Journal of Experimental Psychology:
Learning, Memory, and Cognition
2010, Vol. 36, No. 5, 1171–1206

0278-7393/10/$12.00 DOI: 10.1037/a0019765

Causal Status and Coherence in Causal-Based Categorization

Bob Rehder
New York University

ShinWoo Kim
Kwangwoon University

Research has documented two effects of interfeature causal knowledge on classification. A *causal status effect* occurs when features that are causes are more important to category membership than their effects. A *coherence effect* occurs when combinations of features that are consistent with causal laws provide additional evidence of category membership. In this study, we found that stronger causal relations led to a weaker causal status effect and a stronger coherence effect (Experiment 1), that weaker alternative causes led to stronger causal status and coherence effects (Experiment 2), and that "essentialized" categories led to a stronger causal status effect (Experiment 3), albeit only for probabilistic causal links (Experiment 4). In addition, the causal status effect was mediated by features' subjective *category validity*, the probability they occur in category members. These findings were consistent with a generative model of categorization but inconsistent with an alternative model.

Keywords: categorization, causal knowledge, conceptual representation, causal model theory, generative model

Supplemental materials: http://dx.doi.org/10.1037/a0019765.supp

An important goal in the psychology of concepts is to identify how both empirical information—information that people directly observe—and theoretical beliefs involving explanatory and causal knowledge contribute to how we represent and use categories. Although early research into concepts focused on the effect of empirical information, subsequent research has shown that theoretical beliefs influence many types of category-based judgments, including learning, induction, and categorization (see Murphy, 2002, for a review). In this article, we are concerned with how one type of theoretical knowledge, namely, causal knowledge that relates features of categories, affects one type of category-based judgment, classification.

The presence of causal relationships between category features is pervasive. For example, we know that having claws enables tigers to catch prey, having gills enables fish to breathe, and having a fan causes an automobile's engine to remain cool. Not surprisingly then, numerous studies have investigated how this knowledge affects classification. Some studies have tested naturally occurring categories and the real-world causal knowledge that people already possess about those categories (e.g., Ahn, 1998; Kim & Ahn, 2002; Sloman, Love, & Ahn, 1998). However, to conduct precise tests of alternative models, investigators have turned to artificial categories that are subject to greater experimental control. In these studies, participants are instructed on new types of objects and their features and then are taught causal relations among those features. To assess the effect of those causal relations, participants are then asked to judge the category membership of items displaying various combinations of features. Note that although the categories are artificial, they are intended to be plausible, that is, to consist of features and interfeature causal relations that could conceivably exist (see Rehder, 2010, for a review).

Researchers have tested a number of different causal network topologies, and one network that has proven to be of particular theoretical importance is the three-feature chain (see Figure 1) in which one category feature X causes another feature Y, which in turn causes feature Z. In the first section below, we begin by describing two important empirical results that have been found with chain networks, namely, the *causal status effect* and the *coherence effect*. In the second section, we present two models that have been proposed to account for the effect of causal knowledge on classification and derive their predictions for causal chains. We then present several experiments that test the predictions of the two models.

Two Empirical Effects

The Causal Status Effect

The causal status effect is the phenomenon in which, all else being equal, features that appear earlier in a category's causal network (and thus are "more causal") carry greater weight in categorization decisions. For example, in Figure 1, X is the most causal feature, Z is the least causal, and Y is intermediate. As a consequence, X should be weighed more heavily than Y, which should be weighed more heavily than Z. Of course, the presence of a causal status effect does not imply that features' causal status is the sole factor determining their categorization weight. It is well known that features' weights are also influenced by their perceptual salience (e.g., Lamberts, 1995, 1998) and their *cue validity* (the extent to which they are diagnostic of that category versus

Bob Rehder, Department of Psychology, New York University; ShinWoo Kim, Department of Industrial Psychology, Kwangwoon University.

We thank Woo-kyoung Ahn and Gregory L. Murphy for comments that led to numerous improvements in this article.

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

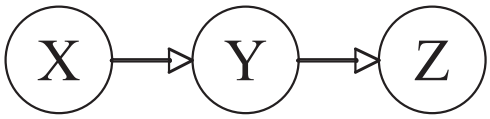

Figure 1. Three-element causal chain.

another; Rosch & Mervis, 1975). However, the claim is that when these factors are controlled, causal status dominates.

Figure 2 presents the results of two empirical studies that provide at least partial support for the causal status effect. In Ahn, Kim, Lassaline, and Dennis (2000, Experiment 1), participants were instructed on artificial categories, such as roobans, that were described as possessing three features: eats fruit (X), has sticky feet (Y), and builds nests on trees (Z). In addition, participants were told about a causal relationship between X and Y ("Eating fruit tends to cause roobans to have sticky feet because sugar in fruits is secreted through pores under their feet.") and Y and Z ("Sticky feet tends to allow roobans to build nests on trees because they can climb up the trees easily with sticky feet."). Participants were then presented with items missing exactly one feature and were asked to rate on a 0–100 scale how likely that item was a category member. For example, a missing-X item had sticky feet (Y) and builds nests on trees (Z) but eats worms instead of fruit (not X). The ratings of missing-X, missing-Y, and missing-Z items are shown in Figure 2A. In fact, an exemplar missing X was rated lower than one missing Y, which in turn was lower than one missing Z, suggesting that X is more important than Y, which is more important than Z. Similar results were found in Sloman et al. (1998, Study 3).

