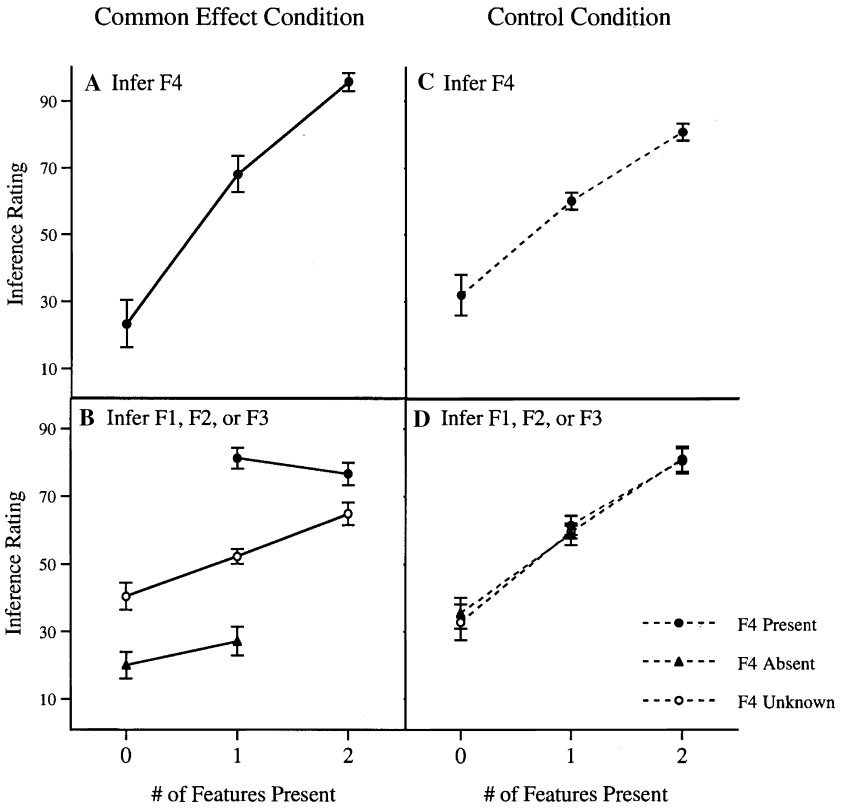


Fig. 13. Feature inference results from Experiment 4.

weight of 7.84), $t(23) = 2.72, p < .01$, a significant difference between whether the common effect was present versus unobserved or absent (weight of 5.4), $t(23) = 8.42$, $p < .0001$, and a significant difference between whether the common effect was unobserved versus absent (weight of 5.4), $t(23) = 5.21, p < .0001$. In addition, the effect of number of causes was different when the common effect was present versus absent or unobserved, $t(23) = -4.70$, $p < .0001$, reflecting the fact that ratings decreased as the number of causes increased when the common effect was present but not otherwise. Finally, the effect of number of features did not differ according to whether the common effect was absent versus unobserved, $t(23) = 1.74, p > .10$, a result consistent with the Underlying Mechanism Model but not the Feature Uncertainty Model.

As a direct test of the effect of number of features present when the common effect was unobserved (crucial for discriminating between our two alternative models), the ratings given to F_4-unobserved items in the common effect condition were submitted to regressions in which the sole predictor was the number of features observed present. As expected, the average weight assigned to this predictor was significantly greater than zero, $t(23) = 4.77, p < .0001$. Again, this is consistent with the Underlying Mechanism Model but not the Feature Uncertainty Model.

As expected, results in the control condition (Fig. 13D) were a simple function of the number of features present (weight of 21.9), $t(23) = 5.75$, $p < .0001$, replicating the results of Experiment 3.

5.3. Discussion

One goal of Experiment 4 was to discriminate between the two models we have advanced as explanations for the nonindependence effect. We tested a common effect network in Experiment 4 because, unlike the common cause network used in Experiments 1–3, for this network the two models make distinct predictions. We found that inferring a cause in a common effect network was influenced by the presence of other causes even when the common effect was unobserved, a result consistent with the Underlying Mechanism Model but not the Feature Uncertainty Model.

Once again, we also observed a nonindependence effect in the control condition, in which inferences were stronger to the extent that other features were present. In fact, this finding provides additional evidence in favor of the Underlying Mechanism Model over the Feature Uncertainty Model, because the Feature Uncertainty Model provides no explanation of nonindependence when features are causally unrelated: Features should be inferentially independent regardless of whether one is confident that observed features are really present or not. In other words, only the Underlying Mechanism Model provides a complete account of the results from all conditions of the first four experiments.

Another goal of Experiment 4 was to demonstrate that participants' feature inferences were consistent with some of the more complex predictions made by the Underlying Mechanism Model, especially those that depend on the direction of causality in the network. To this end, it is instructive to compare performance on the common cause network in Experiment 3 with the common effect network, because those networks have the same structure if one ignores the direction of causality. On the one hand, when a common cause is present, inferences to one of its effects should strengthen with the number of other effects present (Fig. 9D). In contrast, the analogous inference in a common effect network (to a cause when the common effect is present) can *weaken* with the number of other causes present, because of a discounting effect that arises when other causes already provide an explanation for the effect (Fig. 12F). In fact, these predictions are borne out in participants' feature inference ratings (Figs. 10B and 13B). The fact that these inferences were sensitive to the asymmetry in causal relationships provides yet further support for our claim that our participants were engaged in causal reasoning with the network we have called the Underlying Mechanism Model.

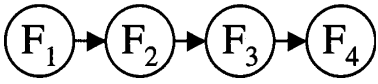
6. Experiment 5

In considering explanations for the nonindependence effect, we have been considering the possibility that participants are not suboptimal reasoners, but rather that they were reasoning with causal knowledge other than that with which we provided

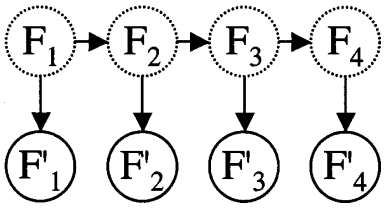
them. In fact, the results of four experiments have provided support for the claim that people's default causal structure for categories includes the presence of an underlying mechanism linking category features. However, because one must always exercise caution in appealing to an unobserved variable (in this case, underlying causal mechanisms) to explain experimental results, it is important to garner support for such a variable from sources which are as diverse as possible. To this end, in Experiment 5 we tested yet a third network, a chain network in which feature F_1 causes F_2 which causes F_3 which causes F_4 (Fig. 14A).

