

Causal Status as a Determinant of Feature Centrality

Woo-kyoung Ahn and Nancy S. Kim

Yale University

Mary E. Lassaline

University of Illinois at Urbana-Champaign

and

Martin J. Dennis

Yale University

One of the major problems in categorization research is the lack of systematic ways of constraining feature weights. We propose one method of operationalizing feature centrality, a causal status hypothesis which states that a cause feature is judged to be more central than its effect feature in categorization. In Experiment 1, participants learned a novel category with three characteristic features that were causally related into a single causal chain and judged the likelihood that new objects belong to the category. Likelihood ratings for items missing the most fundamental cause were lower than those for items missing the intermediate cause, which in turn were lower than those for items missing the terminal effect. The causal status effect was also obtained in goodness-of-exemplar judgments (Experiment 2) and in free-sorting tasks (Experiment 3), but it was weaker in similarity judgments than in categorization judgments (Experiment 4). Experiment 5 shows that the size of the

We thank Susan Carey, Marvin Chun, Susan Gelman, Arthur Markman, Douglas Medin, Steven Sloman, and an anonymous reviewer for their helpful comments on the earlier draft of this article. We also thank Denise Hatton, Tisha Baldwin, Joshua Nathan, Helen Sullivan, and Julia Wenzlaff for collecting data. Some of the stimulus materials used in Experiments 1 and 2 are adapted from the stimulus materials used in Rehder and Hastie (1997) and we thank them for inspiring many of the features and objects used in these studies. This project was supported by a National Science Foundation Grant (NSF-SBR 9515085) and a National Institute of Mental Health Grant (RO1 MH57737) given to Woo-kyoung Ahn, a National Science Foundation Graduate Fellowship to Nancy Kim, and a National Institute of Mental Health Postdoctoral Fellowship (MH10888-01A1) to Mary Lassaline.

Address correspondence and reprint requests to Woo-kyoung Ahn, Department of Psychology, Vanderbilt University, Wilson Hall, Nashville, TN 37240. E-mail: woo-kyoung.ahn@vanderbilt.edu.



causal status effect is moderated by plausibility of causal relations, and Experiment 6 shows that effect features can be useful in retrieving information about unknown causes. We discuss the scope of the causal status effect and its implications for categorization research. © 2000 Academic Press

Features of objects vary in their importance as a basis for categorization. For example, bird DNA might be much more central and important in our concept of "bird" than the ability to fly. In our concept of dogs, having internal organs seems to matter more than wagging the tail. It is easier to imagine an apple that is not round than to imagine an apple that does not grow from a tree. Why are some features more central than others?

One might argue that some stimuli, such as shape, are inherently salient or central in categorizing rigid artifacts (e.g., Landau, Smith, & Jones, 1988; but see also Gelman & Ebeling, 1998; Soja, Carey, & Spelke, 1991). However, this idea of constraints based on specific dimensions cannot be the complete answer because the centrality of the same feature can vary depending on the category. For instance, the feature *curvedness* is more important for boomerangs than for bananas (Medin & Shoben, 1988).

Moreover, the centrality of features does not seem to depend exclusively on how many members of the same kind have that feature (Rosch & Mervis, 1975), or category validity (the probability that an object has a certain feature given that it belongs to a certain category; e.g., the probability that an object has wings given that it is a bird). For instance, "square" has a category validity of zero for both basketballs and cantaloupes but people are more likely to accept a square cantaloupe than a square basketball (Medin & Shoben, 1988).

Recent similarity-based models of categorization (e.g., Kruschke, 1992; Lamberts, 1995, 1998; Nosofsky, 1984, 1986) pay more attention to the issue of feature weighting because, as pointed out by Murphy and Medin (1985), any two objects in the world can be judged as similar or dissimilar to each other without appropriate constraints on feature weighting. In general, these models have focused on perceptual salience of dimensions and cue validity (or diagnosticity; the probability that an object belongs to a certain category given that it has a certain feature; e.g., the probability that an object is a bird given that it has wings). For instance, Kruschke's ALCOVE (1992) can learn attention strengths (i.e., feature centrality) as a function of diagnosticity of features. Still, diagnosticity or perceptual salience alone cannot account for cases such as the difference between a square cantaloupe and a square basketball because the feature *roundness* is equally diagnostic for and perceptually salient in cantaloupes and basketballs.

As one critical part of the groundwork needed to develop a comprehensive theory of feature centrality, this article investigates one way of constraining feature centrality, namely the causal status of features in a category. We argue that the causal status of features can account for categorization judg-

ments made even in cases in which cue validity and category validity are not helpful. Before presenting the main claim, we first describe related ideas about feature centrality that form a backdrop to the current claim.

Theory-Based Categorization

One answer to the issue of feature weighting is provided by the theory-based categorization view. The idea is that the centrality of any given feature is determined by its importance in the principles or theories underlying the categories (Murphy & Medin, 1985). Categories are connected in many complex ways that resemble structured theories. For example, our concept of “boomerang” is connected with other concepts such as “throwing,” “air,” “speed,” and so on, all of which are intricately connected as in a scientific theory. Features that play an important part in commonsense or naive theories seem more essential in categorization than those that do not. For example, the feature *curvedness* plays a role in a theory of physics, and it thereby becomes a central feature for categorizing objects as boomerangs. In the case of bananas, however, curvedness does not play a critical role in a naive biological theory, and it is therefore not so central.

Even young children’s inductions seem to be influenced by their naive theories (Gelman, 1988; Gelman & Wellman, 1991; Keil, 1989). For example, Keil (1989) presented subjects, ranging from kindergartners to adults, with a description of a raccoon who went through surgery to make it look like a skunk, but still bears live raccoons and was born from a raccoon. In spite of these changes in its surface features, even 2nd graders believed that the animal described is a raccoon. In contrast, kindergartners’ inductions were based on perceptual appearance, presumably because they did not yet have appropriate domain theories. In sum, one way of explaining why some features are more essential than others is to examine each feature’s role in the domain theories.

Although the previous theory-based approaches have all discussed the importance of causal background knowledge, the exact mechanism by which this knowledge influences feature weighting has rarely been articulated (Gelman & Kalish, 1993; Murphy, 1993). That is, previous approaches have not specified precisely how to determine whether a feature plays a central role in one’s domain theory or how this should be computationally implemented. The main goal of the current study is to provide one way to operationalize feature centrality in terms of one’s causal background knowledge and to empirically test this formulation in various contexts.

Causal Status Hypothesis

In accordance with the theory-based view, we assume that concepts consist of richly structured features rather than of a set of independent features (Murphy & Medin, 1985). As argued by Rosch (1978), natural categories seem to consist of correlated features rather than independently occurring features.

For example, the category “chair” has correlated features of “has legs,” “has a seat,” and “can be sat on.” Later, Murphy and Medin (1985) expanded this notion by arguing that “people are not only sensitive to feature correlations, but they can deduce reasons for those correlations, based on their knowledge of the way the world works. Perhaps, then, the connection between those features is not a simple link, but a whole causal explanation for how the two are related” (p. 300). Similarly, we assume that the majority of features of existing concepts are directionally connected, not just correlated, because people have naive theories about how these features are connected (Carey, 1985; Wellman, 1990; see also Ahn, 1998; Kim & Ahn, 1999; Sloman, Love, & Ahn, 1998 for empirical demonstrations of interconnected features in real-life categories). Of the many types of asymmetric, directional relations (e.g., causal, temporal, or physical support relations), the current study focuses only on causal relations because numerous researchers of theory-based categorization have viewed causality as a central component in theorylike conceptual representation (Carey, 1985; Gelman & Kalish, 1993; Wellman, 1990).

Taking this theory-based approach, our main hypothesis is that the position of a feature within a causal structure determines its centrality, the degree to which a feature’s presence or absence affects the likelihood that an object belongs to a certain category. Specifically, our claim is that people regard cause features as more important and essential than effect features in that cause features affect category membership decisions more than do effect features.

Consider a real-life example of this causal status effect. Suppose a clinical graduate student is forming a new concept, “borderline personality disorder,” by learning its typical symptoms, such as “frantic efforts to avoid abandonment,” “identity disturbance,” and “recurrent suicidal behavior.” Following the Diagnostic and Statistical Manual of Mental Disorders (4th ed., American Psychiatric Association, 1994) guidelines, the student might make a diagnosis of borderline personality if a patient displays five of nine criterial symptoms. Note that in this case, it does not matter whether a missing symptom is “suicidal behavior” or “fear of abandonment,” assuming identical cue and category validities for these symptoms. On the other hand, the student might learn a causal structure for the symptoms, such that recurrent suicidal attempts are made *because* of a fear of abandonment. Then, even if a patient does not display “suicidal behavior,” it might not matter much for the diagnosis because this symptom is a mere effect. However, if a cause symptom is missing, as in the case where a patient displays suicidal behavior not in an attempt to avoid abandonment from other people but for some other reason such as depression, then this might substantially lower the likelihood of that patient’s having borderline personality disorder.

Or consider the previous example of a square basketball versus a square cantaloupe, in which the category validities and cue validities of “square-

ness'' are both zero, yet the feature of roundness is more important for basketballs than for cantaloupes (Medin & Shoben, 1988). This puzzle can be explained by considering the causal status of roundness as a feature in the respective categories of basketballs and cantaloupes. In the case of basketballs, roundness enables certain functional features of the ball, such as bouncing with a predictable trajectory and being able to go through a round hoop. In the case of cantaloupes, roundness is not usually considered to enable any other features; that is, if a cantaloupe were square, no other features of it (i.e., taste, texture, or function of protecting the seeds) would be affected. In other words, the feature of roundness is more *causally* central in basketballs than it is in cantaloupes, and according to our causal status hypothesis, that is why this feature is also more *conceptually* central in basketballs than it is in cantaloupes. In this way, the causal status hypothesis may predict how, and in which direction, the same feature can affect categorization judgments differently within different categories. Before introducing the details of our theory, the next section provides a rationale for the causal status effect. This will serve as a basis for some of the predictions we make below.

Why Would People Weigh Causes More Than Effects?

One major reason to expect the causal status effect comes from the literature on psychological essentialism. Essentialism in the purely philosophical sense states that objects have essences that make them the objects they are (Kripke, 1972; Locke, 1894/1975; Putnam, 1977). Whether or not this metaphysical claim is true, Medin and Ortony (1989) proposed that people act as if things have essences, the doctrine called "psychological essentialism." According to this view, essences in one's conceptual representations are believed to generate or constrain surface features of objects (Gelman & Wellman, 1991; Medin & Ortony, 1989). Hence, the causally deepest known properties of entities might be people's best guess as to what essences might be. For instance, the male essence might be male hormones which cause surface features of "male" including certain heights and facial hair. Although surface features might be useful in identifying instances of the concept, "the more central properties are best thought of as constraining or even generating the properties that might turn out to be useful in identification" (Medin & Ortony, 1989, p. 185). Because cause features by definition serve to constrain or generate other properties, those features may fill the role of essences in people's representations of objects, lacking more complete metaphysical knowledge.

We also believe that the causal status effect should occur because of the higher (perceived) predictability of cause features compared to effect features. One enormous advantage to having concepts is that we can infer or predict nonobvious properties (e.g., "is dangerous") based on category membership information (e.g., wolf). Thus, a "good" category is considered to be the one that allows rich inductive inferences (e.g., Anderson, 1990).

In particular, the more that underlying causes are revealed, the more inductive power the concept seems to gain. For instance, discovering a cause of a symptom such as nausea (e.g., Is it caused by bacteria or pregnancy?) allows doctors to determine the proper course of treatment and also to make a better prognosis of the condition (e.g., Will it lead to fever or to a new baby?). In contrast, merely learning the effect of the symptom (e.g., nausea usually causes a person to throw up) does not necessarily help us to come up with a treatment plan. Similarly, understanding the motive of a person's nice behavior (e.g., does he want promotion or is he genuinely nice?) would allow us to predict many more behaviors of the person than discovering the consequence of the person's nice behavior (e.g., people were impressed) would.

In fact, people do seem to believe that a cause feature has more predictive power than an effect feature, whether or not this is indeed the case. For instance, Tversky and Kahneman (1982) report that people feel more confidence in predicting an effect from a cause (e.g., predicting the son's height from the father's height) than predicting a cause from an effect (e.g., predicting the father's height from the son's height) even when the two probabilities, $P(\text{Effect}|\text{Cause})$ and $P(\text{Cause}|\text{Effect})$, should be equal. They also found that people actually give higher estimates of $P(\text{Effect}|\text{Cause})$ than $P(\text{Cause}|\text{Effect})$. Given findings such as these, we speculate that people believe that causal features provide more predictions and that they would therefore give more weight to cause features than to effect features in categorization.

A Previous Empirical Demonstration of the Causal Status Effect

Ahn (1998) presented the initial demonstration of the causal status effect in natural kinds and artifacts. The main purpose of this study was to investigate why some studies (e.g., Barton & Komatsu, 1989; Gelman, 1988) have shown that different features are central for natural kinds and artifacts: in natural kinds internal or molecular features are more conceptually central than functional features, but in artifacts functional features are more conceptually central than internal or molecular features. Ahn (1998) argued that the mechanism underlying this phenomenon is the causal status effect. That is, in natural kinds, internal/molecular features tend to cause functional features (e.g., cow DNA determines whether cows give milk), but in artifacts, functions intended by the designer determine its compositional structure (e.g., chairs are intended to be used for sitting, and for that reason, they are made of a hard substance). Experiments 1 and 2 in Ahn (1998) examined real-life categories used in previous studies (Barton & Komatsu, 1989; Malt & Johnson, 1992). Participants were asked to draw causal relations among features within the same category. At the same time, they judged centrality of features as measured by the degree to which a feature impacts categorization when it is missing. It was found that across natural and artifactual kinds, the more

features any particular feature caused, the more influential the feature was in categorization. In addition, Ahn (1998) directly manipulated the causal status of features using artificial stimuli and showed that when a compositional feature caused a functional feature, a compositional feature was more influential in categorization of both natural and artifactual kinds, whereas the opposite was true when the causal direction was reversed.

Although the basic phenomenon of the causal status effect was demonstrated in Ahn (1998), a full theoretical specification of the causal status hypothesis, as a conceptual and empirical research tool in knowledge-based categorization, has not yet been offered. The main goal of the current study is to provide an articulation of the scope and boundary conditions of the causal status hypothesis, empirical tests of these claims, and discussion of their theoretical implications in the larger context of the concepts literature. In the following section, we discuss the predictions and scope of the causal status hypothesis.

What the Causal Status Hypothesis Does and Does Not Predict

Our specification of the scope of the causal status hypothesis covers a number of issues: when it is manifested, what it predicts when causal depth or the plausibility of causal relations or feature centrality are varied, what the role of effect features is within our framework, and whether the causal status effect is a structural or content-based constraint. We discuss each of these factors in turn below.

Manifestation of the causal status effect. We claim that the causal status effect is deeply rooted in our categorization processes. Thus, we propose that a variety of phenomena involved in conceptual representation or categorization should be sensitive to the causal status of features as a result of its influence on feature centrality.

The most straightforward measure of feature centrality would be the degree to which a feature determines category membership. Following the format used by Medin and Shoben (1988) and Barton and Komatsu (1989), among others, we ask participants in Experiments 1 and 5 to judge the degree to which an object belongs to a target category if it were missing a target feature. This task was successfully used by Ahn (1998) to demonstrate the causal status effect, but it was the only measure of feature centrality in that study. The current study utilizes a number of additional measures to examine other important phenomena.

One of these is the question of what people believe the natural structure of categories to be. As explained earlier, the causal status effect is expected to occur partly because of people's belief that categories have essences. As in Medin and Ortony (1989), we assume that essences are believed to be the deepest cause of an entity's properties. If this is true, we would expect people to create categories based on matching causes rather than matching effects when asked to free-sort objects in any way that looks natural to them (e.g.,

Imai & Garner, 1965). Experiment 3 examines this issue using free-sorting tasks.