A second study that provides partial support for the causal status effect was conducted by Rehder and Kim (2006). Once again participants were instructed on artificial categories. For example, participants were told about Myastars (a type of star) that had five features, three of which were related in a causal chain. An example of three Myastar features and their causal relations is presented in Table 1. After being instructed on Myastars, participants were shown a series of stars and were asked to rate the category membership of each. To assess the importance of features, Rehder and Kim performed regression analyses on those ratings. The left panel of Figure 2B presents the feature regression weights from that study (averaged over

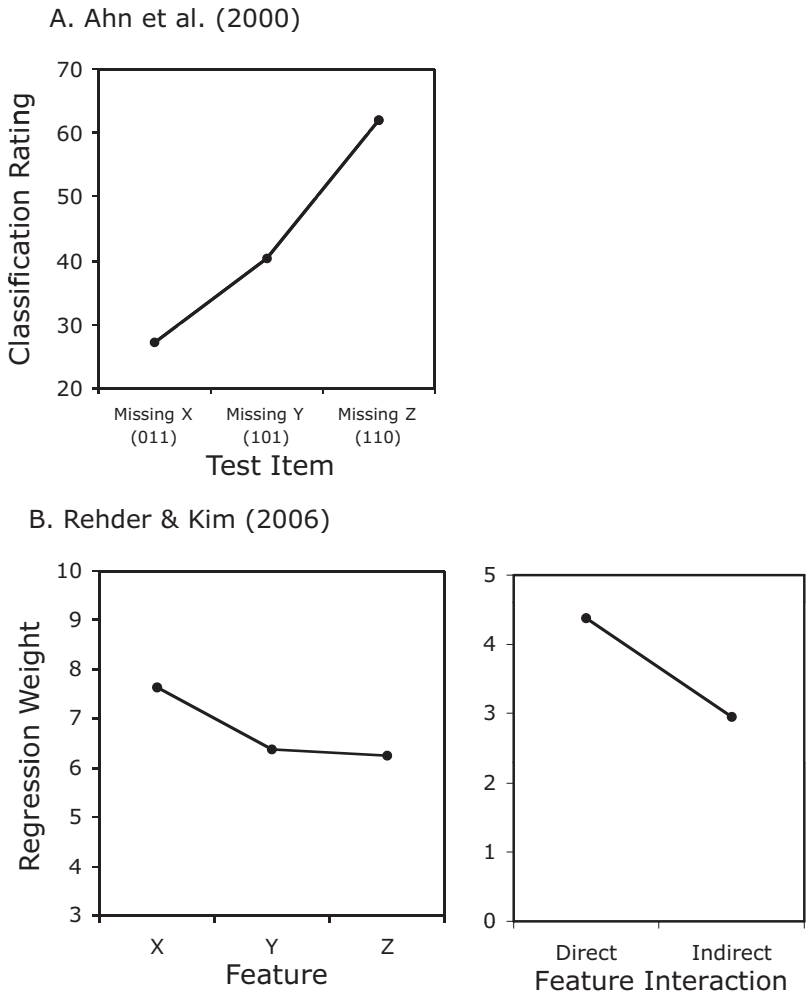

Figure 2. Previous results from (A) Ahn et al. (2000, Experiment 1) and (B) Rehder and Kim (2006).

Table 1
Features and Causal Relationships for Myastars, an Artificial Category

Feature	Description
X	Very hot
Y	High density
Z	Large number of planets
Causal relationship	**Description**
X → Y	A hot temperature causes the star to have high density. At unusually high temperatures heavy elements (such as uranium and plutonium) become ionized (lose their electrons), and the resulting free electrons and nuclei can be packed together more tightly.
Y → Z	High density causes the star to have a large number of planets. Helium, which cannot be compressed into a small area, is spun off the star, and serves as the raw material for many planets.

the chain conditions from three experiments). Note that whereas in Figure 2A lower scores mean more important features, in Figure 2B more important features are indicated by higher regression weights. In fact, the figure indicates that feature X was weighed more heavily than feature Y. Unlike Ahn et al.'s (2000) study, however, the weights of features Y and Z did not differ. This partial causal status effect has also been found for four-feature causal chains (a larger weight on the chain's first feature and smaller but equal weights on the remaining ones; Rehder, 2003b, Experiment 1).