Fig. 15A presents the predictions of the chain network for feature inference problems in which values on three dimensions are observed and the value on the fourth dimension must be inferred (as in Experiments 1 and 2). For purposes of these predictions, we consider the "distance" of each observed feature from the to-be-inferred feature: features that are separated from the unobserved dimension by one causal link (i.e., immediate neighbors) versus those that are separated by two or three links. Specifically, the number of features present at each distance is computed by scoring 1 point for each feature present and −1 for each one absent. (E.g., 1x10, has 2, −1, and 0 features at distances 1, 2, and 3, respectively; x011 has −1, 1, and 1 features at the three distances; and so on). In Fig. 15A, the predictions as a function of the number of observed features at distance 1 are collapsed over the numbers of features at distances 2 and 3; predictions for observed features at distance 2 are collapsed over the numbers at distances 1 and 3; and so on. As expected, Fig. 15A indicates that inferences to a characteristic feature should be stronger to the extent that its immediate neighbor(s) in the causal chain are present. But it also demonstrates how the causal Markov condition predicts that features that are not immediate neighbors (those 2 or 3 causal links away) should have no influence, because they are screened off by the immediate neighbors.

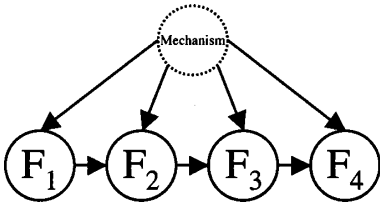
The Underlying Mechanism Model version of the chain network is presented in Fig. 14C, and the predictions of that model are presented in Fig. 15C. Unlike the simple chain model, the Underlying Mechanism Model predicts that features at



(A) Chain Model.



(B) Feature Uncertainty Model for Chain network.



(C) Underlying Mechanism Model for Chain network.

Fig. 14. Alternative chain models.

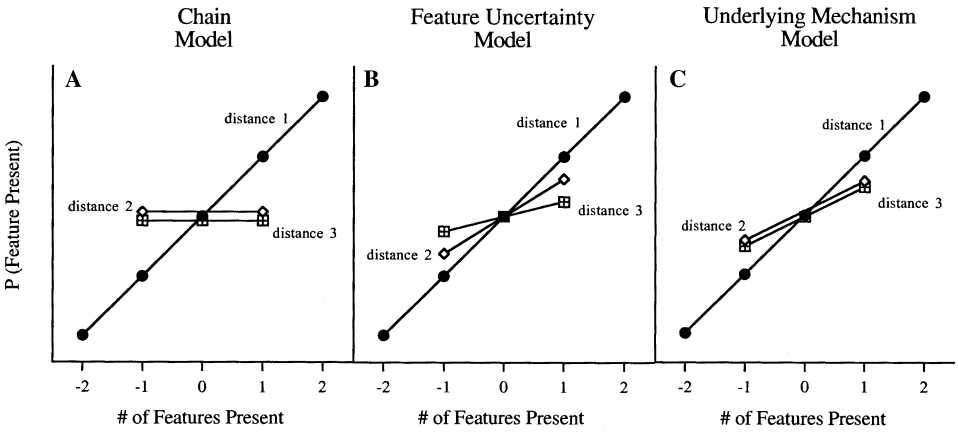


Fig. 15. Normative predictions for Experiment 5.

distances 2 and 3 have predictive value for the unobserved feature (that is, it predicts a nonindependence effect), because of indirect inferences through the underlying mechanism. For example, if dimension 1 is unobserved, then one can infer F_1 not only from F_2, but also from F_3 and F_4 by reasoning to the underlying mechanism and then to F_1. Importantly, the model also predicts that the inferential value of F_3 and F_4 should be the same, because they both involve reasoning through two causal links.

Another advantage of the chain model is that it provides yet one more test of the Underlying Mechanism Model against the Feature Uncertainty Model. The Feature Uncertainty version of the chain network is presented in Fig. 14B, and its predictions are presented in Fig. 15B. In contrast to the Underlying Mechanism Model, the Feature Uncertainty Model predicts that the influence of observed features should be a decreasing function of their distance from the to-be-inferred dimension. For example, if dimension 1 is unobserved, then one can infer F_1 from F'_2 by reasoning across two links (F'_2 to F_2 to F_1), from F'_3 by reasoning across three links (F'_3 to F_3 to F_2 to F_1), and from F'_4 by reasoning across four links (F'_4 to F_4 to F_3 to F_2 to F_1). Thus, F'_2 should have greater influence than F'_3, which in turn should have greater influence than F'_4.

6.1. Method

6.1.1. Materials

The materials used in Experiment 5 were identical to those in Experiments 3 and 4, except that chain condition participants were taught the three causal links that make up a chain network: $F_1 \rightarrow F_2$, $F_2 \rightarrow F_3$, and $F_3 \rightarrow F_4$ (see Appendix A).

6.1.2. Participants

Thirty-six Northwestern undergraduates received course credit for their participation.

6.1.3. Design

Participants were randomly assigned in equal numbers to either the chain or the control condition, and to one of the six categories. The order of the classification and inference tasks was randomized for each participant.

6.1.4. Procedure

The procedure was identical to that in Experiment 1. The 32 possible category members in which one feature is unobserved and each of the other three features is either present or absent were presented (in random order) during the feature inference test.

6.2. Feature inference results

The results are presented in Fig. 16 as a function of the number of features present at each distance from the to-be-inferred dimension. Three aspects of the results in the chain condition should be noted (Fig. 16A). First, as expected, features that were immediate neighbors of the unobserved dimension had a large influence on the inference ratings. (We take the S-shaped pattern of responding in this condition to result from ceiling and floor effects in which one immediate neighbor provides near maximal support for the presence/absence of the unobserved feature.)

Second, features that were two or three causal links away from the to-be-inferred dimension also had an influence. This result represents an (apparent) violation of the causal Markov condition with yet a third causal network topology. However, this result is predicted by both the Underlying Mechanism and Feature Uncertainty Models.