In addition, we propose that the graded structure of categories is also determined by features' causal status. One of the most well-established phenomena in categorization research is that members of a category vary in how good an example they are of their category (e.g., Barsalou, 1985; Rosch & Mervis, 1975; Smith, Shoben, & Rips, 1974). What determines the goodness of exemplars? Rosch and Mervis (1975) used category validity of features ("the number of items in a category that had been credited with that attribute," p. 578) as a measure of family resemblance and showed that this measure was highly correlated with subjects' typicality ratings. That is, having features shared by many members of the same category makes an exemplar "good." We argue that even when category validity is held constant, the causal status of features can also determine the goodness of exemplars. Note that category validity and causal status of features are independent constructs in that high category validity does not necessarily entail high causal status (Ahn & Sloman, 1997; Keil, Smith, Simons, & Levin, 1998; Sloman et al., 1998). For example, almost all tires are black, but this feature has low causal status because it does not cause other features of a tire. We expect causal status to determine independently the graded structure of categories. It is because the more causally central a feature is, the greater impact it should have on the distance of an object to the central tendency or an ideal, which in turn, determines graded structure (Barsalou, 1985). Experiment 2 in the current study provides the first demonstration that features' causal status determines typicality ratings, even when the category validity of features is held constant.

Another construct that has been frequently associated with categorization and conceptual representation is similarity. According to some similarity-based theories of categorization (e.g., Hampton, 1998; Posner & Keele, 1968; Rosch, 1978), the causal status of a feature should have an equivalent effect on both categorization and similarity judgments because these models assume that categorization is based on similarity. In contrast, we predict that the causal status of features will be more important in categorization than in similarity judgments. This is because psychological essentialism and inductive inferences, constructs which we argued earlier are central to the causal status effect, are more important in categorization than in similarity. As discussed earlier, the causal status effect is expected to occur in categorization because people believe objects in the same category share the same essence which causes their surface features and because cause features are believed to have more inductive power than their effects. However, when people make judgments of similarity, the goal is more diverse. Whereas some types of similarity are used for inductive inferences (e.g., Gentner, 1989; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990), "mere appearance" matches such as "a planet is like a round ball" have virtually no predictive

utility (e.g., Gentner & Medina, 1998). Unless the goal of the similarity judgment is specified, the relative weighting of a feature can vary with the stimulus context and task (e.g., Goodman, 1972; Murphy & Medin, 1985). Therefore, in making similarity judgments the need is not as strong as in categorization to give greater weight to deeper cause features. Experiment 4 directly investigates the differential impact of causal status in categorization versus similarity judgments.

To summarize, the causal status effect is expected to be manifested when categorizing transfer items after learning a novel concept, when free-sorting objects, and when judging goodness of exemplars. These results will demonstrate how deeply embedded the causal status effect is in diverse aspects of our conceptual representations, above and beyond a simple demonstration of methodological generality. In addition, we propose that this bias toward weighing cause more than effect will be stronger in categorization than in similarity judgments, suggesting that the causal status effect might be due to psychological essentialism present only in people's categorization judgments.

Causal depth. We argue that the causal status effect is not just limited to the difference between the deepest cause of an entity and the surface effect features. Instead, feature centrality should be thought of as a continuum along a causal chain within a category (Medin & Ortony, 1989). The basis of this claim is directly derived from the causal status hypothesis. A feature that causes another feature can also be an effect feature of some other feature. For instance, "having gills" allows fish to swim and, at the same time, "having gills" in fish is an effect or consequence of "fish DNA." Because a feature is more central than its effect but less central than its cause, it is logical to predict that the deeper a cause is in a causal chain, the more conceptually central that feature is.

This continuous nature of feature centrality is one of the main departures of the causal status hypothesis from one reading of essentialism. As discussed earlier, essentialists argue that an object has an essence, the very being of what it is (Kripke, 1971; Locke, 1894/1975; Putnam, 1975). Applying this metaphysical account to how people might categorize things in the world, one might develop a strong version of essentialism that states that people treat essences as defining features of a category. That is, people treat essences as necessary and sufficient features of a kind in that if an object has an essence of a certain kind, the object must be a member of the kind, and if an object is a member of a certain kind, the object must possess the essence. Therefore, according to this strong version of essentialism, categorization is all-or-none rather than graded; that is, nonessential properties do not determine reference (see Braisby, Franks, & Hampton, 1996; Diesendruck & Gelman, 1999; Kalish, 1995; Keil, 1989; and Malt, 1994 for similar descriptions of essentialism). According to the causal status hypothesis, however, even if a feature is not the deepest cause, it is expected to play a role in categorization

judgments. Moreover, nonessential features are expected to influence categorization as a function of their causal status. Experiments 1 and 2 will provide the first empirical demonstration of this graded nature of the causal status effect.

Plausibility of causal relations and feature centrality. Causal relations vary in plausibility. For instance, severed brake lines are a plausible precursor to a car accident, but doing homework is not a plausible precursor to an allergic reaction (Fugelsang & Thompson, in press). We propose that the effect of causal status on feature centrality is moderated by the plausibility of the causal relations involved such that the more plausible causal relations are, the stronger the causal status effect is expected to be. There are many ways in which a causal relation becomes plausible, such as by acquiring an understanding of the underlying mechanism (Ahn, Kalish, Medin, & Gelman, 1995). Compatibility with preexisting causal background knowledge should also influence plausibility of newly acquired knowledge. Medin (1989) describes, as an example, that the germ theory, when first introduced, was not easily accepted because people had a prior belief that the size of a cause should be similar to the size of its effect, and they could not believe that such tiny entities as germs could result in disastrous effects. Experiment 6 tests the hypothesis that the degree of compatibility between prior knowledge and new knowledge can determine the size of the causal status effect. If our predictions are verified, this study will also demonstrate that the causal status effect is not fixed, but rather changes dynamically as one's causal background knowledge evolves.

The role of effect features. The causal status hypothesis does not necessarily imply that surface features (or the most terminal effect features) never affect categorization. Instead, the claim is that effect features matter *less* than their cause features in making categorization judgments. Consider the case of chicken pox, in which patients always exhibit pox marks on the skin and fever. Moreover, both these symptoms are caused by the chicken pox virus, which, like both symptoms, is always present in chicken pox. Suppose that a clinician finds that a patient has a fever and the chicken pox virus, but does not have pox marks. This absence of pox marks might make a clinician somewhat less likely to make a diagnosis of chicken pox because the patient does not display this highly characteristic symptom. However, note that if another hypothetical patient has pox marks and a fever, but a blood test shows that the person does not have the chicken pox virus, then it would be even much less likely that the clinician would diagnose this patient with chicken pox. Thus, our hypothesis is that the absence of effect features can matter in categorization because any missing features would matter, but the absence of (or discrepancy in) cause features matters even more. That is, the causal status hypothesis is about this difference between cause and effect features.

What role, then, does an effect feature play in categorization? The effect

features can be useful especially when deeper causes are not immediately observable. For instance, when we categorize people as males and females, we normally do not have information about their sex chromosomes. Even in such cases, people are willing to categorize based only on surface features (e.g., hair length and voice). In these cases when people cannot directly access information about cause features, however, they may infer the specific cause features based on their causal background knowledge. For instance, physicians infer the cause of one's illness based on the patient's symptoms (i.e., surface features). As in Bloom (1996), we also infer the designer's intention for an artifact based on its perceptual features. For instance, when observing an object with four legs and a seat, we infer that the designer of that object intended to create a chair. Sometimes, people might not even know what the essences are. Nonetheless, based on surface features, they infer a sort of "essence placeholder," filled with beliefs that experts exist who would know what the essences are (Medin & Ortony, 1989). In this case, surface features provide information about the nature of the essence placeholder, such as what sort of essence it is assumed to be, or what type of experts might know about the essences (see also Murphy & Medin, 1985, and Gelman & Medin, 1993, for similar discussion of the role of surface features).

It should be emphasized, however, that this role of effect features in serving as a heuristic guide for inferring cause features or assuming an essence placeholder should not be confused with feature centrality. For example, we might use the chemical composition of blood to infer the abnormality of a liver because the liver abnormality causes the changes in blood chemistry. Still, if we could somehow obtain more direct information that the blood test result was not due to a liver abnormality but occurred for some other reason, then the presence of the effect feature would be much less influential in a diagnosis of liver disease. Experiment 5 contrasts these two situations: one in which the cause feature is explicitly missing, and the other in which the information about the cause feature is simply unavailable. In accord with our views put forth in the above discussion, we predict that the causal status effect will disappear in the latter situation.

Structural rather than content-based constraints. The causal status hypothesis is a structural constraint rather than a content-based one; that is, a feature's centrality is determined by its causal status (i.e., structure of features) independent of its contents, such as the dimension on which it is measured (e.g., function or color) or attribute (e.g., red). Although this does not necessarily preclude the possibility of the role of feature content in categorization, a strong test of the causal status hypothesis would be to see whether the effect can be obtained even with features that are not usually thought of as "deeper" features. In general, essences and deeper features are believed to be inaccessible, internal, or hidden features, such as male hormones or DNA, whereas surface features are assumed to be obvious, perceptual fea-

tures. However, as Gelman and Wellman (1991) pointed out, essences are not necessarily insides (although they coincide in many cases). Likewise, features that cause other features do not have to be internal or hidden features, and features that are caused by other features do not have to be perceptual features. Indeed, it is illogical to assume that a cause feature is necessarily a hidden or internal feature because that same feature can be also an effect feature of another feature, as discussed before. Throughout the experiments reported in this article, we compare a condition in which features are causally related with a control condition in which features are not causally related in order to ensure that any effects obtained were not due to feature content.

Summary of Theoretical Claims and Overview of Experiments

To summarize the theoretical claims made so far, we propose that the causal status effect (1) is deeply rooted in categorization processes and therefore influences various aspects of concepts and categorization (e.g., categorization of new transfer items, free-sorting, and typicality judgments), (2) is less likely to affect similarity judgments because such judgments do not draw upon psychological essentialism, (3) is continuous in that a feature is more central than its effect but less central than its cause, (4) is a function of the plausibility of causal relations and can therefore change dynamically as causal background knowledge changes, (5) can disappear if cause features are not explicitly denied, and (6) can occur fairly independently of the content of the features. The current article reports six experiments investigating these issues.

Experiment 1: Causal Depth

Experiment 1 tests the hypothesis that in a causal chain, the deeper a cause is, the more central that feature is in influencing categorization judgments. Participants first learned three characteristic features of a target category (e.g., “Animals called ‘roobans’ tend to eat fruits, have sticky feet, and build nests on trees”). The features were selected in such a way that participants would be unlikely to have a priori knowledge about causal relations among them. In the causal background knowledge condition, participants learned that the three features, say X, Y, and Z, form a single causal chain (e.g., X causes Y and Y causes Z). For instance, participants learned that “Eating fruits tends to cause roobans to have sticky feet because sugar in the fruits is secreted through pores under their feet. Having sticky feet tends to allow roobans to build nests on trees because they can climb up the trees easily with sticky feet.” Hence, in this condition, a feature was a fundamental cause of all features (feature X, henceforth), a cause of one of the features but an effect of another feature (feature Y), or an effect of other features (feature

Z). The centrality of features was measured by asking participants to judge the membership likelihood of a new item missing one of the three features.

The hypothesis is that in the causal background condition, X would be judged to be more central than Y, which in turn would be judged to be more central than Z. In contrast, in the control condition, no causal background knowledge was provided. The difference between the control condition and the causal background knowledge condition will reveal the amount of impact that causal relations have on feature centrality. For instance, a null effect of features in the control condition would insure that the results from the causal background condition are not due to the content of the features.

Method

Participants. Fifteen undergraduate students at Yale University participated in this study either in partial fulfillment of requirements of an introductory psychology course or for payment of \$7.00 for participating in this experiment and other unrelated experiments.

Procedure. Each participant received 12 problems. For each of the 12 problems, participants were first told that a target category tends to have three features (X, Y, and Z, henceforth). Participants in the control condition did not receive any further information. Participants in the causal background knowledge condition, in contrast, learned that feature X causes feature Y and feature Y causes feature Z ($X \rightarrow Y \rightarrow Z$, henceforth). All participants were then presented with a description of a new item and asked to judge the likelihood that the item belonged to the target category. The new item was always missing exactly one of the characteristic features of the target category.

For example, one of the problems used in the causal background condition was the following. (The information that was not present in the control condition is written in italics here. However, it was presented in plain text to the participants in the causal background knowledge condition. Note that this item also includes one of the three test questions, which are described in more detail in the next section.)

Animals called "roobans" tend to eat fruits, have sticky feet, and build nests on trees.

In addition, biologists know the following.

Eating fruits tends to cause roobans to have sticky feet because sugar in fruits is secreted through pores under their feet.

Having sticky feet tends to allow roobans to build nests on trees because they can climb up the trees easily with sticky feet.

The above information can be summarized as follows: eats fruits \rightarrow has sticky feet \rightarrow builds nests on trees

Suppose an animal has the following characteristics.

likes to eat—worms

has feet that are—sticky

builds nests—on trees

How likely is it that this animal is a rooban?

The responses were made on a 0 to 100 scale, with 0 being "definitely *unlikely*" and 100 being "definitely *likely*". The experiment was programmed using Psyscope 1.1 (Cohen, MacWhinney, Flatt, & Provost, 1993) and run on Power PC Macintosh computers. At each trial, only one problem was presented on a computer screen. The background information (characteristic features as well as causal background information) was displayed on the screen while participants were answering the relevant question. Participants could correct their answer before proceeding to the next question, but not afterward. Questions concerning the same cate-

gory were blocked, and the order of questions within the same block was randomized across the participants. The order of categories was counterbalanced across participants. The order in which features were presented within each category and problem was held constant for both conditions.

Design and materials. For generality, four categories from four different domains were selected: animals, diseases, tribes, and cars. The actual features used in each category are summarized in Table 1. In the causal background knowledge condition, Feature X served as the fundamental cause, Feature Y served as the intermediate cause, and Feature Z served as the terminal effect. The instructions that were unique to the causal background knowledge condition are also provided in the last column of Table 1.

For each participant, two categories were used for the causal background condition, and the other two categories were used for the control condition. A latin-square design was used to determine which categories were used for the causal background condition or the control condition varied for each participant.

For each category, three questions were developed: Missing-X, Missing-Y, and Missing-Z. Each question presented a novel object that has only two of the target category's characteristic features. In two of the categories (disease and car), missing features were explicitly denied (e.g., the patient does not exhibit blurred vision and the car uses butane-free fuel), and in the other two categories (animal and tribe), missing features were described using implicit negatives (e.g., the animal eats worms and the tribe relies on hunting). Followed by this description of the novel object, a question was presented asking the likelihood that this object is a member of the target category.

Results and Discussion

The results are summarized in Fig. 1, with error bars indicating standard error. The results from the control condition indicate that three features within each category are equally central when they are not causally related. However, when the same features were causally related in the causal background knowledge condition, the likelihood judgments varied as a function of the missing feature's causal status. When an object was missing its fundamental cause in the causal chain (Missing X), the mean likelihood of being a target category member was lower than when an object was missing its intermediate cause in the causal chain (Missing Y), which in turn was lower than when an object was missing its terminal effect (Missing Z).

Subject analysis. A repeated-measures ANOVA with missing problem type and background knowledge conditions as the within-subject variables was carried out on each subject's average response on each problem type in each background knowledge condition. There was a reliable interaction effect, $F(2, 28) = 12.29$, $MS_e = 146.03$, $p < .001$. There was no main effect of background knowledge, but there was a reliable main effect of missing problem type, $F(2, 28) = 11.57$, $MS_e = 252.37$, $p < .001$, because the effect of missing problem type was strong in the causal background knowledge condition.