Finally, the critical comparison concerns the influence of features two versus three causal links away. In fact, whereas the Feature Uncertainty Model predicts that the

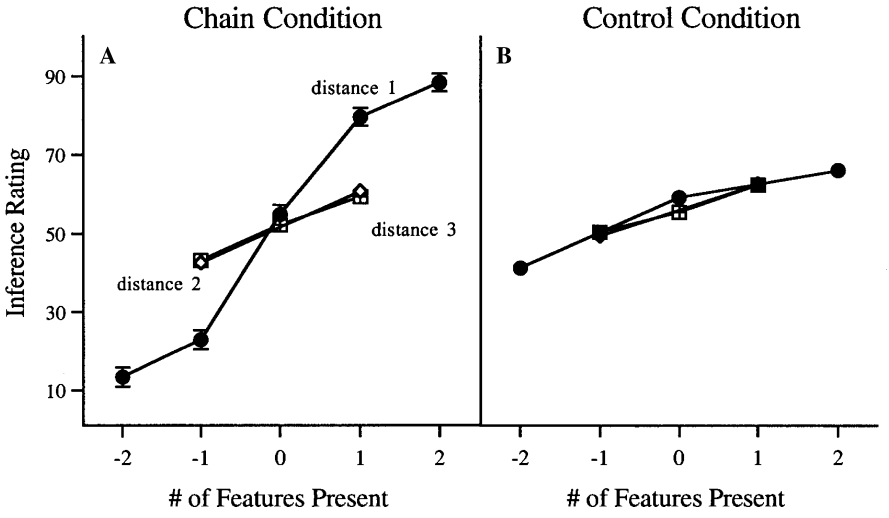


Fig. 16. Feature inference results from Experiment 5.

influence of features at distance 2 should be greater than that at distance 3 (Fig. 15B), participants' judgments were equally influenced by information at those two distances, a result consistent with the Underlying Mechanism Model (Fig. 15C).

Each participants' ratings were predicted with a regression equation with three predictors corresponding to the distances. The regression weight associated with the distance 1 predictor (21.9) was significantly greater than those for distances 2 (9.1) and 3 (8.1), both $ps < .0001$. Moreover, the difference between the distance 2 and distance 3 predictors did not approach significance ($t < 1$). As expected, and consistent with the results of the first four experiments, results in the control condition were a simple function of the overall number of features present (Fig. 16B).

6.3. Discussion

Experiment 5 assessed participants' pattern of feature inferences when features are arranged in a causal chain. As in the first four experiments, inferences appeared to violate the causal Markov condition, because observed features two or three causal links away influenced inferences even when those features were screened off by immediate neighbors. However, the results were consistent with the predictions of the Underlying Mechanism Model in which reasoners assume that features are linked by underlying causal relationships. Thus, the findings from the current experiment mean that the predictions of the Underlying Mechanism Model have been confirmed in five experiments testing three network topologies with six sets of materials drawn from three distinct domains (and one set of blank materials in which the domain was unspecified).

The results of Experiment 5 were also inconsistent with the predictions of the Feature Uncertainty Model, because the influence of features two causal links away from the to-be-inferred dimension was the same as that of features three links away. This finding represents another failure of the Feature Uncertainty Model, which was also unable to account for the results with a common effect network (Experiment 4), or the presence of a nonindependence effect in any of our experiment's control conditions. That is, of the models we have considered, only the Underlying Mechanism Model is able to uniquely account for all the experimental results.

7. General discussion

The purpose of this article was to discover how theoretical category knowledge—specifically, knowledge of the causal relations that link the features of categories—supports one of the primary functions of our conceptual system, namely, the ability to go beyond the given and make inferences about the unobserved. As a starting hypothesis we adopted the Bayesian network view of mental representation of causal knowledge, one which affords a wealth of predictions regarding how the observed features of an exemplar support the inference of other, currently unobserved (or unobservable) features. In the following section we summarize our findings regarding causal knowledge and the inference of unobserved features. We then consider the nature of the Underlying Mechanism Model, its generality, and how it comes to be

acquired. We close by discussing the implications our findings have for psychological essentialism, category-based property induction, and causal reasoning more generally.

7.1. Summary of causal-based feature inference

To start, we think that it is important to stress that the straightforward causal reasoning theory represented by Bayes' nets generally predicted participants' feature inferences quite well. First, of course, participants were much more willing to infer the presence of a feature to the extent that its causes and/or its effects were also present. These results contribute to a collection of findings demonstrating the importance of theoretical or explanatory knowledge in a variety of feature inference tasks. For example, Lassaline (1996) found that the projection of a new property from one category to another was stronger when causal knowledge supportive of that property was provided (also see Sloman, 1994). Rehder and Ross (2001) found that the learning of a category via a feature inference task proceeded more rapidly when features were related on the basis of prior knowledge. However, so far as we know, the current study is the first to address the specific role of causal knowledge in inferring the presence of an unobserved feature on a known stimulus dimension.

By itself, of course, the finding of stronger inferences between causally related features may be unsurprising, as virtually any model that represented interfeature semantic relations in some form would predict that the presence (absence) of one feature would imply the presence (absence) of the other. However, participants' inferences exhibited some of the subtler patterns of inference predicted by Bayes' nets. For example, whereas we found a discounting effect in the common effect network of Experiment 4 in which inferences to a cause were *weaker* to the extent that other causes of the common effect were present, the analogous inferences with a common cause network (to an effect when the common cause was present) were stronger when other effects were already present. These results add to the collection of findings that people are sensitive to the asymmetries inherent in causal relations, and how those asymmetries manifest themselves with networks with multiple variables. For example, Waldmann et al. (1995) found that the ease of learning a category using a supervised learning paradigm depended on whether the correlational structure of the training examples matched the causal structure that the participants were led to believe that the category possessed (either a common-cause or a common-effect structure). Similarly, Rehder (2003a) found that common-cause and common-effect structures manifested asymmetries in how people classify new exemplars.

Taken together, these successes suggest that Bayes' nets provide a useful framework within which to understand people's inferences in light of their causal beliefs. Nevertheless, our experiments also appeared to produce a robust violation of a key property of Bayes' nets, the causal Markov condition. According to this condition, variables should be treated as conditionally independent when they are "screened off" from one another. Instead, we found a nonindependence effect whereby inferences to characteristic features were stronger to the extent that the category member already possessed other characteristic features.