Planned comparisons were carried out to examine whether each adjacent step in the causal chain led to a reliable increase in the likelihood judgments. In the causal background knowledge condition, ratings on Missing-X problems were significantly lower than those on Missing-Y, $t(14) = -2.16$, $p <$

TABLE 1
Features and Causal Background Information Used in Experiment 1

Domains	Types of features			Causal background information
	X	Y	Z	
Animal (Roobans)	Eat fruits	Have sticky feet	Build nests on trees	Eating fruits tends to cause roobans to have sticky feet because sugar in fruits is secreted through pores under their feet. Having sticky feet tends to allow roobans to build nests on trees because they can climb up the trees easily with sticky feet.
Disease (Covition)	Blurred vision	Headache	Insomnia	Blurred vision tends to cause Covition patients to have a headache. A headache tends to cause Covition patients to suffer from insomnia.
Tribe (Hino)	Farming	Many leaders	Monotheistic	Relying on farming tends to cause Hino tribes to have many leaders because large-scale farming requires specialized decisions that must be coordinated by many leaders. Having many leaders, in turn, tends to cause Hino tribes to be monotheistic because unity under a single deity prevents squabbling and fighting for power among the many tribe leaders.
Car (Romanian Rogo)	Use butane-laden fuel	Hot engine temperature	Loose gas gasket	Butane-laden fuel in a Romanian Rogo tends to cause hot engine temperatures. The butane in the fuel burns at a hotter temperature than normal gasoline. Hot engine temperatures in a Romanian Rogo tends to cause a gas gasket to become loose. The heat in the engine makes the rubber around the gas gasket melt and become loose.

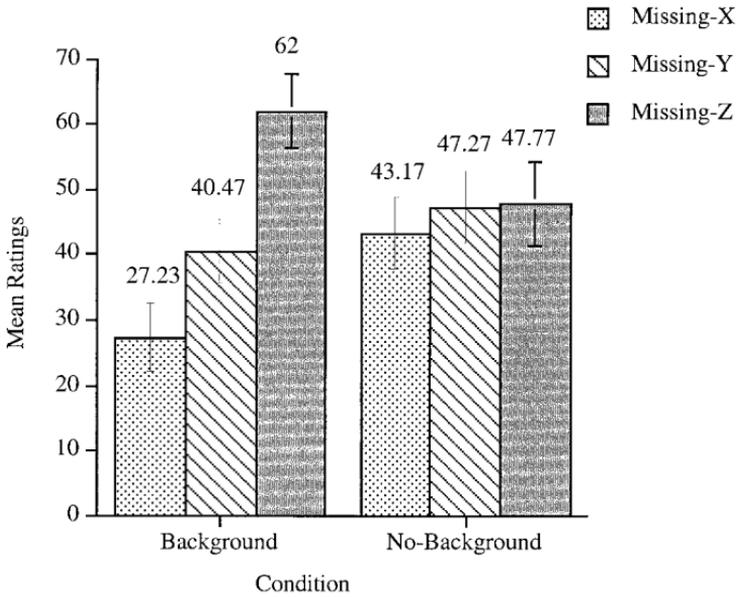


FIG. 1. Mean membership likelihood judgments of Experiment 1.

.05. Also, in the causal background knowledge condition, ratings on Missing-Y problems were significantly lower than those on Missing-Z, $t(14) = -5.53$, $p < .001$. However, in the no-background knowledge condition, no pairwise comparisons showed reliable differences.

Domain effect. Was the causal status effect more pronounced in a particular domain (i.e., item)? Table 2 shows the mean ratings broken down by each domain. As can be seen from this table, all four domains showed the causal status effect in the causal background knowledge condition. A 2 (background knowledge) \times 3 (missing problem type) \times 4 (domain) ANOVA

TABLE 2
Mean Percentage Ratings from Experiment 1

Category	Background knowledge condition			No-background knowledge condition		
	Missing-X	Missing-Y	Missing-Z	Missing-X	Missing-Y	Missing-Z
Animal	27.4	35.9	62.0	29.3	35.8	47.0
Car	28.3	41.2	71.4	43.3	52.3	64.3
Disease	26.7	49.2	66.5	48.3	51.7	52.2
Tribe	25.8	37.5	43.3	47.1	47.1	32.8

Note. Mean percentage ratings are broken down by stimulus category. Each item is missing feature X, Y, or Z, with the background knowledge that X causes Y, which causes Z (background knowledge condition) and without background knowledge (no-background knowledge condition).

was carried out with the missing problem type as a within-subject variable and the rest as between-subject variables.¹ The only statistically significant results were the main effect of missing problem type, $F(2, 104) = 16.30$, $MS_e = 368.69$, $p < .001$, and the interaction effect between background knowledge and missing problem type, $F(2, 104) = 7.19$, $MS_e = 368.69$, $p < .001$. Most importantly, the three-way interaction effect was not significant, $F(6, 104) = 0.23$, $MS_e = 368.69$, $p > .90$, indicating that the causal status effect did not depend on the domain. No other effects were statistically reliable.

To summarize, Experiment 1 provides the first empirical demonstration that the causal status effect is a matter of degree. That is, the most fundamental cause feature was judged to be more central than the intermediate cause in the causal chain, which, in turn, was more central than the most marginal effect feature. Thus, within a category, feature centrality forms a continuum along a causal chain.

EXPERIMENT 2: CAUSAL STATUS EFFECT IN GOODNESS-OF-EXEMPLAR JUDGMENTS

We propose that the causal status of features will also determine goodness-of-exemplar judgments such that exemplars with more central features will be judged to be better members of the category. As discussed earlier, Rosch and Mervis (1975) found that category validity is a good measure of typicality ratings. Experiment 2 attempts to show that causal status of features also determines typicality ratings, independent of category validity.

Experiment 2 also controls for the category validity of features, which was a necessary step to rule out one possible alternative interpretation for Experiment 1: Participants might have assumed that the category validities of features vary as a function of causal status. Whereas category validities of features can be an important determinant of feature centrality, high category validities do not necessarily entail high causal status, as discussed earlier. Note, however, that most cause features in basic-level categories seem to also have high category validities. For example, birds' genetic codes are causally central, but they are also present in all birds. This observation suggests that, due to this correlation in real-life categories, participants in Experiment 1 (which did not explicitly specify category validities of features) might have assumed that the cause features must have high category validities. Following this logic, it could be argued that these inferred high category validities of the cause features determined their conceptual centrality rather

¹ As explained earlier, different subjects received different combinations of domain and background knowledge, so that the interaction between the two experimental factors is confounded with within-subject variance. Ignoring within-subject variance allows, therefore, a test of the interaction at the expense of statistical power.

than causal status per se. Therefore, in Experiment 2 we also investigate whether the causal status effect occurs even when category validities are held constant.

In Experiment 2, participants observed 12 representative samples of the target category before they made judgments on transfer items. The actual category validities of three characteristic features of a target category were held constant. After observing 12 samples, participants were asked to judge frequencies of each feature. No difference in frequencies of three features would ensure that the causal status effect is not due to misperceived category validities of features. Afterward, they made typicality ratings of items missing the deepest cause, the intermediate cause, or the terminal effect.

Methods

Participants. Twenty-seven undergraduate students at Yale University participated in this study either in partial fulfillment of requirements of an introductory psychology course or for payment of \$7.00 for participating in this experiment and other unrelated experiments.

Materials and procedure. The materials and procedure of Experiment 2 were identical to those of Experiment 1, except that in Experiment 2, (1) participants observed 12 samples of each target category and made frequency judgments for each of the features and (2) they judged the goodness of membership of the transfer items.

For each category, participants observed 12 samples. They were told that they would first see descriptions of 12 examples of a target category they were about to learn. They were further told that "these are representative samples of [the target category to be learned]." The 12 samples consisted of 6 exemplars with all three characteristics of the target category, 2 exemplars of Missing-X, 2 exemplars of Missing-Y, and 2 exemplars of Missing-Z. Therefore, the category validity of each of the three features was .83. The correlations among features were also held constant. The 12 samples were presented in a randomized order, one at a time. Participants read each description at their own pace.

After observing 12 samples, participants judged frequencies of the three characteristic features. They were asked, "How often did the following characteristic appear in the samples you saw?" and were presented with the target feature. Participants answered the question by entering a number on a keyboard.

In the control condition, this observation/frequency judgment phase was inserted before participants were presented with transfer items, and in the causal background knowledge condition, it was inserted before they learned causal relations in the causal background knowledge. The reason for this procedure, in the causal background knowledge condition, was that previous studies (e.g., Sloman et al., 1998) found that background knowledge itself can affect how feature frequencies are perceived. Because the goal of Experiment 2 is to equate perceived category validities of features, the observation phase was inserted before learning causal relations.

It should be also noted that in learning category validities of features, participants directly experienced exemplars of each category rather than were presented with summary statistics. This procedure was adopted because previous studies suggested that the impact of base-rate information increases when participants directly experience it on a trial-by-trial basis (e.g., Koehler, 1996; Medin & Edelson, 1988). Hence, participants in Experiment 2 should be more, not less, likely to use the frequency information than if they had been given summary statistics. This provides a stronger test of causal status as pitted against use of category validities.

The second difference from Experiment 1 was the type of judgment to be made concerning each transfer item. Instead of judging membership likelihood, participants in Experiment 2 judged goodness of membership (e.g., "How good an example of a Romanian Rogo would

TABLE 3
Mean Frequency Judgments of the Three Characteristic Features
in Each Condition in Experiment 2

Conditions	Type of features		
	X	Y	Z
Causal background knowledge condition	9.22 (1.08)	9.31 (1.04)	9.19 (1.14)
Control condition	9.39 (1.20)	8.87 (1.57)	8.78 (1.52)

Note. Standard deviations are in parentheses.

you consider this car?"). Participants answered each question on a 9-point scale, where 1 was "not good at all" and 9 was "very good."

Results and Discussion

Mean frequency judgments are provided in Table 3. A 2×3 repeated-measures ANOVA showed that there was no reliable main effect of background knowledge conditions, no reliable main effect of feature type, and no reliable interaction effect between the two factors; all p 's $> .10$. Although participants' mean frequency judgments were not accurate (means in all six conditions differed statistically from actual frequency of 10; all p 's $< .01$), the critical result is that there was no systematic difference in perceived frequencies with respect to the background knowledge condition or the type of features. For instance, the results indicate that participants were aware that feature X (9.22) was as frequent as feature Y (9.31) and feature Z (9.19) in the causal background knowledge condition (all p 's $> .50$).

Even though participants knew that the category validities of the characteristic features were equated, the causal status effect was replicated. The mean goodness of membership ratings are summarized in Fig. 2 with the error bars indicating standard error. The results from the control condition indicate that three features within each category are equally central when they are not causally related. However, when the same features were causally related in the causal background knowledge condition, the goodness judgments varied as a function of the missing feature's causal status. When an object was missing its fundamental cause in the causal chain (Missing X), the mean likelihood of being a target category member ($M = 3.80$) was lower than when an object was missing its intermediate cause in the causal chain (Missing Y, $M = 4.48$), which in turn was lower than when an object was missing its terminal effect (Missing Z, $M = 5.74$).

Subject analysis. A repeated-measures ANOVA with missing problem type and background knowledge conditions as the within-subject variables was carried out on each subject's average response on each problem type in each background knowledge condition. There was a reliable interaction effect, $F(2, 52) = 6.34$, $MS_e = 1.61$, $p < .01$. There was a reliable main effect

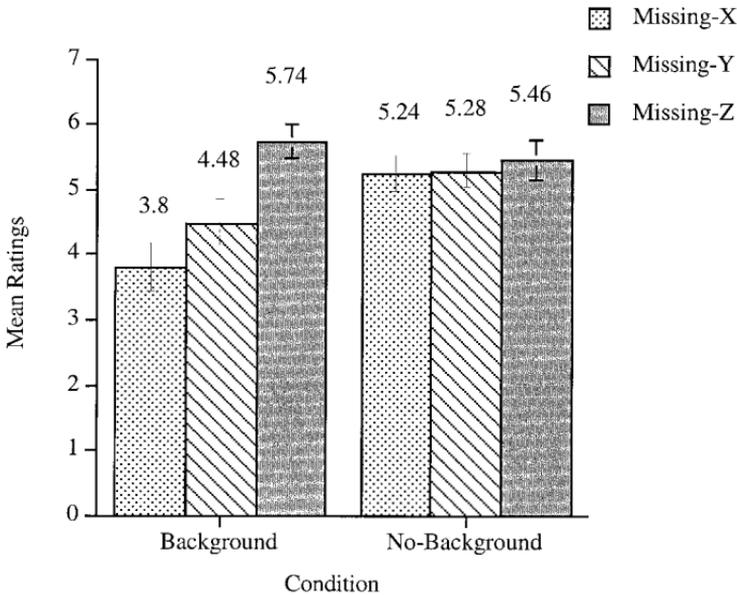


FIG. 2. Mean goodness of membership judgments in Experiment 2.

of background knowledge, $F(1, 26) = 27.06$, $MS_e = 0.64$, $p < .001$, because the causal background knowledge condition overall led to lower goodness ratings ($M = 4.67$) than the control condition ($M = 5.33$). There was a reliable main effect of missing problem type, $F(2, 52) = 12.93$, $MS_e = 1.27$, $p < .001$, because Missing-X ($M = 4.52$) led to lower goodness judgments than Missing-Y ($M = 4.88$), which in turn led to lower goodness judgments than Missing-Z ($M = 5.6$). Both reliable main effects should be interpreted in light of the interaction effect because they occurred mainly because of lower ratings of Missing-X items and Missing-Z items in the causal background knowledge condition.

Planned comparisons were carried out to examine whether each adjacent step in the causal chain led to a reliable increase in goodness-of-membership judgments. In the causal background knowledge condition, ratings on Missing-X problems were significantly lower than those on Missing-Y, $t(26) = -2.25$, $p < .05$. Also, in the causal background knowledge condition, ratings on Missing-Y problems were significantly lower than those on Missing-Z, $t(26) = -3.74$, $p < .001$. However, in the no-background knowledge condition, the same pairwise comparisons showed no reliable differences.

Domain effect. It was examined whether the causal status effect was more pronounced in a particular domain. Table 4 shows the mean ratings broken down by each domain. A 2 (background knowledge) \times 3 (missing problem type) \times 4 (domain) ANOVA was carried out with missing problem type as a within-subject variable and the rest as between-subject variables. Again,

TABLE 4
Mean Goodness-of-Exemplar Ratings from Experiment 2 for Each Category

Category	Background knowledge condition			No-background knowledge condition		
	Missing-X	Missing-Y	Missing-Z	Missing-X	Missing-Y	Missing-Z
Animal	3.54	3.92	5.54	5.80	5.27	6.07
Car	3.64	4.00	5.79	4.50	4.93	5.64
Disease	4.21	5.43	6.40	5.00	6.23	5.62
Tribe	4.36	4.71	5.14	5.43	4.71	4.57

Note. Mean ratings are broken down by stimulus category. Each item is missing feature X, Y, or Z, with the background knowledge that X causes Y, which causes Z (background knowledge condition) and without background knowledge (no-background knowledge condition).

there was a causal status effect as shown by the significant interaction effect between missing problem type and background knowledge, $F(2, 198) = 8.89$, $MS_e = 2.30$, $p < .001$. Most importantly, this causal status effect was not dependent on the domain as shown by no significant three-way interaction effect, $p > .90$. In addition, there was a significant main effect of background knowledge, $F(1, 99) = 5.10$, $MS_e = 6.67$, $p < .05$, a significant main effect of missing problem type, $F(2, 198) = 13.35$, $MS_e = 2.30$, $p < .001$, and a significant interaction effect between the domain and missing problem type, $F(6, 198) = 7.30$, $MS_e = 2.30$, $p = 0.01$. Other effects did not reach statistical significance.

To summarize, the deeper the cause an item was missing, the worse an exemplar it was judged to be. Thus, the results showed that the causal status of features determined the goodness of exemplars even when perceived base rates of features were held constant.