Stated informally, we think what is going on is this. When people see an exemplar they first classify it into the category with which its combination of features is most consistent. After classification, however, reasoners limit their attention to only those aspects of the exemplar that are directly relevant to the inference task. Of course, this includes any interfeature causal relations involving the unobserved dimension that the reasoner may have explicit knowledge of (e.g., one predicts wings given flight and vice versa). But in addition, to the extent that the exemplar has most or all of the category's characteristic features, it will also be considered a *well functioning* category member. That is, the many characteristic features are taken as a sign that the exemplar's underlying causal mechanisms functioned (and/or are continuing to function) properly or normally for members of that kind. And if the exemplar's underlying mechanisms are operating normally, then they are likely to have produced a characteristic value on the unobserved dimension. Conversely, when an exemplar has several uncharacteristic features, it is a sign that the causal mechanisms have operated (are operating) in some way that is unusual and unexpected for members of that kind, and thus may have produced an unusual value on the unobserved dimension.

We formalized this explanation in terms of the Underlying Mechanism Model in which the apparent violations of the causal Markov condition are understood in terms of correct reasoning via an underlying mechanism that links all category features. Importantly, this reasoning apparently does not depend on concrete knowledge regarding the nature of the category's underlying mechanism. We considered the possibility that participants augmented their causal models of categories on the basis of knowledge associated with particular domains, such as the causal mechanisms that govern biological kinds, that produce artifacts, and that lead to the formation of nonbiological natural kinds. But the finding of a nonindependence effect even when the kind of the category was left unspecified (in Experiment 2) indicates that this reasoning does not depend on domain-specific knowledge like this. Instead, we take the finding of a nonindependence effect even for "blank" categories as evidence that (a) people have a domain-general bias to assume the presence of an underlying mechanism even without knowing what that mechanism is, and (b) despite the schematic or skeletal nature of this knowledge, it is sufficient to lead reasoners to infer characteristic features from other characteristic features.

7.2. The nature of knowledge of underlying mechanism

Although we believe our studies have made considerable progress in identifying the explanation for the nonindependence effect, a number of open questions remain. One concerns the exact structure of the underlying mechanism that links observable features. According to the Underlying Mechanism Model, people assume that categories possess a single underlying mechanism that varies in how well it functions, producing as a result either many or few characteristic features. But there are networks with hidden mechanisms with more complex topologies that would produce the same nonindependence effect. For example, features might be related via a hierarchy of underlying mechanisms rather than just one. Or, features might be causally linked to one another directly rather than indirectly. One proposal related to this lat-

ter idea is that natural kinds consist of *homeostatic feature clusters* (Boyd, 1999; Kornblith, 1993) in which members of kinds exist because their features work to maintain one another via mutual causal support. Construed as a mental representation (rather than ontology) of any category (not just natural kinds), homeostatic feature clusters would produce a nonindependence effect, because a characteristic feature on an unobserved dimension would be more likely to the extent that many other characteristic features were present to produce it. Finally, people may be committed to the idea that features causally linked, but their knowledge may be too vague and ill-formed to be expressed in terms of any specific network topology.

Another open question concerns the nature of the cognitive processes that give rise to the nonindependence effect. Our presentation of the Underlying Mechanism Model assumed that our participants' category knowledge included a representation of underlying mechanism plus the interfeature causal relations that we provided, and that they then engaged in explicit causal reasoning with this causal model to predict an unobserved feature. But another possibility is that people's feature inferences are a result of two processes: causal reasoning (which respects independence) and another which encodes a general expectation of within-category correlations. An important consequence of the Underlying Mechanism Model is that, because the number of characteristic features displayed by exemplars will vary depending on how well their underlying mechanisms are (or have been) functioning, these features will be correlated with one another *within* the category, that is, even when only members of the kind are considered. A cognitive process that encodes this expectation would lead people to infer characteristic features on the basis of other characteristic features (and to do so even when within-category correlations are in fact absent in a given category's observed data, Yamauchi & Markman, 2000).

7.3. The generality of the bias toward underlying mechanism

Another important issue concerns the generality of the bias to view category features as related by underlying mechanisms. Our conclusion that this bias is universal was based on the finding of a nonindependence effect for not only all three of the category types we tested (biological kinds, nonliving natural kinds, and artifacts), but also when the kind of the category was left unspecified (in Experiment 2). But we can envisage a couple of ways in which the assumption of underlying mechanism may be less general than we have suggested. One possibility is that in Experiment 2 participants exhibited a nonindependence effect not because they applied an abstract and schematic Underlying Mechanism Model, but rather because they made use of specific domain knowledge via analogy. For example, participants may have assumed that the experimental category ("Daxes") was some sort of biological kind, and thus assumed that Daxes had the underlying causal mechanism associated with those kinds. Another possibility is that the assumption of underlying mechanism, while abstract, is restricted to certain general classes of categories, such as spatiotemporally bound objects (and that Daxes were conceived of as, if not specifically biological, at least some kind of object). However, studies conducted in the first author's laboratory provide evidence that the nonindependence effect obtains for cat-

egories that are not spatio-temporally bound objects, and for which the use of analogy is unlikely. Undergraduate participants were instructed on a novel type of economic system, weather system, or society whose features formed a common-cause network, and violations of screening off were found: Effect features led participants to infer the presence of other effect features even when the common cause was observed (Roofeh, 2003). Apparently, just like in the current experiments, participants assumed that the features were causally linked such that they provided each other mutual inferential support. Moreover, because we think it is unlikely that these undergraduates know much about the underlying causal principles that govern the domains of economics, meteorology, and sociology (or that they conceived of these systems via analogy as biological), it seems that the only possibility is that they assumed that features were causally linked even without knowing what the mechanisms might be. Indeed, when combined with those of the current article, these additional findings mean that a nonindependence effect has been found in nine novel categories drawn from six distinct domains, a result which provides strong evidence for the claim that the assumption of underlying mechanism holds across at least a wide range of category types.