EXPERIMENT 3: FREE-SORTING

In the two experiments reported so far, the causal status effect was obtained after participants learned preestablished categories. Will it also occur when participants are free to sort objects without any criterion about what the categories should look like? The task used in Experiment 3 is a standard match-to-sample task that has been frequently used in studies on categorization and similarity (e.g., Tversky, 1977; Medin, Goldstone, & Gentner, 1993). Consider the triad in Fig. 3 under the Causal Condition. In this triad, Jane (Target) is depressed because she has low self-esteem, Susan is depressed because she has been drinking, and Barbara is defensive because she has low self-esteem. Note that compared to the Target, one case (Susan) has a matching effect but a differing cause (Matching-Effect) and the other case (Barbara) has a matching cause but a differing effect (Matching-Cause). Participants in Experiment 3 were asked which item should be categorized with the Target. If the causal status effect occurs, matching on a cause feature

Causal Condition

		Target
		Jane is depressed. Jane has low self-esteem, which is the reason why she is depressed.
A		B
Barbara is defensive. Barbara has low self-esteem, which is the reason why she is defensive.		Susan is depressed. Susan has been drinking, which is the reason why she is depressed.

Non-Causal Condition

		Target
		Jane is depressed. Jane has low self-esteem, which is NOT the reason why she is depressed.
A		B
Barbara is defensive. Barbara has low self-esteem, which is NOT the reason why she is defensive.		Susan is depressed. Susan has been drinking, which is NOT the reason why she is depressed.

FIG. 3. Sample stimulus materials used in Experiment 3 for the Causal and the Noncausal conditions. In the Causal condition for this sample, Option A is the Matching-Cause item and Option B is the Matching-Effect item.

will be considered more important than matching on an effect feature and consequently, participants will prefer the Matching-Cause case over the Matching-Effect case.

Even if the above results are obtained, however, several alternative explanations are possible, and therefore, two control measures were taken in Experiment 3. First, the above results might occur because the features we used as the cause features might happen to be more salient than the features we used as the effect features. Suppose the items in a triad are schematically described as Target ($P \rightarrow Q$), Option A ($P \rightarrow R$), and Option B ($S \rightarrow Q$), where each letter in parentheses stands for a feature of each item, and an arrow indicates the causal direction of the relationship between two features. The argument here would be that people might sort Option A with the target because P is more salient than Q, not because P is more causal than Q. To eliminate this possibility, a control condition was employed in which all features and tasks were identical to those used in the experimental condition, except that the causal relations among features were explicitly denied (see the Noncausal condition in Fig. 3 for an example). If the salience or the content of the matching feature is the only reason for selecting the matching-cause option over the matching-effect option, then the same pattern of the results should be obtained even when the causal relations are explicitly de-

nied. The difference between this Noncausal condition and the Causal condition would indicate the amount of the causal status effect.

Second, the above predicted results could be obtained because of differences in the similarity of mismatching features. Going back to the above schematic presentation of a triad, suppose Q and R (i.e., the mismatching features between Target and Option A) are more similar than P and S (i.e., the mismatching features between Target and Option B) are. If so, the Target would look more similar overall to Option A than to Option B, not because of the causal status effect, but because of differences in the similarity between the mismatching features. To control for this possibility, a pretest was conducted by asking an independent group of subjects to judge the similarity between Q and R and the similarity between P and S. Features for the main experiment were selected in such a way that these two similarities were equated.

Methods

Participants. Forty-eight undergraduate students at Yale University participated in this experiment in partial fulfillment of introductory psychology course requirements.

Materials and design. Six sets of problems were developed. Each problem consisted of a triad of cases with one target and two options, A and B. Each case in a triad was described in terms of two features. The target and Option A shared one feature and the target and Option B shared another feature. For instance, one target case was "Ann lost 7 pounds last month. Ann has food allergies." Option A for this triad was "Cathy lost 7 pounds last month. Cathy had her wisdom teeth pulled." Option B for this triad was "Kim has rashes. Kim has food allergies." Table 5 lists all six sets of features used in Experiment 3.

As explained in the introduction of this experiment, it is important to ensure that any effect was not due to similarity of the nonmatching features. Thus, the features shown in Table 5 were selected through a pretest. Fourteen participants were presented with the 12 pairs of nonmatching features and asked to rate each pair for how similar the two features were to each other (on a scale of 1–7, where 1 = *very dissimilar* and 7 = *very similar*). For instance, they rated similarity between having food allergies and having a wisdom tooth pulled out (i.e., mismatching features between Ann and Cathy in the above example) and similarity between losing 7 pounds and having rashes (i.e., mismatching features between Ann and Kim). Six paired *t* tests were run to compare the two sets of ratings for each item, and revealed no significant, or even marginally significant, differences between them (all p 's > .1).

For each triad, two types of problems were developed, Causal and Noncausal. The problems in the Causal condition explicitly stated that the one feature is the reason for another feature within each of the three cases in each set. The top half of Fig. 3 shows an example triad in the Causal condition. The problems in the Noncausal condition explicitly stated that one feature is not the reason for another feature in all three cases, as shown in the bottom half of Fig. 3. Since the problems in the Causal and the Noncausal conditions contained identical features and the only difference between the two conditions was whether the features were causally related, any differences between the Causal condition and the Noncausal control condition should be attributable to the effect of causal status.

In addition, two orders of presentation for each type of problem (Causal or Noncausal) were developed. In one version, the first feature within each of the three cases shown in Table 5 was stated first (e.g., "Ann lost 7 pounds last month. Ann has food allergies, which is [not] the reason why she lost 7 pounds last month" where "not" in brackets appeared only in the

TABLE 5
Stimulus Materials Used in Experiment 3

Triad	Target	Option A	Option B
1	Himolu birds have slow digestion. Himolu birds have iron sulfate in their blood.	Raconi birds have slow digestion. Raconi birds have low levels of the hormone secretin.	Semuto birds build nests quickly. Semuto birds have iron sulfate in their blood.
2	Noros beetles have a short flight response. Noros beetles have a large quantity of ACh (neurotransmitter) in the head ganglia.	Rakof beetles have a short flight response. Rakof beetles have a spiky exoskeleton.	Telig beetles have a high body weight. Telig beetles have a large quantity of ACh (neurotransmitter) in the head ganglia.
3	Mark cannot remember his past at all. Mark has been working in a highly polluted factory for the last 10 years.	Brian cannot remember his past at all. Brian was born with a defective hipposampus.	Ed is blind. Ed has been working in a highly polluted factory for the last 10 years.
4	Jane is depressed. Jane has low self-esteem.	Susan is depressed. Susan has been drinking.	Barbara is defensive. Barbara has low self-esteem.
5	Mary donates \$200 to a charity every month. Mary enjoys helping others.	Joanne donates \$200 to a charity every month. Joanne wants to maximize her tax deductions.	Jennifer works as a personal injury lawyer. Jennifer enjoys helping others.
6	Ann lost 7 pounds last month. Ann has food allergies.	Cathy lost 7 pounds last month. Cathy had her wisdom teeth pulled.	Kim has rashes. Kim has food allergies.

Note. Whether Option A or Option B shares the first or the second feature of the target case was counterbalanced across all participants in the experiment. For simplicity, Table 5 depicts Option A as sharing the first feature of the target and Option B as sharing the second feature of the target.

Noncausal condition). In the other version, the second feature in Table 5 was stated first (e.g., "Ann has food allergies. Ann lost 7 pounds last month, which is [not] the result of her having food allergies" where "not" in brackets appeared only in the Noncausal condition).

Each problem was printed on a single page and began with the question, "Which one would you categorize with the target, A or B?" Below this question, the Target was presented in the center. Below the Target, Options A and B were placed side by side (see Fig. 3 for a sample layout). In half of the sets, the option sharing the first feature (or the cause feature in the Causal condition) was presented on the left as Option A and in the other half, it was presented on the right as Option B.

For each participant, a booklet containing six problems, three from the Causal condition and three from the Noncausal condition, was prepared. One set of booklets had Triads 1, 3, and 5 in Table 5 for the problems in the Noncausal condition and Triads 2, 4, and 6 for the problems in the Causal condition. The other set of booklets had Triads 1, 3, and 5 for the Causal condition and Triads 2, 4, and 6 for the Noncausal condition. This way, all participants were presented with both the Causal and the Noncausal problems, but the same participant never saw the same stimuli in both Causal and Noncausal conditions. Following a 2 (which problems are presented as causal) X 2 (Matching-Cause option on Left or Right) X 2 (feature 1 or 2 presented first) latin-squares design, eight sets of booklets were prepared. Each subject received one of the eight sets of booklets.

Procedure. Participants received a booklet containing instructions and six problems, each of which was arranged in the manner explained above. They were asked to answer the question "which one would you categorize with the target, A or B?" by circling either Option A or B. The order of the six problems was completely randomized for each participant.

Results and Discussion

In the Noncausal condition, participants' responses were essentially split in half across the two options in that the option sharing the first feature with the target was preferred 54.2% of the time over the other option. However, in the Causal condition, the preference for this option was increased to 73.6%. This increase is remarkable in that the options were almost identical across the two conditions, the only difference being whether the common feature between the option and the target was serving as a cause feature. Hence, this near 20% increase in preference strongly supports the claim that merely changing the causal status of the features can change the centrality of the features and hence the categorization of the animal or person to whom the features belong.

For the statistical analyses, two scores were calculated for each participant: the number of times the matching cause was chosen in the Causal condition, and the number of times the corresponding matching "Noncause" was chosen in the Noncausal condition. Since there were three problems for each condition, each of the four scores could range between 0 and 3. A paired *t*-test showed that scores from the Matching-Cause chosen in the Causal condition ($M = 2.21$, $SD = 0.82$) were significantly higher than scores from the matching noncause chosen in the Noncausal condition ($M = 1.63$, $SD = .98$), $t(47) = 3.23$, $p = .002$. An item analysis revealed that the results were in the same direction in all six sets. Moreover, when only the first response of each participant was analyzed, the results remained consistent (in the Causal

condition, 82.1% of responses were Matching-Cause compared with 55.0% in the Noncausal condition).

To summarize, Experiment 3 tested whether the causal status hypothesis would occur when the task was to free-sort exemplars into categories that were not prespecified. The results demonstrate that people prefer to categorize based on causes rather than effects.

The task posed in this experiment (i.e., whether to categorize based on matching cause or matching effect) mimics real-life situations in which people are faced with the task of constructing new categories when causal relations among features are discovered. For instance, at the current stage of scientific knowledge, the causes of many disorders (e.g., infant depression; Guedeny, 1997) are controversial. Hence, at this point, such disorders are diagnosed based on the manifestation of symptoms (e.g., psychomotor retardation). One might argue that this might be an example of a case where classification is based on effect features. However, in this case, the causal status hypothesis does not apply because the cause features are unknown. For the sake of argument, suppose scientists have recently specified the exact mechanisms underlying infant depression and found that there are two different, specific causes for infant depression. Would doctors still classify patients based on symptoms or would they now classify patients based on causes? By providing two potential causes for the same effect and also two effects for the same cause in a single triad, Experiment 3 examines how laypeople facing this more well-defined task classify objects.

The results show that once causal relations are clearly specified, people classify based on causes. For instance, consider Fig. 3, in which Option A and the Target case both describe people who are depressed. In the Noncausal condition, 54% of the participants categorized Option B with the Target. In the Causal condition, however, participants were told that the cause of depression differed between Option B and the Target. Once it was made explicit that the cause of depression differed from the Target, their preference for Option B dropped by 17%. A full 71% of the participants in the Causal condition preferred Option A, the Matching-Cause item, despite the fact that this item has a different symptom (defensiveness). This type of categorization based on causes is consistent with that utilized by the *DSM-IV* (1994), which distinguishes a major depressive episode from symptoms “due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hypothyroidism; p. 327).” The results from Experiment 3 suggest that once the differences in causes are revealed, laypeople also prefer to categorize objects based on these causes.

EXPERIMENT 4: SIMILARITY JUDGMENTS

So far, we have been concerned primarily with how causal status affects feature weighting in categorization. To what extent, however, does this

causal status effect hold for other types of judgments besides categorization? One type of judgment that may be particularly informative is similarity. According to some similarity-based theories of categorization (e.g., Hampton, 1998; Posner & Keele, 1968; Rosch, 1978), the causal status of a feature should have an equivalent effect in both categorization and similarity judgments because these models assume that categorization is based on similarity. Preliminary studies by Ahn and Dennis (1997) show that the causal status effect also occurs in similarity judgments. They used the match-to-sample task in Experiment 3 and found that participants judged that objects sharing a common cause were more similar to each other than objects sharing a common effect.

Even if the causal status effect occurs in similarity judgments, however, there are reasons to expect that it might not be as robust in similarity judgments as in categorization. In Ripps' (1989) study, for example, a fictitious bird called a "sorp" ate chemically contaminated vegetables and changed its appearance to that of an insect. After this accidental change, participants were still more likely to categorize the object as a bird than to judge that the object was similar to a bird. Conversely, if a sorp went through an essential change as a result of maturation, categorization judgments were more affected than similarity judgments. This dissociation between categorization and similarity judgments may be explainable if the causal status of a feature does not play as important a role in similarity judgment as in categorization.² Within our causal status hypothesis framework, the "essence" of an object may be viewed as a hidden causal feature in the above studies (Medin & Ortony, 1989). Thus, Ripps' study suggests that changes in causal features (i.e., essence) matter more in categorization judgments than in similarity judgments. Furthermore, as discussed in the introduction, we would expect psychological essentialism, which serves as a basis for the causal status effect, to operate in categorization and not in similarity judgments.

Experiment 4 directly tests the hypothesis that the dissociation between categorization and similarity judgments can occur due to the differential importance of causal status in the two tasks. Participants were shown sets of three objects (a target object and two options) as in Experiment 3. Participants were asked either to categorize one option with the target or to decide which option was more similar to the target. One way of showing that the causal status effect is weaker in similarity judgments than in categorization judgments is to impose a counteracting force against causal status and to examine which task still shows the causal status effect despite the counteracting force. To implement this manipulation, new object descriptions were constructed with causal structures such that the matching-cause options would have fewer shared features with the target than the matching-effect

² We do not wish to suggest that causal status is the only factor leading to similarity-categorization dissociations. Ripps (1989), for example, also found this dissociation by manipulating the variability of categories.

option. More specifically, the target object had three features, one causing two surface features ($A \rightarrow B$ and C). In the matching-effect option, both effect features were shared with the target ($F \rightarrow B$ and C), and in the matching-cause option, only the single causal feature was shared with the target ($A \rightarrow D$ and E). Table 6 shows the materials used in this study.

If the causal status effect is stronger in categorization than in similarity judgments, the option that shares the smaller number of features with the target but shares a cause feature with the target ($A \rightarrow D$ and E) should be more likely to be selected in categorization judgments than in similarity judgments. That is, when this counteracting force (i.e., a smaller number of shared features) is imposed against causal status, the causal status effect is expected to be weakened in similarity judgments as compared to categorization judgments.

Methods

Participants. Fifty-two undergraduate students at Yale University participated in this experiment in partial fulfillment of requirements for an introductory to psychology course.

Materials and procedure. Six sets of materials were developed. Two sets were natural kinds, two were artifacts, and two were artworks. In each set, there were three objects: a target and two options. The target had three features, and one of the features (feature A) caused the other two features (features B and C). The target is schematically represented as $A \rightarrow B$ and C in Table 6. One of the options shared the cause feature with the target (Option $A \rightarrow D$ and E) and the other option shared the two effect features with the target (Option $F \rightarrow B$ and C). Table 6 shows the actual items used for each domain.

As in Experiment 3, each set of materials was presented on a separate page. At the top of each page was either a similarity or a categorization question. The similarity question read “Which one is more similar to the Target, A or B?” and the categorization question read “Which one should be categorized with the Target, A or B?” After the similarity or categorization question, the objects were laid out in a triad, such that the target was placed at the top and the two options, labeled as A or B, were placed side by side at the bottom. The location of the two options, right or left, was counterbalanced across participants.

All participants received all six sets of materials in a randomized order. Twenty-five participants were asked the categorization question for all six sets; 27 were asked the similarity question for all six sets. Participants responded by circling either A or B in the booklet.

Results and Discussion

Overall, the expected dissociation between categorization and similarity was obtained. When asked to judge similarity to the target ($A \rightarrow B$ and C), participants chose the matching-effects option (Option $F \rightarrow B$ and C , 64.2% of the time) over the matching-cause option (Option $A \rightarrow D$ and E , 35.8% of the time). That is, for similarity judgments, there was a tendency to prefer the option that shared two surface features with the target, and the causal status effect disappeared. In contrast, when asked to categorize the objects, participants preferred the option that shared a deeper feature despite the fact that this option shared fewer features with the target. Given the categorization task, participants chose the matching-cause option (Option $A \rightarrow D$ and

TABLE 6
Materials Used in Experiment 4

Set of materials	Target A → B and C	Option F → B and C	Option A → D and E
Object			
Artifact 1	This object has a rubber platform and vibrates smoothly because it was designed to relax pregnant mares during labor.	This object has a rubber platform and vibrates smoothly because it was designed to massage one's back.	This object has a record player and a cooling fan because it was designed to relax pregnant mares during labor.
Artifact 2	This object has a high-intensity light bulb and a pouch that can contain liquid because it was designed to kill bugs.	This object has a high-intensity light bulb and a pouch that can contain liquid because it was designed to be used in a photograph studio.	This object has a sweet, smelly patch and an x-ray generator because it was designed to kill bugs.
Artwork 1	This painting has four pillars and is red because the painter intended to draw a dog.	This painting has four pillars and is red because the painter intended to draw a cat.	This painting has one rectangle and is blue because the painter intended to draw a dog.
Artwork 2	This sculpture is made of metal and consists of six cubes stacked up because the sculptor intended it to symbolize pollution.	This sculpture is made of plastic sheets covering up a globe because the sculptor intended it to symbolize pollution.	This sculpture is made of metal and consists of six cubes stacked up because the sculptor intended it to symbolize a family.
Natural kind 1	This plant has needle leaves, and produces tiny pink flowers in the spring because it has a DNA structure called Valva.	This plant has needle leaves, and produces tiny pink flowers in the spring because it has a DNA structure called XA10.	This plant has wide leaves, and produces tiny white flowers in the fall because it has a DNA structure called Valva.
Natural kind 2	This animal has a block-shaped head, is red, and has 13 teeth because this animal has a genetic code, XB12.	This animal has a block-shaped head, is red, and has 13 teeth because this animal has a genetic code, Zebura 36.	This animal has a cylinder-shaped head, is blue, and has 23 teeth because this animal has a genetic code, XB12.

E, 59.3% of the time) over the matching-effects option (Option F \rightarrow B and C, 40.7% of the time).³

Independent *t*-tests were performed by coding the choice of Option A \rightarrow D and E (matching-cause option) as 1, and the choice of Option F \rightarrow B & C (matching-effects option) as 0. These scores were then summed for each participant to create a total score indicating preference for the matching-cause option across all items. Because each participant made categorization or similarity judgments for all six items, the score can range from 0 to 6. Indeed, preference for the matching-cause option in similarity judgment ($M = 2.15$, $SD = 1.66$) was significantly lower than in categorization judgment ($M = 3.56$, $SD = 1.66$), $t(50) = 3.07$, $p < .01$.

Table 7 shows the percentage of choices between the two options, collapsed across all items, along with the percentage of choices for each of the six items, classified by domain. As can be seen in this table, the basic pattern of an increased preference for deep features in categorization was obtained in all six items.

The results from this experiment show that causal status is not of equal importance in the two types of judgments. Specifically, the effect of causal status is less strong in similarity judgment than in categorization. The two effect features shared between a pair of objects led people to judge those objects as more similar than the pair of objects with a single shared cause feature. Such a bias toward a greater number of shared features was much weaker in categorization because people have stronger preference to perform this task based on the causal status of the shared features than when judging similarity among objects.

The current results also have important implications for the usefulness of similarity-based models. Like previous demonstrations of the dissociation between similarity and categorization (e.g., Keil, 1989; Rips, 1989), the current results pose problems for a strong version of similarity-based models, which argue that categorization is based on perceived similarity among objects. This nonmonotonicity between similarity and categorization has often been attributed to differential weighting of features in the two tasks. However, previous demonstrations of this dissociation have also been criticized by supporters of the similarity-based view as mainly relying on a single paradigm, namely the transformation paradigm, in which a creature has metamorphosed from one form to another form (Hampton, 1998). These critics have further argued that such transformations are not common in natural categories, especially in nonbiological kinds. Experiment 4 bypassed this problem by using a different paradigm, in which the causal status of matching and

³ One might wonder why the causal status effect in the categorization task appears somewhat weaker in this experiment compared to the previous experiments. Recall that in Experiment 4, the number of matching features counteracts the causal status effect in that the matching cause item has a smaller number of matching features than the matching effect item does.

TABLE 7
 Percentage of Choices for Categorization and Similarity Judgments for Each Set of Materials in Experiment 4

Set of materials	Categorization		Similarity	
	Option F \rightarrow B and C	Option A \rightarrow D and E	Option F \rightarrow B and C	Option A \rightarrow D and E
Overall	40.7	59.3	64.2	35.8
Artifact 1	40.0	60.0	70.4	29.6
Artifact 2	20.0	80.0	37.0	63.0
Artwork 1	60.0	40.0	77.8	22.2
Artwork 2	28.0	72.0	44.4	55.6
Natural kind 1	48.0	52.0	85.2	14.8
Natural kind 2	48.0	52.0	70.4	29.6

mismatching features was manipulated. Thus, it provides what may be the most direct evidence to date for differential feature weighting in similarity versus categorization judgments.

In addition, it should be also pointed out that the stimulus materials involving artworks and the artist's intention behind them were selected to relate to the literature on intention in naming (Bloom & Markson, 1998; Gelman & Ebeling, 1998). Bloom and Markson (1998) showed that when two intended referents had identical shapes, children preferred to name the pictures according to the intended representation. Gelman and Ebeling (1998) showed that when the same picture was produced either intentionally or accidentally, intentional representation led to higher rates of naming responses than did accidental representation. It remains an open question as to whether the preference for intention is due to its causal status or whether the causal status is due to the intention. We speculate that the preference for intention in naming representational objects may be a special case of the causal status effect because the causal status effect has also been found in non-representational objects as well as objects involving no intentionality. Future research can empirically determine this issue by directly pitting causal status against intention.

EXPERIMENT 5: WHEN INFORMATION ABOUT CAUSES IS UNAVAILABLE

So far, we have measured the centrality of features either by explicitly omitting them (Experiments 1 and 2) or by replacing them with different features (Experiments 3 and 4). Quite often, however, features can be "missing" simply because they are unobservable or information about their presence or absence is otherwise unavailable. For instance, most of the time we can only observe the perceptual appearance and movement of an animal. Yet, people seem quite confident in categorizing animals without knowing their genetic code. Is this a counterexample to the causal status hypothesis?

We propose that it is not in that unavailable information about features can be inferred through background knowledge about interproperty relations. This process is similar to the classic phenomenon on the effects of schema or scripts in which people made inferences about the presence of missing features by using their existing schema (e.g., Bower, Black, & Turner, 1979; Brewer & Treyens, 1981). For instance, after reading a story involving a restaurant scene, people falsely reported that the passage contained a sentence about the protagonist paying for the meal when in fact such a sentence was not present. Our proposal is also similar to the notion that analogical inference works by carrying over a missing structure from one domain to another given a partial correspondence (e.g., Clement & Gentner, 1991; Markman, 1997).

If unavailable features are inferred from available information, we would

expect the causal status effect to seem to disappear. People, having inferred the presence of the missing feature, will treat missing-cause and missing-effect cases as though all features are present. In abstract notation, suppose “?” indicates unavailable information and “ \rightarrow ” indicates a causal direction. If people generally believe $X \rightarrow Y$, then $? \rightarrow Y$ (missing-cause) and $X \rightarrow ?$ (missing-effect) would be both inferred as $X \rightarrow Y$. Consequently, there would be no difference in categorization judgments between missing-cause and missing-effect cases. Going back to our previous example, we can confidently categorize animals just by looking at their perceptual appearance and movement because people infer the missing feature (e.g., “flamingo genes”) based on their causal background knowledge. If this is the case, then our confident category decision does not stem from surface features alone, but rather from surface features plus the inferred causal features. Thus, we argue that such cases do not necessarily contradict the causal status hypothesis. By providing evidence that hidden features are inferred through one’s background knowledge, Experiment 5 attempts to show that people’s confidence in making categorization judgments based on surface feature information is not necessarily a counterexample of the causal status effect.

Experiment 5 uses a task similar to that used in Experiments 1 and 2 (judging membership likelihood of objects missing a feature) and manipulates the way in which the feature is missing: it is either explicitly missing or unavailable. The Explicit condition is similar to Experiments 1 and 2 in that the missing feature is explicitly said to be absent. (In abstract notation, a missing cause item is $\text{no-}X \rightarrow Y$ and a missing effect item is $X \rightarrow \text{no-}Y$ in the Explicit condition.) This situation is similar to a case in which we explicitly confirm by autopsy that a patient does not have any defect in her amygdala. As in Experiment 1, the causal status effect is expected such that $\text{no-}X \rightarrow Y$ would be less likely to be a member of the target category than $X \rightarrow \text{no-}Y$. In the Unavailable condition, participants are told that we do not have any available information about the presence of the target feature (i.e., $? \rightarrow Y$ and $X \rightarrow ?$). This situation would be like a case in which we do not yet know through autopsy whether a patient has any defect in her amygdala. In general, we hypothesize that $? \rightarrow Y$ and $X \rightarrow ?$ will be both inferred as $X \rightarrow Y$ and thus there will be no causal status effect. Because the causal status effect will occur only in the Explicit condition, we predict an interaction effect between the causal status of the missing feature (Cause or Effect) and the way information is missing (Explicit or Unavailable). Specifically, higher ratings on the missing-cause items in the Unavailable condition ($? \rightarrow Y$) than in the Explicit condition ($\text{no-}X \rightarrow Y$) would provide more direct evidence that people actually inferred the underlying cause in the Unavailable condition. In addition, we predict a main effect of the way in which information is missing such that the Unavailable condition will lead to overall higher likelihood judgments than the Explicit condition. If people infer missing features in the Unavailable condition (i.e., resulting in $X \rightarrow Y$ in

both missing-cause and missing-effect items), they will be perceived as having more of the target features than the objects in the Explicit condition, where no features could be inferred. This inference will lead to a better perceived match between the target object and the learned category.

Method

Participants. Sixteen undergraduate students at Yale University participated in this study either in partial fulfillment of requirements of an introductory psychology course or for payment of \$7.00 for participating in this experiment and other unrelated experiments.

Design and materials. Two variables were manipulated. The first variable was the Causal Status of the missing feature. Unlike in Experiments 1 and 2, only the first cause (Missing-cause) and the last effect (Missing-Effect) in the three-feature causal chain were manipulated. The second variable was the Missing Information condition, which refers to the manner in which information about missing symptoms was given. In half of the problems (the Unavailable condition), participants were told that “we do not have any information about whether” the new object has the target feature. In the other half of the problems (the Explicit condition), they were told that the new case “does not have” the target feature. Both factors were varied within participants to form four types of problems. Within each problem type, two problems were developed, one using novel diseases consisting of three symptoms and the other using novel artifact objects consisting of three components. Eight problems were developed by crossing the two levels of Causal Status of features, the two levels of Missing Information, and two levels of content materials.

In each problem, three characteristic features of a novel category were described (e.g., “If a person has symptoms K, T, and M, the person has Disease Ziso 75% of the time”; “If an object has components M, Q, and Y, the object is a jigraw 75% of the time”). Then participants learned how the three features were causally related (e.g., “the scientists have found that symptom K causes symptom T, and symptom T causes symptom M”; “The component M determines the operation of the component Q, and the component Q determines the operation of the component Y”). The three features always formed a single causal chain as in the above example.⁴ Then a question asking the membership likelihood of a novel case missing one of the three features was presented (e.g., “Suppose Kim has symptoms T and M, but we do not have any information about whether she has symptom K or not. How likely is it that Kim has Ziso?” for the Unavailable condition).

Procedure. All participants received all eight problems in a completely randomized order. The experiment was programmed using MacProbe (Hunt, 1994) and was presented on Power PC Macintosh computers.

Results and Discussion

As predicted, the causal status effect occurred only when the missing value was explicitly missing in the transfer items. In the Explicit condition, the difference between the Missing-Cause and the Missing-Effect problems was 15.3%, whereas in the Unavailable condition, this difference was only 1.3% (see Fig. 4 for the overall means of the four conditions). This same pattern of results was obtained for both the medical domain problems (the difference being 23.0% in the Explicit condition and 2.6% in the Unavailable condition)

⁴ Because the three features used in Experiment 5 were blank properties (e.g. symptom T), a control condition with no causal relations was not utilized in Experiment 5, as it was assumed that a priori centrality of the features would be approximately equal.

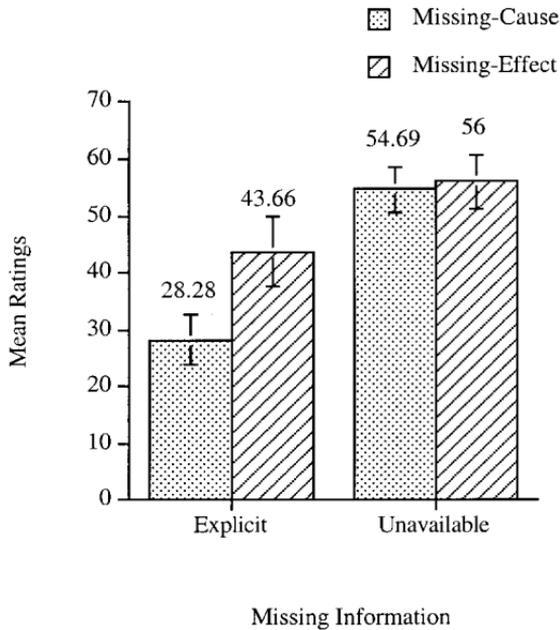


FIG. 4. Results of Experiment 5: The differential effect of causal status of features as a function of the type of missing features.

and the artifact problems (the difference being 11.5% in the Explicit condition and 2.0% in the Unavailable condition).

An ANOVA with Causal Status and Type of Missing Information (Explicit or Unavailable) as within-subject variables indicated that this interaction effect was significant, $F(1, 15) = 7.60$, $MS_e = 104.09$, $p < .05$. Planned comparisons indicated that in the Explicit condition, the ratings on the Missing-Cause items ($M = 28.3$, $SD = 17.8$) were significantly lower than the ratings on the Missing-Effect items ($M = 43.7$, $SD = 24.8$), $t(15) = -2.55$, $p < .05$, whereas in the Unavailable condition, there was no causal status effect, $p = .8$. The main effect of Causal Status was not statistically reliable, $F(1, 15) = 2.83$, $MS_e = 393.9$, $p = .11$, because the causal status effect occurred only in the Explicit condition. Finally, the hypothesis that the Unavailable condition would lead to overall higher likelihood judgments than the Explicit condition was supported. The mean rating in the Unavailable condition (55.4%) was reliably higher than the mean rating in the Explicit condition (35.9%), $F(1, 15) = 16.64$, $MS_e = 360.99$, $p < .001$. This result suggests that people indeed inferred missing features in the Unavailable condition. Planned t tests for dependent samples indicate that for both the Missing-Cause and the Missing-Effect items, the Unavailable condition was higher than the Explicit condition; $t(15) = 4.48$, $p < .001$, and $t(15) = 2.55$, $p < .05$, respectively.