Still, additional research will be required to determine whether the nonindependence effect and the assumption of underlying mechanism hold for the full panoply of category types that have been investigated, including social categories (Fiske, 1998), scripts (Schank & Abelson, 1977), mental events (Rips & Conrad, 1998), ad hoc or goal-based categories (Barsalou, 1983), and relational categories (Gentner, 1981). Our prediction is that the nonindependence effect is likely to hold for categories that exhibit a family resemblance structure in which category members tend to be very similar on the basis of shared features (social categories, scripts, mental events, etc.) because this structure may serve as a cue that invokes the assumption of underlying mechanism. On the other hand, it is less clear how this assumption applies to ad hoc and relational categories whose members are usually quite dissimilar (e.g., members of the goal-based category ''things to take out of the house in case of a fire,'' such as babies and family photographs). An underlying mechanism that leads one to infer characteristic features from other characteristic features is less applicable to a category with no (or few) characteristic features.

7.4. Acquiring knowledge of underlying mechanisms

Another important question concerns how the Underlying Mechanism Model is acquired in the first place. One possibility is that it is learned through experience. For example, we have already suggested that people's representation of biological kinds, artifacts, and nonliving natural kinds may include specific beliefs about the underlying mechanisms that bring rise to the features of such kinds. These beliefs may originate with first-hand experience with categories of these types, and/or through formal instruction. Then, noting the prevalence of underlying mechanism associated with different kinds of categories, people may generalize that all (or at least many) types of categories possess underlying mechanisms. Alternatively, or in addition, people may observe the presence of within-category correlations for

many categories, and then generalize this expectation to most new categories. Either way, the result is a nonindependence effect for many categories that people encounter.

Another possibility, of course, is that these generalizations may not be learned but rather may have been programmed into the human cognitive architecture by evolution. Consistent with this idea is research showing that even young children have beliefs regarding the underlying mechanisms associated with kinds—at least for some types of categories. For example, three-year-olds will attribute the autonomous movement of animals to something intrinsic to the animal itself (Gelman & Gottfried, 1996). Similarly, Inagaki and Hatano (1993) have found evidence that young children are *vitalists* who assume that biological kinds possess some sort of vital force or energy. But whereas this research emphasizes the special beliefs associated with living kinds, our findings suggest that, at least for adults, a belief in underlying mechanism (or an expectation of within-category correlations) holds for many kinds in addition to biological organisms (i.e., artifacts, nonliving natural kinds, economies, weather systems, and so on). Our own guess is that children would exhibit a nonindependence effect every bit as domain-general as adults, but additional research will be required to confirm this suspicion.

A final issue concerns how a schematic or skeletal representation of underlying mechanism might interact with new knowledge that a person acquires (e.g., through formal education) about a particular kind's underlying mechanism. On the one hand, the new knowledge might be simply added to the learner's existing causal model for the category, leaving the original schematic representation in place. Another possibility is that, because the schematic representation functions as a kind of "mechanism placeholder" (Ahn, Kalish, Medin, & Gelman, 1995), it gets "filled in" or replaced by the new knowledge. This latter possibility predicts, for example, the absence of a nonindependence effect when reasoners possess a wealth of knowledge of underlying mechanisms, and knowledge of the state of those mechanisms for a particular inference problem.

7.5. Relationship to psychological essentialism

Our claim that people assume the presence of underlying mechanism in categories is related to the view known as *psychological essentialism* (Gelman, 2003; Medin & Ortony, 1989). According to essentialism, people view categories as being organized around underlying properties (essences) that are shared by all category members and by members of no other categories. Like our Underlying Mechanism Model, essentialism invokes underlying causes or mechanisms responsible for generating observable features. And like that model, it also claims that people often have little specific knowledge about the nature of the essence or the means by which it produces observed features (possessing instead an "essence placeholder").

The important difference between the two theories is that they explain different kinds of uniformity and variation. Essences give a reasoner a way of understanding why members of a category differ from members of other categories. That is, essentialism explains between-category variation. Because an essence is understood to be

present and absolute in all members, essentialism does not predict or explain systematic within-category variation in observable features. Underlying mechanisms, in contrast, are understood to vary in well-functioningness among category members, and the model therefore predicts within-category clustering of observable features.

We summarize the complementary relation we envision between the two theories with the causal network presented in Fig. 17. In this network, the essence node is a binary variable that encodes an object's status as a category member. This binary representation captures the key intuition behind essentialism that category membership is all-or-none. The mechanism node, in contrast, is the pathway by which the essence causes the observable features. It is a continuous variable which represents how well the underlying mechanisms associated with that category are, or have been, functioning for the current object. This structure—which we claim is the default causal structure that gets superimposed on people's mental representation of most categories—has complementary implications for observable features. The essence, absolute in all category members, produces within-category uniformity by enabling the causal mechanisms associated with a particular category. By functioning either well or poorly, those causal mechanisms, in turn, produce within-category variability in the form of within-category feature clustering—and the nonindependence effect.

Since Rosch it has been generally accepted that the features of objects appear in clusters which reflect the underlying causal regularities that govern the world. We believe, however, that causal regularities generate more observable structure than just feature clusters. Because the causal histories of members of a kind operate more or less successfully the features of that cluster will be correlated with one another *within* the cluster. And, one way or another—either through explicit causal reasoning or an expectation of within-category correlations, either with innately provided knowledge structures or ones acquired through experience—people have internalized this fact, using characteristic features as a sign that still more characteristic features are present.

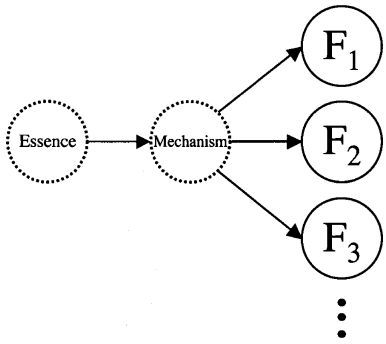


Fig. 17. Integrating psychological essentialism and the underlying-mechanism model.

7.6. Relationship to causal reasoning

Although the present experiments have investigated how people reason with causal knowledge in the context of categories, this leaves open the question of how people reason causally more generally, that is, with variables that are not construed as features of categories. On the one hand, the assumption of an underlying mechanism, like the assumption of an essence, seems more reasonable for members of a natural category—which often do share deep, internal similarities—than for instances of other kinds of causal networks. On the other hand, it is at least conceivable that the nonindependence effect (and, by hypothesis, the assumption of an unobserved common cause like an underlying mechanism) is characteristic of reasoning about any set of variables related in a causal network (e.g., see Cartwright, 1993, for arguments related to this idea; and Hausman & Woodward, 1999, for a response). This is, of course, an empirical question.

Still, there is reason to think that category-ness is important to the nonindependence effect. One important piece of evidence is the fact that features were predictive of one another in the control condition of each of our five experiments, despite the fact those features were not described as causally related. That is, the nonindependence effect is induced by mere knowledge that variables are features of a category. Moreover, the magnitude of violations of screening off in, say, Experiment 1 was no greater when features were related in a common cause network than when they were unrelated (i.e., in the control condition, see Fig. 3), suggesting that, as predicted by the Underlying Mechanism Model, membership in an explicit causal network does not make an additional contribution to feature nonindependence above and beyond that which obtains for unrelated category features.

7.7. Relationship to causal-based categorization

As discussed earlier, in addition to feature inferences, the Bayes' net approach has also been successfully applied to how people make classification decisions in light of causal knowledge about categories. Rehder (2003a, 2003b) has proposed that people judge category membership on the basis of whether the particular combination of features displayed by an object was likely to have been *generated* by the category's causal model. This account has been shown to explain why categorizers are sensitive to inter-feature consistency (e.g., cause and effect features both present or both absent), and also to features that have many causes (i.e., a common effect in a common effect network) because such features are especially likely to have been generated. Moreover, as in the current research, to provide a full explanation of the categorization results, it was assumed that participants augmented the category's causal model with an assumption of underlying mechanism. For example, to account for the *causal status effect* in which features that occur earlier in a causal network have more influence on categorization decisions (Ahn, 1998; Ahn et al., 2000), Rehder (2003b) assumed that root features (i.e., those for which no explicit causes are provided) were generated by the category's underlying mechanism. However, this assumption is somewhat different from the one embodied in the Underlying Mechanism Model which assumes that

all (not just root) features are causally related to the underlying mechanism. One way to resolve this discrepancy would be to assume that nonroot features are less strongly related to the underlying mechanism, just because an explicit cause has been provided for them, a proposal which would provide a uniform explanation for a wide range of judgments of both categorization and feature inference.

7.8. Relationship to category-based property induction

Finally, considerations of underlying causal mechanism have also become important in research on category-based property induction. On the one hand, research has shown that the extent to which a new property is projected to a category depends on the causal mechanisms thought to have produced that property. For example, Proffitt, Coley, and Medin (2000) found that tree experts generalized diseases on the basis of their knowledge of the mechanisms by which diseases can be transmitted among trees (also see Lassaline, 1996). That is, as in our experiments, causal knowledge is used powerfully when available. In the absence of causal knowledge, however, a typicality effect is found in which more typical category members like robins support stronger generalizations about birds than do atypical members like ostriches (Osherson, Smith, Wilkie, López, & Shafir, 1990; Rips, 1975; Sloman, 1993). We think that this effect, like our nonindependence effect, can be understood in terms of causal reasoning with underlying causal mechanisms. We suggest that reasoners prefer more typical premises (robins) because they are assumed to manifest the underlying causal mechanisms that are characteristic of the category (birds). Once a novel feature is attributed to a category's normal causal mechanisms, it is judged likely to be present in all category members.

8. Conclusion

Our research has established two facts regarding how people infer unobserved features of category members. The first of these is that feature inference is strongly influenced by the presence of explicit interfeature causal knowledge. Our results show that this influence reflects the asymmetry of causal relationships, and is consistent with a normative view of causal reasoning as defined by Bayes' nets.

The second finding is of a nonindependence effect in which the presence of characteristic features implies the presence of other characteristic features. We have argued that this effect arises because people take characteristic features as diagnostic of the normal operation of underlying causal mechanisms. As a result, when a category member is discrepant in its observed features, that discrepancy is taken as a sign that something has gone awry with that exemplar's normal mechanisms and hence is likely to possess uncharacteristic features. We suggest that the domain-general assumption of causal mechanisms is critical to understanding not only how people infer features, but also their categorization decisions, their propensity to

generalize new properties, and, we suspect, their performance in a wide variety of category-related tasks.

Appendix A. Materials

Description of the cover story, attributes, attribute values, causal relationships, and blank properties for each of the six categories is presented below.

A.1. Kehoe Ants

On the volcanic island of Kehoe, in the western Pacific Ocean near Guam, there is a species of ant called Kehoe Ants. For food, Kehoe Ants consume vegetation rich in iron and sulfur.

A.1.1. Features

(F_1) Some Kehoe Ants have blood that is very high in iron sulfate. Others have blood that has low levels of iron sulfate.

(F_2) Some Kehoe Ants have an immune system that is hyperactive. Others have a suppressed immune system.

(F_3) Some Kehoe Ants have blood that is very thick. Others have blood that is very thin.

(F_4) Kehoe Ants build their nests by secreting a sticky fluid that then hardens. Some Kehoe Ants are able to build their nests quickly. Others build their nests slowly.

A.1.2. Causal relationships

($F_1 \rightarrow F_2$). Blood high in iron sulfate causes a hyperactive immune system. The iron sulfate molecules are detected as foreign by the immune system, and the immune system is highly active as a result.

($F_1 \rightarrow F_3$). Blood high in iron sulfate causes thick blood. Iron sulfate provides the extra iron that the ant uses to produce extra red blood cells. The extra red blood cells thicken the blood.

($F_1 \rightarrow F_4$). Blood high in iron sulfate causes faster nest building. The iron sulfate stimulates the enzymes responsible for manufacturing the nest-building secretions, and an ant can build its nest faster with more secretions.

($F_2 \rightarrow F_3$). A hyperactive immune system causes thick blood. A hyperactive immune system produces a large number of white blood cells, which results in the blood being thicker.

($F_2 \rightarrow F_4$). A hyperactive immune system causes faster nest building. The ants eliminate toxins through the secretion of the nest-building fluid. A hyperactive immune system accelerates the production of nest-building secretions in order to eliminate toxins.

($F_3 \rightarrow F_4$). Thick blood causes faster nest building. The secreted fluid is manufactured from the ant's blood, and thicker blood means thicker secretions. Thicker secretions mean that each new section of the nest can be built with fewer application of the fluid, increasing the overall rate of nest building.

A.2. Lake Victoria Shrimp

Lake Victoria Shrimp are found in Lake Victoria, Africa. The concentration of algae that are rich in choline is unusually high in some parts of Lake Victoria.

A.2.1. Features

(F_1) Lake Victoria Shrimp use acetylcholine (ACh) as a brain neurotransmitter. Some Lake Victoria Shrimp have an unusually high amount of ACh. Others have a unusually low amount of ACh.