Initially, we started out with an observation that seemingly contradicts the causal status hypothesis. That is, people seem to be confidently categorizing objects based only on surface features when the underlying cause is unavailable to them. However, we argued that this occurs because people inferred the presence of a causally related but hidden feature. If people in the above situation were categorizing objects based *only* on surface features, then there should not have been any difference between items with hidden cause (i.e., Missing-Cause item in the Unavailable condition) and items with explicitly missing cause (i.e., Missing-Cause item in the Explicit condition). It is because not inferring any underlying cause should be the same as the explicit denial of the underlying cause. However, our results showed that the Missing-Cause item led to much higher rating in the Unavailable condition than in the Explicit condition.

Another pattern of results from Experiment 5 is also consistent with our hypothesis. When the presence of the symptom was explicitly denied (i.e., in the Explicit condition), missing a cause led to a lower membership likelihood than missing an effect, replicating the causal status effect. However, when the presence of a missing feature was not explicitly denied (i.e., in the Unavailable condition), the causal status effect vanished, presumably because participants could make inferences about the presence of the features causally connected with other features. Furthermore, as predicted, the overall ratings in the Unavailable condition were higher than in the Explicit condition. These findings support our hypothesis that participants would treat transfer items in the Unavailable condition essentially as though neither causes nor effects were missing. Thus, Experiment 5 demonstrates evidence supporting the notion that people infer the presence of a causally related but hidden feature. Therefore, categorization in the absence of an explicit observation of ultimate causes is not a counterexample of the causal status hypothesis.

The current experiment only dealt with the type of situation in which people have specific causal knowledge. Sometimes, however, people might not know actual essential properties and simply hold beliefs that there are experts who really know what essences are (Medin & Ortony, 1989). For instance, people might categorize objects as trees based on surface features such as “has a trunk” or “has leaves” along with the belief that these objects also have “tree essence.” However, they cannot have any explicit causal knowledge about the tree essence because there is in some sense no biological essence for trees.⁵ In this case, we speculate that the inference about a tree essence (whatever that might be) came from the surface features. That is, surface features serve as evidence for the essence placeholder even if its

⁵ In fact, many trees are more closely related to other types of plants than they are to each other. We thank Arthur Markman for this example.

actual properties are unknown. Thus, we suggest that the underlying mechanism is the same as in the current experiment.

Finally, the current results also address an important methodological detail in measuring feature centrality. That is, unless the presence of a target feature is explicitly denied, we might not be measuring the centrality of that target feature because that feature's presence could be inferred through interproperty relations.

EXPERIMENT 6: EFFECT OF CONFLICTING CAUSAL BACKGROUND KNOWLEDGE

The final experiment examines a moderating factor for the causal status effect, namely plausibility of causal relations. In particular, this experiment investigates how compatibility between existing and newly acquired causal background knowledge affects the size of the causal status effect by way of influencing plausibility of causal relations among features. If a new piece of causal background knowledge conflicts with an already existing piece of causal background knowledge, it seems reasonable to expect that the causal status of the features involved will cancel out due to this conflict. The most extreme example of this type would occur when a person is told that feature W causes feature X, but the person does not accept this statement because it runs counter to his or her prior beliefs. In this case, there is no reason to expect the causal status effect. The converse is also likely. A person might be told that feature Y causes feature Z, a new piece of information which supports, rather than conflicts with, this person's previously formed belief system. This person, whose prior causal background knowledge does not conflict with new causal information, is more likely to exhibit the causal status effect than a person who has less trust in the newly given causal information. In Experiment 6, we attempt to investigate these two opposite situations by providing causal background knowledge that is generally consistent or inconsistent with laypeople's existing background knowledge.

The preexisting background information that Experiment 6 relies upon is a general lay theory about how the abnormal symptoms of diseases are related to germs or genetic defects. Although there might be a few exceptional diseases, a common sense notion is that in general, viral infections or genetic mutations usually cause abnormal symptoms rather than the other way around. The idea is that such a priori background information should determine the plausibility of new causal background information. Suppose a person learns that Virus XB12 causes Symptom Y, a new piece of information consistent with the person's preexisting naive theory (Canonical condition). Then, the two sources of consistent background knowledge together would elevate or augment the causal status effect. In contrast, suppose that although a person believes that viral infections cause abnormal symptoms, the person is instead told that Symptom Y causes infection with Virus XB12 (Reverse

condition). Such conflicting pieces of knowledge would lower the plausibility of the causal relations in the new piece of knowledge, thereby weakening the causal status effect.

Another way to view our predictions for Experiment 6 is based on, among other theories, Wellman's (1990) argument about the distinction between framework theories and specific theories. According to him, framework theories define the ontology of the domain and place limits on the kinds of information and causal mechanisms that specific theories can incorporate. For example, behaviorism is a framework theory, and the Rescorla–Wagner model is a specific theory (Rescorla & Wagner, 1972). Roughly drawing on this distinction, it can be suggested that causal relations among features in a specific category (e.g., Virus XB12 causes Symptom Y) are like specific theories, and people's existing naive theories about the medical domain are like framework theories. Just as framework theories constrain specific theories, we expect that people's general beliefs about disorders will affect the plausibility of the specific causal information provided during the experiment. As a result, we expect the causal status effect to be strengthened or weakened depending on whether this plausibility is strengthened or weakened, respectively.

Method

Participants. Forty-six undergraduate students at Yale University participated in this experiment in partial fulfillment of Introductory to Psychology course requirements.

Design and materials. Three novel diseases were used. (1) Disease Yorva was described to have one viral infection ("Virus XB12") and two symptoms ("low insulin level" and "shortness of breath"). (2) Disease Surpa was described to have one genetic mutation ("mutation on gene TCR alpha-1") and two symptoms ("swollen liver" and "internal hemorrhage"). It was assumed that people have a priori framework theories involving viral infection and genetic mutation. (This assumption was tested in a pretest that will be presented below.) Hence, these two diseases served as problems where specific theories can be manipulated to be consistent or inconsistent with respect to these framework theories. Details about the nature of these manipulations are described below. (3) Disease Xeno was described to have three symptoms: "blurred vision," "headache," and "insomnia." These three symptoms were selected in such a way that people would not have any consensual framework theories involving these symptoms. (Again, this assumption was tested in a pretest presented below.) For instance, it seems as reasonable to say that blurred vision might cause a headache as it is to say that a headache can cause blurred vision. Hence, Disease Xeno served as a control condition where no framework theories exist.

All of the characteristic features were described in a manner similar to that in Experiment 5. For instance, in Disease Yorva, participants were told, "Scientists have found that if a person is infected by Virus XB12, has a low insulin level, and has shortness of breath, the person has Disease Yorva 75% of the time."

In each of the three diseases, there were two causal relation conditions depending on the order in which the three features in each disease were laid out in a single causal chain. In the Canonical order condition, which was designed to conform to laypeople's framework theories, the virus served as the cause for other symptoms of Yorva (e.g., "Virus XB12 causes a low

insulin level, and a low insulin level causes shortness of breath”), and the genetic mutation served as the cause for other symptoms of Surpa (e.g., “A mutation on gene TCR alpha-1 causes a swollen liver, and a swollen liver causes internal hemorrhage”). For Xeno, where all features were symptoms, one symptom (blurred vision) arbitrarily served as the deepest cause for the other symptoms. In the Reverse order condition, the viral infection was an effect feature of other symptoms for Yorva (“shortness in breath causes a low insulin level, and a low insulin level causes infection of Virus XB12”), and the genetic mutation was an effect feature of other symptoms for Surpa (“internal hemorrhaging causes a swollen liver, and a swollen liver causes a mutation on gene TCR alpha-1”). Thus, this condition was designed to conflict with laypeople’s framework theories of diseases for Yorva and Surpa. For Xeno, the symptom that served as an effect feature in the Canonical condition (insomnia) served as the deepest cause for the other two symptoms.

To ensure that the Reverse versions of the genetic mutation and viral infection feature type scenarios were indeed more implausible to subjects than their respective Canonical versions, a separate pilot study was run. Sixteen undergraduate students at Yale University were presented with each of the six scenarios described above and were asked how much they believe that the kind of causal relations described in each scenario could occur in the real world. Participants were asked to select their response for each scenario on a scale of 1–7 (where 1 = *disbelieve very strongly* and 7 = *believe very strongly*). A 3 (Feature Type; Symptom, Virus, or Gene) \times 2 (Causal Order; Canonical or Reverse) ANOVA revealed that the interaction effect was significant ($p = .01$). Paired *t*-tests (two-tailed) were then run to compare ratings of the Canonical and Reverse conditions for each feature type. For the symptom feature type, the Canonical and Reverse versions did not differ with respect to plausibility (mean ratings of 5.1 and 4.8, respectively; $p = .6$). For the virus and gene feature types, the Canonical version was significantly more plausible to participants than its corresponding Reverse version (for virus, mean ratings of 4.4 and 2.6, respectively; $p = .002$; for gene, mean ratings of 5.3 and 2.6, respectively; $p < .001$). Thus, the only scenarios that participants found implausible were the Reverse versions of the virus and gene feature types.

The final independent variable was the causal status of missing features. For each order condition in each disease category, two questions were asked as in the Explicit condition of Experiment 5: Missing-Cause and Missing-Effect. Again, participants’ answers to these questions served as the critical dependent measure.

To summarize, the experiment was a 3 (Feature Type: Symptom, Virus, or Gene) \times 2 (Causal Order; Canonical or Reverse) \times 2 (Missing Feature: Missing-Cause or Missing-Effect) factorial design. Feature Type and Missing Feature were within-subject variables, and Causal Order was a between-subjects variable because the same sets of features were used for the two causal order conditions.

Procedure. Each participant received six problems (i.e., Missing-Cause and Missing-Effect problems from each of the three conditions of Feature Type) in a completely randomized order. In each problem, they first learned three characteristic symptoms of a new disease and their causal relations as described in the material section. Afterward, they were presented with either a Missing-Cause or a Missing-Effect question as in Experiment 5. Twenty-three randomly selected participants were assigned to the Canonical condition, and another randomly selected 23 were assigned to the Reverse condition.

Predictions. In the Canonical condition, the causal status effect was predicted to occur in all three conditions of Feature Type (symptom, gene, and virus). In addition, it was predicted that compared to the control condition involving three symptoms, the causal status effect would be even stronger in the virus and the gene conditions, where a priori framework theories exist and are consistent with specific theories. In the Reverse condition, however, the causal status effect was predicted to occur only in the symptom condition, where no framework theory exists and therefore no conflict with specific theories can occur. For the virus and gene condi-

tions, however, the causal status effect was expected to disappear due to the conflict between the framework theory and the specific theory.

Results and Discussion

The results are summarized in Fig. 5, broken down for the Canonical condition and the Reverse condition. In the Canonical condition, the mean likelihood for the Missing-Cause problems was much lower than that for the Miss-

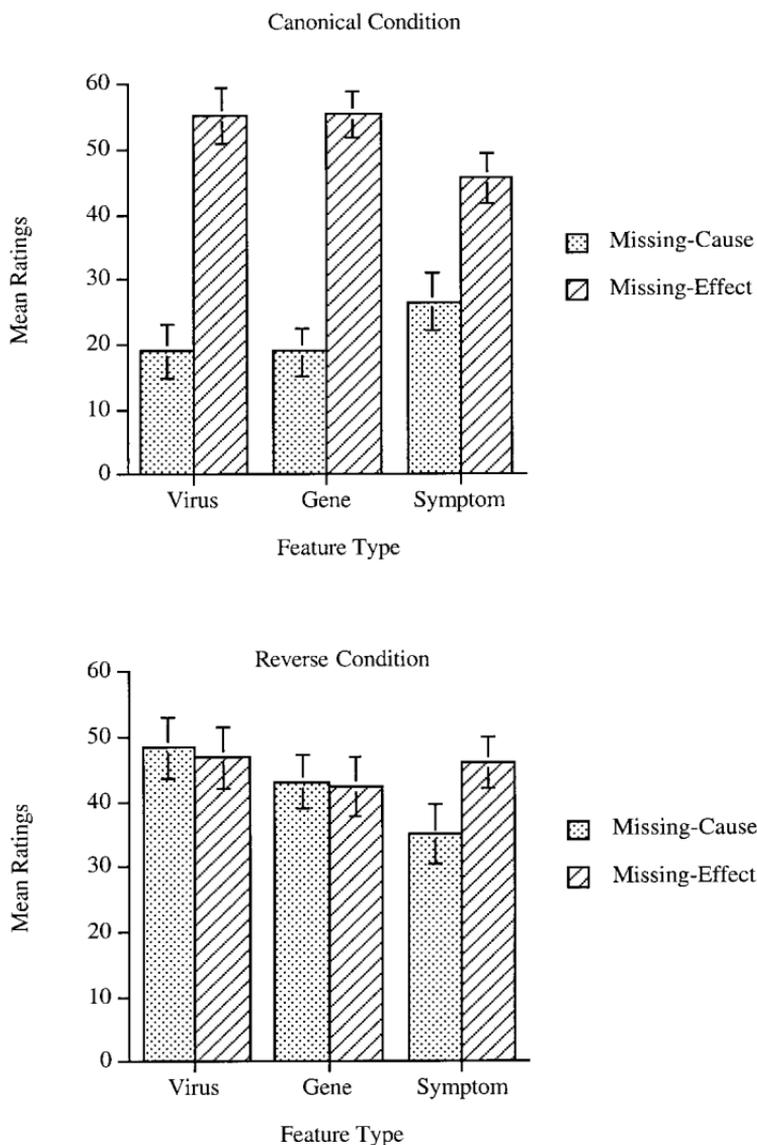


FIG. 5. Results from Experiment 6.

ing-Effect problems across all three types of features. Furthermore, the causal status effect was much stronger when the cause feature was a viral infection or a genetic mutation, as is generally the case in real-life diseases. More specifically, in the Canonical condition, the difference between the Missing-Cause and the Missing-Effect problems was 36.1% for the Virus feature type and 36.6% for the Gene feature type, but only 19.0% for the Symptom feature type. In the Reverse condition, the causal status effect occurred only in the Symptom feature type (11.1% difference between the Missing-Cause and the Missing-Effect). There was virtually no difference between the Missing-Effect and the Missing-Cause conditions when the virus infection or the genetic mutation, both of which ordinarily serve as a cause for other symptoms in a disease, served as an effect feature in the experimental materials.

ANOVA analyses support the reliability of the all of the above descriptions. A $3 \times 2 \times 2$ ANOVA with Feature Type and Missing Type as within-subject variables and Causal Order as a between-subject variable showed a reliable three-way interaction effect, $F(2, 88) = 7.229$, $MS_e = 231.4$, $p < .01$. That is, in the Canonical condition, the Virus and Gene problems led to a stronger causal status effect than the Symptom problem did, whereas in the Reverse Condition, the Virus and Gene problems did not show any causal status effect, though the Symptom problem did. Overall, there was a reliable main effect of Causal Status with Missing-Cause being much lower than Missing-Effect, $F(1, 44) = 28.0$, $MS_e = 691.9$, $p < .001$. In addition, there was a reliable two-way interaction effect between Causal Order and Missing Type, $F(1, 44) = 18.9$, $MS_e = 691.9$, $p < .001$. This interaction demonstrates that the ratings from the Missing-Cause problems in the Canonical condition were much lower than those in the Reverse condition, whereas the ratings from the Missing-Effect problems did not differ between the two conditions.