(F_2) Lake Victoria Shrimp have an a "flight" response in which they flee from potential predators. The flight response consists of an electrical signal sent to the muscles which propel the shrimp away from a predator. Some Lake Victoria Shrimp have a flight response which is long-lasting. Others have a short flight response.

(F_3) Some Lake Victoria Shrimp have an accelerated sleep cycle (4 hours sleep, 4 hours awake). Others have a decelerated sleep cycle (12 hours sleep, 12 hours awake).

(F_4) Some Lake Victoria Shrimp have high body weight. Others have a low body weight.

A.2.2. Causal relationships

($F_1 \rightarrow F_2$). A high quantity of ACh neurotransmitter causes a long-lasting flight response. The duration of the electrical signal to the muscles is longer because of the excess amount of neurotransmitter.

($F_1 \rightarrow F_3$). A high quantity of ACh neurotransmitter causes an accelerated sleep cycle. The neurotransmitter speeds up all neural activity, including the internal "clock" which puts the shrimp to sleep on a regular cycle.

($F_1 \rightarrow F_4$). A high quantity of ACh neurotransmitter causes a high body weight. The neurotransmitter stimulates greater feeding behavior, which results in more food ingestion and more body weight.

($F_2 \rightarrow F_3$). A long-lasting flight response causes an accelerated sleep cycle. The long-lasting flight response causes the muscles to be fatigued, and this fatigue triggers the shrimp's sleep center.

($F_2 \rightarrow F_4$). A long-lasting flight response causes a high body weight. The shrimp are propelled over a greater area of the lake, and find more new food sources as a result.

($F_3 \rightarrow F_4$). An accelerated sleep cycle causes a high body weight. Shrimp habitually feed after waking, and shrimp on an accelerated sleep cycle wake three times a day instead of once.

A.3. Myastars

In certain parts of the known universe there exists a large number of stars called Myastars. Myastars are formed from clouds of helium.

A.3.1. Features

(F_1) Some Myastars are constructed from ionized helium. Others are constructed from normal helium.

(F_2) Some Myastars are very hot. Others have a low temperature.

(F_3) Some Myastars are extremely dense. Others have low density.

(F_4) Some Myastars have a large number of planets. Others have a small number of planets.

A.3.2. Causal relationships

($F_1 \rightarrow F_2$). Ionized helium causes the star to be very hot. Ionized helium participates in nuclear reactions that release more energy than the nuclear reactions of normal hydrogen-based stars, and the star is hotter as a result.

($F_1 \rightarrow F_3$). Ionized helium causes the star to have high density. Ionized helium is stripped of electrons, and helium nuclei without surrounding electrons can be packed together more tightly.

($F_1 \rightarrow F_4$). Ionized helium causes the star to have a large number of planets. Because helium is a heavier element than hydrogen, a star based on helium produces a greater quantity of the heavier elements necessary for planet formation (e.g., carbon, iron) than one based on hydrogen.

($F_2 \rightarrow F_3$). A hot temperature causes the star to have high density. At unusually high temperatures heavy elements (such as uranium and plutonium) become ionized (lose their electrons), and the resulting free electrons and nuclei can be packed together more tightly.

($F_2 \rightarrow F_4$). A hot temperature causes the star to have a large number of planets. The heat provides the extra energy required for planets to coalesce from the gas in orbit around the star.

($F_3 \rightarrow F_4$). High density causes the star to have a large number of planets. Helium, which cannot be compressed into a small area, is spun off the star, and serves as the raw material for many planets.

A.4. Meteoric sodium carbonate

A special form of sodium carbonate (Na_2CO_2) is found in meteors that land on earth. Molecules of "meteoric" sodium carbonate differ from molecules of normal sodium carbonate that are found on earth in that they have been exposed to intense X-rays in space.

A.4.1. Features

(F_1) Some meteoric sodium carbonate molecules are radioactive, i.e., theta particles get emitted from the nuclei of the sodium (Na) atoms. Other meteoric sodium carbonate molecules are nonradioactive.

(F_2) Some molecules of meteoric sodium carbonate have their five atoms arranged in an eight-bond pyramid (four atoms at the base of the pyramid, and one at the "peak"). Other molecules of meteoric sodium carbonate have their five atoms arranged in a normal five-bond ring, as in normal sodium carbonate found on earth.

(F_3) Some molecules of meteoric sodium carbonate are positively charged. Others have a negative charge.

(F_4) Some molecules of meteoric sodium carbonate are very reactive (tend to enter into chemical reactions). Others have a low level of reactivity.

A.4.2. Causal relationships

($F_1 \rightarrow F_2$). Radioactivity causes the molecule to take on a pyramid structure. Theta particles provide the extra energy required to form the additional atom-to-atom bonds required for the pyramid.

($F_1 \rightarrow F_3$). Radioactivity causes the molecule to have a positive charge. Theta particles are negatively charged, and so leave the molecule with a positive charge after they are emitted.

($F_2 \rightarrow F_3$). The pyramid structure causes the molecule to have a positive charge. Atoms are packed close together in the pyramid structure, and so are able to share electrons. Because they can be shared, the molecule has fewer negatively charged electrons, and hence the molecule has an overall positive charge.

($F_1 \rightarrow F_4$). Radioactivity causes the molecule to be reactive. Theta particles break up surrounding molecules and hence accelerate the natural rate of chemical reactions.

($F_2 \rightarrow F_4$). The pyramid structure causes the molecule to be reactive. Once one atom of the pyramid is involved in a chemical reaction, the remaining atoms break apart, providing the raw material for further reactions.

($F_3 \rightarrow F_4$). Having a positive charge causes the molecule to be reactive. The molecule attracts negatively charged subparts of other molecules, which breaks up the other molecules, and causes chemical reactions.

A.5. Romanian Rogos

The Romanian Motor Company, located in Bucharest, Romania, manufactures an automobile called a Rogo which is designed to run on fuel refined locally in Romania. Depending on where it is refined, the fuel may or may not have butane (C_4H_{10}), a naturally occurring hydrocarbon, blended in with the gasoline.

A.5.1. Features

(F_1) Some Rogos are filled with gasoline laden with butane. Other Rogos are filled with gasoline with no butane.

(F_2) The fuel filters of Rogos have gaskets. Some Rogos have fuel filter gaskets that are extra loose. Other have tight fuel filter gaskets.

(F_3) Some Rogos have a hot engine temperature. Others have a low engine temperature.

(F_4) Some Rogos have a high amount of carbon monoxide in their exhaust. Others have a low amount of carbon monoxide in their exhaust.

A.5.2. Causal relationships

($F_1 \rightarrow F_2$). Butane-laden fuel causes loose fuel filter gaskets. The butane tends to corrode the rubber out of which gaskets are made, and so the gaskets do not fit tightly.

($F_1 \rightarrow F_3$). Butane-laden fuel causes hot engine temperature. The butane in the fuel burns at a hotter temperature than normal gasoline.

($F_1 \rightarrow F_4$). Butane-laden fuel causes high amounts of carbon monoxide in the exhaust. Butane contains more carbon than normal gasoline, and so more carbon is available to bind with oxygen to form carbon monoxide.

($F_2 \rightarrow F_3$). A loose fuel filter gasket causes hot engine temperature. Loose gaskets allow more air to be mixed in with the fuel, meaning that the gas is more fully burned, resulting in the engine running hotter than normal.

($F_2 \rightarrow F_4$). A loose fuel filter gasket causes high amounts of carbon monoxide in the exhaust. Loose fuel filters allow more air into the gas–air mixture, providing the oxygen which binds with carbon to form carbon monoxide.

($F_3 \rightarrow F_4$). Hot engine temperature causes high amounts of carbon monoxide in the exhaust. The heat provides the energy required for the carbon to bind with the oxygen.

A.6. Neptune Military Personal Computers

The power supplies for the Neptune Military Personal Computers are made from tungsten mined in southern Utah, some samples of which are magnetic.

A.6.1. Features

(F_1) Some Neptune Personal Computers have a power supply that is magnetic and extends a magnetic field. Others have a normal nonmagnetic power supply that extends no magnetic field.

(F_2) Neptune Personal Computers have an internal clock based on a crystal oscillator that determines how fast the computer runs. Some Neptune Personal Computers have a clock speed that is too fast. Others have a slow clock speed.

(F_3) Some Neptune Personal Computers run at an unusually high temperature. Others run at a low temperature.

(F_4) Some Neptune Personal Computers have a screen image that is unusually bright. Other have a screen image that is unusually dim.

A.6.2. Causal relationships

($F_1 \rightarrow F_2$). Magnetic power supplies cause the computer to have a fast clock speed. The magnetic field interferes with the natural phase transitions of the crystal oscillator, the result being that the crystal oscillator emits square waves at a faster rate.

($F_1 \rightarrow F_3$). Magnetic power supplies cause the computer to run at a hot temperature. The magnetic field influences the copper atoms in electrical wire to orient themselves perpendicularly to the flow of electricity, increasing the resistance, and resulting in more heat being generated.

($F_1 \rightarrow F_4$). Magnetic power supplies cause the computer to display a bright image. The magnetic field concentrates the electron beam which strikes the phosphor on the computer screen, leading to an image that is slightly smaller but brighter.

($F_2 \rightarrow F_3$). A fast clock speed causes the computer to run at a hot temperature. With a faster clock speed the computer runs faster, performs more operations, and generates more heat as a result.

($F_2 \rightarrow F_4$). A fast clock speed causes the computer to display a bright image. The clock controls how fast the image is "repainted" on the screen. A faster clock means that the phosphors on the screen's surface are being irradiated with electrons more often, leading to a brighter image.

($F_3 \rightarrow F_4$). Hot temperature causes the computer to display a bright image. Heat increases the efficiency of the cathode ray tube, leading to a more energized electron beam and a brighter screen.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition, 69*, 135–178.

Ahn, W., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16*, 81–121.

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*, 299–352.

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*, 361–416.

Ahn, W., Marsh, J. K., Luhmann, C. C., & Lee, K. (2002). Effect of theory-based feature correlations on typicality judgments. *Memory & Cognition, 30*, 107–118.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13*, 263–308.

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

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

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

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 141–185). Cambridge: Cambridge Universtiy Press.

Burnett, R. C., Medin, D. L., Ross, N. O., & Blok, S. V. (in press). Ideal is typical. *Canadian Journal of Experimental Psychology*.

Cartwright, N. (1993). Marks and probabilities: Two ways to find causal structure. In F. Stadler (Ed.), *Scientific philosophy: Origins and development*. Dordrecht: Kluwer.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of Social Psychology* (pp. 357–411). Boston MA: McGraw-Hill.

Gelman, S. A. (2003). *The essential child: The origins of essentialism in everyday thought*. New York: Oxford University Press.

Gelman, S. A., & Gottfried, G. (1996). Causal explanations of animate and inanimate motion. *Child Development, 67*, 1970–1987.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory, 4*, 161–178.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 3–32.

Hadjichristidis, C., Sloman, S. A., Stevenson, R., & Over, D. (2004). Feature centrality and property induction. *Cognitive Science, 28*, 45–74.

Hausman, D. M., & Woodward, J. (1999). Independence, invariance and the Causal Markov Condition. *British Journal for the Philosophy of Science, 50*, 521–583.

Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind-body distinction. *Child Development, 64*, 1534–1549.

Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary approach* (pp. 234–262). Oxford: Clarendon Press.

Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. Cambridge, MA: MIT Press.

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

Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology, 40*, 87–137.

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

Malt, B. C., Ross, B. H., & Murphy, G. L. (1995). Predicting features for members of natural categories when categorization is uncertain. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 646–661.

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

Medin, D. L. (1983). Structural principles in categorization. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analyses* (pp. 203–230). Hillsdale, NJ: Erlbaum.

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.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242–279.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10*, 517–532.

Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review, 102*, 331–355.

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology, 27*, 148–193.

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

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: 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., & 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., & Ross, B. H. (2001). Abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1261–1275.

Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*, 665–681.

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., & Conrad, F. G. (1989). Folk psychology of mental activities. *Psychological Review, 96*, 187–207.

Roofeh, D. (2003). *The effect of attribute prediction on nonobject categories with causal knowledge*. Unpublished Undergraduate Honors thesis.

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

Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 736–753.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.

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 judgements of likelihood. *Cognition, 52*, 1–21.

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189–228.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

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

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

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 776–795.