Two separate ANOVA's were conducted for the Canonical and Reverse conditions, with Missing Type and Feature Type as within-subject variables. In the Canonical condition, regardless of Feature Type, ratings for Missing-Cause were lower than those for Missing-Effect by 30.5%, $F(1, 22) = 41.3$, $MS_e = 779.1$, $p < .001$. In addition, there was a reliable interaction effect, $F(2, 44) = 4.1$, $MS_e = 285.6$, $p < .05$. As discussed before, this interaction effect was due to the fact that the causal status effect was stronger in the Virus and Gene problems than in the Symptom problem. In the Reverse condition, the only reliable effect was an interaction effect, $F(2, 44) = 3.2$, $MS_e = 177.2$, $p < .05$. This was found because the causal status effect occurred only in the Symptom condition, as indicated by a planned t test comparing the Missing-Cause and Missing-Effect ratings in the Symptom condition, $t(21) = 2.7$, $p < .05$.

To summarize, Experiment 6 demonstrated that the causal status effect changes dynamically as a function of compatibility between existing knowl-

edge and newly added knowledge. When the experimentally provided causal relations were consistent with an already existing naive theory, rendering them highly plausible, the causal status effect was the strongest. When the experimentally provided causal relations were implausible because they conflicted with an already existing naive theory, the causal status effect disappeared. Finally, when the experimentally provided causal relations were neutral with respect to a naive theory (i.e., symptom conditions), the strength of the causal status effect was in between that of the above two situations. This evidence that the amount of the causal status effect varies as a function of plausibility of causal background knowledge also suggests that the effect observed in the current article is indeed the effect of causal background knowledge, rather than other possible factors, such as demand characteristics or confounds in the discourse saliency of features in the stimulus materials.

GENERAL DISCUSSION

We propose that a cause feature is more conceptually central than its effect feature. Using a variety of measures, the current study has demonstrated for the first time that this causal status effect is deeply rooted in many aspects of categorization processes.

First, participants learned novel categories with characteristic features and judged the membership likelihood of transfer items explicitly missing one of the features (Experiments 1, 6, and the Explicit condition in Experiment 5). This task revealed a robust effect of causal status in that missing cause features lowered the membership likelihood ratings decidedly more than missing effect features.

Second, in Experiment 3 and the categorization condition of Experiment 4, participants were asked to create categories from three unclassified objects. The results showed that when free to sort objects in any way they like, people prefer to create categories based on matching causes rather than on matching effects. Thus, these results reveal the beliefs that people have about how concepts should be structured; they prefer that members in the same category share the same underlying cause rather than effect. As we discussed earlier, this belief is consistent with the idea of psychological essentialism.

Third, in Experiment 2, the causal status of features affected goodness-of-exemplar judgments, such that the deeper the cause that an item was missing, the worse exemplar it was judged to be. Thus, the results suggest that the causal status of features is responsible for one of the most important constructs in the categorization literature, namely the typicality effect. Future research may determine whether the causal status effect will derive typicality effects manifested in reaction-time differences between typical and atypical items.

Fourth, while the causal status effect has been observed in similarity judg-

ments (Ahn & Dennis, 1997) and also in typicality judgments as shown in Experiment 2, it was found to be less robust in similarity judgments than in categorization judgments (Experiment 4). This result supports previous demonstrations of dissociation between categorization and similarity judgments using the transformation paradigm (e.g., Rips, 1989) and suggests that strong versions of similarity-based categorization models may fall critically short of accurately representing human categorization.

In what follows, we discuss other important findings of the current study and open questions. In addition to demonstrating the prevalence of the causal status effect as summarized above, the current study also found that the effect of causal status on features is continuous along a causal chain. We first discuss the implications of this finding for psychological essentialism. Second, we further discuss the implications of the current findings for both the similarity-based approach and the theory-based approach to categorization. Third, the current study also examined conditions under which the causal status effect should be expected to disappear. We also discuss other possibilities and potential counterexamples. Finally, we discuss open questions, including whether the size of the effect might vary across domains and whether the causal status effect is a special case of a more general phenomenon.

Continuous Feature Centrality and Its Implications for Essentialism

The current study provided evidence that features in a causal chain fall along a continuum of centrality as a function of their causal status in the chain. Experiments 1 and 2 showed that the first cause was judged to be more central than its effect feature which, in turn, was judged to be more central than its own effect feature. That is, even among features that are caused by the most fundamental cause in a category, centrality of features varied as a function of their causal status.

As discussed earlier in the introduction, this finding contrasts with a strong version of essentialism which argues that essential properties are necessary and sufficient for categorization and that nonessential properties should not determine categorization. According to this view, features are dichotomized into essential features and nonessential features. In contrast, Experiments 1 and 2 found that even features that are not the deepest cause in a causal chain affect membership likelihood and typicality ratings. Furthermore, instead of revealing a dichotomy between essential and nonessential features, the current study found that feature centrality lies in a continuum.

Yet, it should be remembered that essentialism is critically related to the causal status hypothesis in that both concern the causal potency of essential (or central) features. Furthermore, if essences are the deepest cause in a category, both essentialism and the causal status hypothesis acknowledge the special status of essences in that they are the most central features in the category.

Implications for Categorization Literature

Implications for the similarity-based approach and computational modeling of the causal status effect. As discussed in the introduction, one of the serious problems of the similarity-based models of categorization has to do with the lack of a specific mechanism that can predict constraints in feature weighting. Although the similarity-based approach has successfully developed many formal models, these models tend to be silent about how to determine the feature weighting as a function of background knowledge. The causal status hypothesis provides one well-defined way of constraining feature weights within the framework of the theory-based approach to categorization. An example of computational implementation of the causal status hypothesis comes from Sloman et al. (1998). This model is based on general “dependency” relationships rather than being restricted only to causal relationships. The idea of the model is that the more other features depend on (e.g., are caused by, are determined by, are followed by, etc.) a feature, the more central this feature becomes to the concept. More specifically, the model states that conceptual centrality or immutability of a feature (C) is a function of dependency as follows:

$$C_{i,t+1} = \sum_{ij} A_{ij} C_{j,t} \quad (1)$$

where A_{ij} is the strength of a dependency link from feature j to feature i . This formula states that the centrality of feature i is determined at each time step by summing across the immutability of every other feature multiplied by that feature’s degree of dependence upon feature i . The results from the experiments reported here are consistent with the centrality measures predicted by this model. For instance, suppose feature X causes feature Y, which causes feature Z, and the causal strengths of both relations are 3, and the initial value for feature centrality was an arbitrary value of, say, 1. After two iterations, the centralities of features X, Y, and Z become 16, 7, and 1, respectively. These qualitative differences in feature centrality predicted by the model are consistent with the results found in Experiments 1 and 2.

Of course, this model is just one of many possible ways of computationally modeling the causal status effect (see also Ahn & Kim, 2000 for data that this particular model cannot account for). For now, we simply wish to illustrate that it is possible to computationally model the effect of background knowledge on categorization. Hence, one way of extending an existing model of similarity-based categorization to deal with the problem of feature weighting might be to add a component similar to the above function as an additional module.

Implications for the theory-based approach to categorization. The most important implication of the causal status hypothesis for the theory-based approach to categorization is that it offers a specific mechanism by which

feature weighting based on background knowledge might occur beyond merely showing that feature weights may change due to background knowledge. At various places in the article, we have already explained, in terms of our hypothesis, some existing studies showing the effect of domain theories. Referring to Medin and Shoben's study (1988), we argued that curvedness is a central feature in boomerangs but not in bananas because of its differing causal status within each category. In Experiment 4, we also discussed how Rips's study (1989) on the dissociation between similarity and categorization can be explained by assuming different degrees of the causal status effect between the two tasks. We also briefly discussed that studies showing the importance of intentionality in naming (Bloom & Markson, 1998; Gelman & Ebeling, 1998) are consistent with our hypothesis.

Continuing in this broader context, we speculate that a number of existing findings in the literature may be viewed as having been mediated by a causal status effect to some degree (see Ahn & Kim, 2000 for a more extensive review).

For example, Ross (1997) found that using category knowledge to perform a nonclassification task resulted in changes in feature weighting. In a series of experiments, participants learned to categorize hypothetical patients as having one of two novel diseases on the basis of four relevant symptoms that were predictive of each disease. During this training phase, participants also learned to prescribe one of two treatments for each of the diseases on the basis of two relevant symptoms for each disease. Ross found that learning how to use the categories to prescribe treatment affected later classification decisions, such that symptoms relevant to both categorization and prescription of treatment were more accurately categorized with the appropriate disease than symptoms relevant only to categorization. This effect occurred even though all four symptoms were in fact equally predictive of the disease. This "category use effect" may be seen as an instance of the causal status effect because participants might have assumed that a treatment that is able to cure a disease is generally most likely to act upon the causal symptoms of the disease, not the peripheral effects. For example, one would not expect an ear infection caused by bacteria to be treated effectively with medicines that simply suppress symptoms (e.g., painkillers). Instead, to cure the ear infection, antibiotics that act directly upon the bacteria should be given. Moreover, it may be that treatment-relevant symptoms were considered by Ross's participants to be causally central in a broader sense in that they determined which treatment plan to adopt. Therefore, the fact that Ross's participants felt that symptoms predictive of treatment were most important in diagnosing the disease can be traced to the experiment's implication that these symptoms are the most causal.

In addition, the causal status hypothesis appears to be compatible with findings involving the use of categories in problem solving. For example, Chi, Feltovich, and Glaser's (1981) classic experiment assessed the categori-

cal representations of physics problems in advanced physics graduate students and professors (experts) and undergraduates who had just completed a semester-long course in mechanics (novices). They found that experts consistently based their categorization of physics problems on deeper physics principles, such as conservation of momentum or the work–energy theorem, whereas novices categorized based on the surface features of the problems, such as whether there was an inclined plane or pulley in the problem. Their findings seem to be related to the causal status hypothesis in that both indicate a tendency to not categorize based on surface features when knowledge about a deeper structure is available. That is, experts in Chi et al.'s (1981) study, like the participants in the current study, were aware of both the surface features and the deeper structure, but preferred to categorize based on the latter. It should be noted, however, that in Chi et al.'s study, participants were specifically asked to sort the problems “based on similarities of solution” (p. 124). Thus, it remains to be seen whether they would sort problems by deeper properties (i.e., deeper properties within their understanding) even without these instructions. Given our results from Experiment 3, we would predict that both experts and novices will spontaneously sort problems based on the deepest cause known to them.

Finally, a number of developmental studies also appear to be consistent with the causal status hypothesis. Earlier, we briefly described Keil's study (1989) in which participants were presented with a description of a raccoon painted to look like a skunk. In this study, even young children thought that the origins of animals (e.g., being born from another raccoon) are more important than perceptual appearance (e.g., black with a white stripe on the back) in determining category membership. This result can be interpreted as demonstrating that cause features (origins of animals) are perceived to be more central than effect features (appearance of animals). Similarly, Gelman's (1998) study can be viewed from the perspective of the causal status effect. In this study, children first learned a novel feature for each type of category (e.g., this rabbit has a spleen inside) and were asked whether this feature is generalizable to another instance of the same category. Second graders responded that features referring to substance and internal structure (e.g., “has a spleen inside,” p. 74) were more generalizable for natural kinds, whereas functional features (e.g., “you can loll with it,” p. 75) were more generalizable for artifacts. As Ahn (1998) later demonstrated empirically, this pattern of results might have occurred because substance and internal structure are more causally central in natural kinds, whereas functional features are more causally central in artifacts.

For the sake of parsimony alone, it would certainly be advantageous to explain all of these studies with this single mechanism, the causal status hypothesis. At this point, however, it remains an open question whether all of these phenomena are indeed instances of the causal status effect. More systematic studies need to be conducted by pitting alternative factors against

the causal status of features in order to determine which is a more fundamental mechanism.

Moderating Factors for the Causal Status Effect

Although the causal status effect is robust, it is important to specify its boundary conditions. The current studies examined two situations in which the causal status effect should disappear. This section gives a summary of the results from those experiments, followed by discussions of how other determinants of feature centrality might interact with the causal status effect, whether there are any counterexamples to the causal status effect, and whether domain can moderate the causal status effect.

Unknown causes and plausibility of causal beliefs. Experiment 5 examined the situation in which the presence or absence of a cause is unknown. Although the cause is missing in some sense in this case, membership likelihood judgments were not lowered more than when the presence or absence of an effect feature is unknown. This result most likely occurred because people inferred the presence of the unknown cause based on the given causal background knowledge. Furthermore, Experiment 6 showed that background knowledge from multiple sources can interact in such a way that a feature that is a cause in a specific theory might not be causal in a framework theory, resulting in cancellation of the causal status effect. That is, when newly learned causal background knowledge conflicts with existing causal background knowledge, the causal status effect does not appear.

Other determinants of feature centrality. In addition, it is likely that other determinants of feature centrality may interact with the causal status effect. Before we can discuss how they might interact, it is necessary to first discuss different types of feature centrality and describe studies examining their dissociability.

Sloman et al. (1998) proposed several different kinds of feature centrality, including conceptual centrality (mutability, or the degree to which a feature in a concept can be transformed while maintaining the concept's coherence), category centrality (perceived relative frequency of an instance within a category), diagnosticity (evidence provided by a feature for one category relative to a set of categories), and prominence (how prominent the feature seems to people when thinking about the category). Among these, the type of feature centrality considered in the current study is closest to conceptual centrality. Sloman et al. (1998) demonstrated that these different kinds of feature centrality are empirically dissociable. In addition, studies have shown that conceptual centrality is the only type of centrality that reliably correlated with a feature's status in a dependency structure of concepts [as measured by Eq. (1) above] across various levels of categories (Ahn & Sloman, 1997; Sloman & Ahn, 1999). A general picture emerging from these results is that the causal status effect is likely to be most pertinent to tasks involving conceptual centrality and may be less influential in tasks such as perceptual categoriza-

tion (which would be determined by perceptual prominence) or discrimination tasks (which would be determined by diagnosticity of features).

Although the above studies showed that it is possible to demonstrate independence among different determinants of feature centrality, there are a number of reasons to expect that they might sometimes interact with each other. First, causal background knowledge, which determines conceptual centrality, can affect the perception of probabilities (e.g., category or cue validities). For instance, a feature that is considered more conceptually central might be mistakenly perceived to be more frequent (e.g., Sloman et al., 1998; Spalding & Ross, 1994). Second, the above studies demonstrated that different determinants of feature centrality are empirically dissociable when tasks are carefully chosen such that each taps only on a single type of feature centrality. Real-life categorization tasks, however, tend to be a mixture of different types of categorization. Under such circumstances, different determinants of feature centrality can simultaneously act on categorization. For instance, a feature with high causal status but low category validity might end up having mediocre feature centrality.

In addition, Gentner and her colleagues (e.g., Gentner, 1989) have suggested that relational features are more important than isolated features in analogical reasoning (see also Lassaline, 1996 for its extension to category-based induction). If the number of relations in which a feature participates also determines feature centrality, then in a structure where multiple causes have a single effect in common (the so-called common-effect structure), the effect feature can outweigh one of the cause features because the effect feature participates in more relations than any one of these cause features. (See also Ahn & Kim, 2000, for more discussion on the causal status effect in a common-effect structure especially in relation to Rehder and Hastie, 1997).

Counterexamples to the causal status effect. In the previous sections, we considered three cases in which the causal status effect might disappear and have argued that none of them is a true exception to the causal status hypothesis. First, an effect feature can be useful in retrieving information about non-obvious causes (Experiment 5). Thus, when a cause feature is not explicitly denied, the presence of an effect feature alone can be sufficient for categorization. This does not counter the causal status hypothesis because it does not show that an effect feature is more central than its cause. Second, as examined in Experiment 6, a feature might serve as a cause for another feature in one causal relationship, but at another level, the causal direction might run in the opposite direction. In that case, a feature's causal status would be cancelled out or even reversed at times, and changes in feature centrality should be expected as a result. Third, when other determinants of feature centrality heavily favor the effect feature over its cause, it is reasonable to expect that these multiple influences can override the causal status effect. In particular, this third point is useful in accounting for examples that may seem

to contradict the causal status hypothesis. We will discuss one such example below.

Pneumonia⁶ is a serious infection or inflammation of the lungs, and it can have over 30 different causes, such as bacteria, virus, chemicals, or even inhaled objects like peanuts or small toys. In determining whether a person has pneumonia, the cause for lung inflammation is less important than lung inflammation itself.⁷ As discussed above, however, feature centrality may be determined by other factors, such as category validity and the number of relations in which a feature participates, if the task calls on those other factors. In the case of pneumonia, the effect feature has a category validity of 1 (i.e., all patients with pneumonia have lung inflammations), whereas each of the cause features has a much lower category validity because there are many possible causes for pneumonia. For the feature centrality of each of these causes, the causal status effect is cancelled out because of its low category validity. Furthermore, when there are multiple causes for the same effect as in this example, each causal strength is weakened because each of these causes is not a necessary condition for the effect. Therefore, conceptual centrality predicted by Eq. (1) should be lowered due to a low A_{ij} . Finally, in the case of the common-effect structure, the effect feature has more advantage over the cause features in terms of the number of relations in which it participates (e.g., Gentner, 1989).

Due to all these counteracting forces acting upon the cause features in a common-effect structure (low category validity, low causal strength, and fewer relations to participate in), the centrality of one of these causes in a common-effect structure can end up becoming lower than the centrality of its effect feature. Thus, a failure to observe the causal status effect in common-effect structures as in the above situations is not a counterexample to actual occurrence of the causal status effect because in these situations there are unbalanced counteracting forces. We simply suggest that the causal status effect occurs in all situations, but like any other psychological construct, it cannot be measured when multiple counteracting forces are bombarded against it. Cause can be empirically documented as more central than effect when everything else is roughly equal.⁸

⁶ We thank Frank Keil and Charles Kalish (in the context of "syndromes") for posing this question.

⁷ In diagnosing pneumonia, causes for lung inflammation are also crucial information because they determine different treatment plans. However, in simply determining whether someone has pneumonia, causes would not be as critical as lung inflammation per se because there are many potential causes while lung inflammation is a necessary feature for pneumonia.

⁸ One might wonder how prevalent such situations would be in real-life categorization. A number of studies examining the causal status effect in real-life categories have shown that about 50% of the variance in feature centrality can be accounted for by the causal status of features. Ahn (1998) has shown a correlation of .73 between conceptual centrality and causal centrality using Barton and Komatsu's (1989) stimulus materials (5 artifacts and 5 natural

What is particularly interesting about the case of pneumonia is that the causal status hypothesis is supported when these counteracting forces are removed. Indeed, the American Lung Association (1998) states explicitly that "pneumonia is not a single disease" precisely because there are many causes. Instead, they break the disease down into various types such as bacterial pneumonia, viral pneumonia, mycoplasma pneumonia, and so on based on the type of cause, just as the causal status hypothesis would predict. Furthermore, note that each subtype no longer has a common-effect structure, and within each subtype, the category validity of the cause feature and the effect feature becomes equal. For instance, the probability of *having lung infection* given that a person has bacterial pneumonia is the same as the probability of *being infected with pneumonia bacteria*, given that a person has bacterial pneumonia (i.e., 1). In addition, both of these two features participate in the same number of relations. Now that all the other counteracting forces are removed, the cause feature (e.g., being infected with pneumonia bacteria) becomes more central than the effect feature (e.g., having lung infection). For instance, a patient infected with pneumonia bacteria, but who has not yet developed lung inflammation (missing effect), although difficult to be declared as having bacterial pneumonia, is at least more likely to be considered as specifically having bacterial pneumonia than a patient who has lung inflammation from inhaling a peanut (missing cause).

As a final point to our discussion of pneumonia as a counterexample to the causal status hypothesis, it might be argued that in medical domains, symptoms are frequently thought of as effects and, therefore, any classification based only on symptoms is interpreted as categorization based on effects. However, as noted at various places throughout this article, the causal status of a feature is a relative notion. Therefore, it does not make sense to simply state that *X* is a cause feature or an effect feature; it is a cause of one feature but at the same time, it is an effect of another feature. Symptoms in diseases can cause some other feature (e.g., a stuffy nose can cause a headache), in which case they can be more central than their effect features. Going back to our previous example of pneumonia, lung inflammation is a symptom of pneumonia, but it causes other symptoms too. Because of lung inflammation, oxygen has trouble reaching the blood and, as a consequence, body cells cannot work properly. From this causal chain, we would expect that a patient who has lung inflammation but does not yet show other symptoms caused by it would be more likely to be diagnosed with pneumonia

kinds) and a correlation of .74 using Malt and Johnson's (1992) stimulus materials (12 artifacts). Kim and Ahn (in preparation) also found a high correlation of .73 between causal centrality and conceptual centrality in laypeople's concepts of four psychological disorders. Yet, because it is difficult to say whether these studies randomly selected representative sets of features and categories from real-life situations, more systematic studies are needed.

than a patient who does not have enough oxygen in the blood but does not show lung inflammation. If not, that would count as a counterexample to the causal status hypothesis. Simply showing that a symptom carries weight in a diagnosis judgment does not mean much about whether the causal status hypothesis is right or wrong. In testing the hypothesis, the question to ask is whether that symptom is more critical than its cause or effect. Whether there are any true exceptions to the causal status hypothesis, therefore, remains as a question for future research.

Domain generality. Another potential moderating factor for the causal status effect is domain. Although Putnam (1975) argues that all kinds (both natural and nominal kinds) have essences, Schwartz (1979) claims that for nominal kinds, there is no real essence. Schwartz (1979) argues that nominal kinds, such as “white things,” are conventionally established, and, therefore, if the criterial properties change (e.g., if an object is stained with mud), it no longer belongs to the category. Indeed, Kalish (1998) showed that adults as well as children understand natural kinds as being discovered in the world, whereas artifact kinds, which are similar to nominal kinds in that they are conventionally defined, are understood as being constructed. More importantly, Diesendruck and Gelman (1999) found more essentialist, all-or-none categorization from natural kinds than from artifacts. Thus, if psychological essentialism is responsible for the causal status effect, it might be predicted that differences in presuppositions about essences might affect the degree to which the causal status effect occurs in nominal kinds.

The results obtained so far, however, are somewhat at variance with that prediction. The six experiments reported in the current article utilized stimulus materials from various domains, including diseases and symptoms, artifacts, natural kinds, and social situations. The causal status effect was observed in all these domains. Ahn (1998) specifically investigates the causal status effect across natural kinds and artifacts and across different kinds of features (e.g., molecular and functional). As discussed earlier in the article, previous studies (e.g., Barton & Komatsu, 1989) reported the apparent finding that the centrality of features is determined by whether an object is a natural kind or an artifact (i.e., molecular features are more central for natural kinds, whereas functional features are more central for artifacts). Ahn (1998) argued that such domain differences were obtained because the causal status of features was confounded with conceptual centrality in those materials. For instance, molecular features are conceptually central for natural kinds but at the same time, they are causally central. Ahn (1998) used novel natural kinds and artifacts for tasks similar to the ones used in the current study, with the causal status of features being directly manipulated. These experiments demonstrated that molecular features (e.g., what an object is made of) and functional features (e.g., what an object is used for or what an object does) are weighted more heavily in both natural kinds and artifacts if these features

serve as causes for other features than if they serve as mere effects. That is, when the causal status of features was held constant, domain differences between natural kinds and artifacts disappeared.

Yet, we add cautionary remarks about the domain generality of the causal status effect because it has not yet been systematically tested across a wide variety of domains by using randomly selected categories and features. Whereas it might be possible to obtain the causal status effect across various domains as shown in the current study, the extent to which the effect occurs might vary, especially within the domain of nominal kinds (e.g., Schwartz, 1979). In particular, free-sorting tasks (or creating new categories) might be susceptible to domain differences. For natural kinds (i.e., kinds that naturally occur in the world), new categories seem to be created as deeper underlying causes are revealed. For instance, to laypeople, pneumonia might be a single disease because they do not know that there are multiple possible causes for it, but experts do not treat pneumonia as having a single cause, as discussed earlier. In contrast, in creating nominal kinds, one can choose any criteria and ignore underlying causes because nominal kinds are by definition conventionally fixed. For instance, the selection criteria for new orchestra members usually focus on quality of playing (an effect) rather than the details of how the person was trained (a cause). A category of an orchestra is a nominal kind, and the creator of the category can intentionally set up the criteria to do that.

Must the Relations Be Causal?

Finally, we discuss whether the causal status effect is a special case of a more general phenomenon. Throughout the current study, we have focused only on causal relationships. What about other kinds of relations such as “depends on” or “temporally follows from”? In Sloman et al. (1998), participants were explicitly told that symptom A, for example, does not cause symptom B but that symptom B follows symptom A (temporal dependency) or that the presence of symptom B depends on the presence of symptom A (contingency; e.g., the presence of moustache is contingent upon the presence of mouth). The results showed that temporally preceding features or features on which other features are dependent were judged to be more conceptually central.

What are the implications of these results for the causal status effect? There are at least two possibilities. First, these results might have occurred because participants imposed complex causal interpretations on the temporal dependency and contingency relations (e.g., symptom A might not directly cause symptom B, as was specified, but it might indirectly cause symptom B). In this case, the effect of dependency structure can be viewed as a special case of the causal status effect. Second, it could be that the causal status effect is a special case of a more general phenomenon occurring in any kind of asymmetrical dependency structure. This second possibility does not

threaten the present claim that cause features are more central than their effect features because the key ideas converge. Furthermore, even if the causal status effect is a special case of a more general phenomenon, it nonetheless appears to be a major portion of that general phenomenon, as indicated by studies showing that causal relations alone can account for a large amount of variance in feature centrality for natural categories (e.g., Ahn, 1998; Ahn & Kim, 2000). Indeed, causal relations are prevalent and essential components of relations that features have in our conceptual representations (e.g., Carey, 1985; Wellman, 1990).

Conclusion

Causal reasoning and categorization are two of the most fundamental reasoning processes. They are prevalent in everyday reasoning, and they govern and moderate other cognitive processes. In some sense it is no surprise that these two processes influence each other. The important question is precisely how they affect each other. In expanding previous categorization theories to include the role of causal background knowledge, the present experiments have shown one specific mechanism by which causal knowledge affects categorization: by determining feature centrality.

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., & Dennis, M. (1997). Dissociation between categorization and similarity judgment: Differential effect of causal status on feature weights. In *Proceedings of the Interdisciplinary Workshop on Similarity and Categorization*, University of Edinburgh, 1–7.
- Ahn, W., & Kim, N. S. (2000). The role of causal status of features in categorization: An overview. In D. L. Medin (Ed.), *Psychology of learning and motivation*. San Diego: Academic Press.
- 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., & Sloman, S. (1997). Distinguishing name centrality from conceptual centrality. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, (pp. 1–7). Mahwah, NJ: Erlbaum.
- American Lung Association (1998). Pneumonia [On-line]. Available: <http://www.lungusa.org/diseases/lungpneumoni.html>.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Barsalou, L. (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.
- Barton, M. E., & Komatsu, L. K. (1989). Defining features of natural kinds and artifacts. *Journal of Psycholinguistic Research*, **18**, 433–447.
- Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition*, **60**, 1–29.

- Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science*, **9**, 200–204.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, **11**, 177–220.
- Brewer, W. F., & Treynens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, **13**, 207–230.
- Braisby, N., Franks, B., & Hampton, J. (1996). Essentialism, word use, and concepts. *Cognition*, **59**, 247–274.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Plenum.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121–152.
- Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science*, **15**, 89–132.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, **25**, 257–271.
- Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin & Review*, **6**, 338–346.
- Fugelsang, J. A., & Thompson, V. A. (in press). Strategy selection in causal reasoning: When beliefs and covariation collide, *Canadian Journal of Experimental Psychology*.
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, **20**, 65–95.
- Gelman, S. A., & Ebeling, K. S. (1998). Shape and representational status in children's early naming. *Cognition*, **66**, B35–B47.
- Gelman, S. A., & Kalish, C. W. (1993). Categories and causality. In R. Pasnak & M. L. Howe (Eds.), *Emerging themes in cognitive development* (Vol. 2). New York: Springer-Verlag.
- Gelman, S. A., & Medin, D. L. (1993). What's so essential about essentialism? A different perspective on the interaction of perception, language, and conceptual knowledge. *Cognitive Development*, **8**, 157–168.
- Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the nonobvious. *Cognition*, **38**, 213–244.
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge, UK: Cambridge Univ. Press.
- Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, **65**, 263–297.
- Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, **52**, 125–157.
- Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects*. New York: Bobbs-Merrill.
- Guedeney, A. (1997). From early withdrawal reaction to infant depression: A baby alone does exist. *Infant Mental Health Journal*, **18**, 339–349.
- Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, **65**, 127–165.
- Hunt, S. M. J. (1994). MacProbe: A Macintosh-based experimenter's workstation for the cognitive sciences. *Behavior Research Methods, Instruments, & Computers*, **26**, 345–351.
- Imai, S., & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, **69**, 596–608.

- Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory & Cognition*, **23**, 335–353.
- Kalish, C. W. (1998). Natural and artifactual kinds: Are children realists or relativists about categories? *Developmental Psychology*, **34**.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition*, **65**, 103–135.
- Kim, N. S., & Ahn, W. (1999, November). The influence of naïve causal theories on lay diagnoses of mental illnesses. Poster session presented at the 40th annual meeting of the Psychonomic Society, Los Angeles, CA.
- Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, **19**, 1–53.
- Kripke, S. (1972). Naming and necessity. In D. Davidson & Harman (Eds.), *Semantics of natural language*. Dordrecht: D. Reidel.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22–44.
- Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, **124**, 161–180.
- Lamberts, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**, 695–711.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, **3**, 299–321.
- Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22**, 754–770.
- Locke, J. (1894/1975). *An essay concerning human understanding*. Oxford, UK: Oxford Univ. Press.
- Malt, B. C. (1994). Water is not H₂O. *Cognitive Psychology*, **27**, 41–70.
- Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory and Language*, **31**, 195–217.
- Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science*, **21**, 373–418.
- Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, **44**, 1469–1481.
- Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General*, **117**, 68–85.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, **100**, 254–278.
- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–196). New York: Cambridge Univ. Press.
- Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, **20**, 158–190.
- Murphy, G. L. (1993). A rational theory of concepts. In G. V. Nakamura, D. L. Medin, & R. Taraban (Eds.), *The psychology of learning and motivation* (Vol. 29, pp. 327–359): Academic Press.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289–316.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology*, **64**, 104–114.

- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39–57.
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E., (1990). Category based induction. *Psychological Review*, **101**, 185–200.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353–363.
- Putnam, H. (1975). Is semantics possible? In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds*. Ithaca, NY: Cornell Univ. Press.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton–Century–Crofts.
- Rehder, B., & Hastie, R. (1997). The roles of causes and effects in categorization. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates (pp. 650–655). NJ: Erlbaum.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In A. O. Stella Vosniadou (Ed.), *Similarity and analogical reasoning* (pp. 21–59). New York: Cambridge Univ. Press.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.
- Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573–605.
- Ross, B. H. (1997). The use of categories affects classification. *Journal of Memory and Language*, **37**, 240–267.
- Schwartz, S. P. (1979). Natural kind terms. *Cognition*, **7**, 301–315.
- Sloman, S., & Ahn, W. (1999). Feature centrality: Naming versus imaging. *Memory & Cognition*, **27**, 526–537.
- Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, **22**, 189–228.
- Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, **81**, 214–241.
- Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, **38**, 179–211.
- Spalding, T. L., & Ross, B. H. (1994). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 1251–1263.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, **84**(4), 327–352.
- Tversky, A., & Kahneman, D. (1982). Causal schemas in judgments under uncertainty. In D. Kahneman & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 117–128). Cambridge, NY: Cambridge Univ. Press.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293–315.
- Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

(Accepted June 12, 2000)