These results notwithstanding, several questions about the causal status effect remain. Although the two studies just reviewed provide support for a causal status effect, they exhibit both quantitative and qualitative differences. First, the size of the effect differed dramatically across the two studies. Whereas the difference between the missing-X and missing-Z items was 35 points on a 100-point scale in Ahn et al. (2000) and was highly significant after testing 15 participants, that difference was only 4 points in Rehder and Kim (2006) (and reached significance at the .01 level only by pooling 192 participants from three experiments). The second question of course concerns why a full causal status effect obtained in some studies (e.g., Ahn et al., 2000; Sloman et al., 1998) but not others (e.g., Rehder, 2003b; Rehder & Kim, 2006). Conceivably, these differences could be due to how the causal links were described. For example, whereas Ahn et al. described the causal relationships in probabilistic terms by use of the phrase "tends to" (e.g., "Sticky feet *tends to* allow roobans to build nests on trees"), Rehder and Kim omitted any information about the strength of the causal link, allowing participants to interpret the links as deterministic (the cause always produces the effect; see Table 1). It may be that a causal status effect only obtains with probabilistic links. Experiments that follow assess this and a number of other hypotheses regarding the conditions that promote a causal status effect.

The Coherence Effect

One important difference between the studies just reviewed concerns how feature weights were assessed. Whereas Ahn et al. (2000) asked participants to rate the category membership of only three test exemplars (missing-X, missing-Y, and missing-Z), Rehder and Kim (2006) asked them to rate all possible exemplars that could be formed on the category's binary dimensions. The latter method allows for an assessment of not only feature weights but also feature *interactions*, that is, how important certain *combinations* of features are for forming good category members. It does so by introducing interaction terms in the regression equation that code, for example, whether features X and Y are both present or both absent versus one present and the other absent. The regression weight on the interaction term that is then derived represents the importance to participants' categorization rating of dimensions X and Y having the same value (present or absent) or not.

In fact, research has established that participants are quite sensitive to whether potential category members exhibit not just the right features considered individually but also the right combination of features, a phenomenon known as the coherence effect. For example, the interaction weights from Rehder and Kim's (2006) chain conditions are presented in the right-hand panel of Figure 2B. In the figure, the three two-way interaction weights for a three-feature chain (XY, YZ, and XZ) have been grouped into two types, namely *direct* interactions between features that are directly causally related (XY and YZ) and the *indirect* interaction between the features that are indirectly related (XZ). As the figure shows, both types of interaction weights were greater than zero. The direct interaction weight of about 4 indicates, for example, that categorization ratings were 4 points higher (all else being equal) when a test item possessed either both X and Y or neither one, and 4 points lower when it possessed one but not the other. Apparently, participants were sensitive to the interfeature correlations one would expect in light of the causal relations, so that an item was more coherent and thus a better category member if it maintained the expected correlations between X and Y, and Y and Z, and was incoherent and, thus, a worse category member if it broke those correlations. The lower weight on the indirect interaction term as compared with the direct terms reflects the fact that X and Z should also be correlated, albeit not as strongly, because of the indirect relation between them (more about this below). This sensitivity to expected correlations induced by causal knowledge has been found in numerous studies (Marsh & Ahn, 2006; Rehder, 2003b), including those testing other causal network topologies (Rehder, 2003a; Rehder & Hastie, 2001; Rehder & Kim, 2006). Coherence effects in which specific configurations of features are treated as a sign of especially good (or bad) category members also occur for interfeature relations that are not explicitly causal (Murphy & Wisniewski, 1989; Rehder & Ross, 2001; Wisniewski, 1995).

Two Theoretical Models

We now present two computational models that have been offered as accounts of the effects of causal knowledge on categorization. Both models specify a rule that assigns to an object a measure of its membership in a category on the basis of that category's network of interfeature causal relations. Neither model denies the existence of other effects on classification, such as the presence of contrast categories, the salience of particular features, or empirical information

about how features are distributed amongst category members that people observe firsthand. Rather, the claim is that causal relations will have the predicted effects when these factors are controlled.

The Dependency Model

One model that has been offered as an account of the causal status effect is Sloman et al.'s (1998) *dependency model*. According to the dependency model, features are more important to category membership (i.e., are more *conceptually central*) to the extent they have more *dependents*, that is, features that depend on them (directly or indirectly). A causal relation is an example of a dependency relation in which the effect depends on its cause. According to the dependency model, feature i's weight or centrality, c_i, can be computed from the iterative equation,

$$c_{i,t+1} = \Sigma d_{ij} c_{j,t}, \quad (1)$$

where $c_{i,t}$ is i's weight at iteration t, and d_{ij} is the strength of the causal link between i and its dependent j.

Figure 3A presents the dependency model's feature weights for a three feature causal chain as a function of link strength d. The figure shows that the dependency model indeed predicts a causal status effect for most values of d. For example, when c_Z is initialized to 1 and each causal link has a strength of 2, after two iterations the weights for X, Y, and Z are 4, 2, and 1, respectively. Stated qualitatively, the dependency model predicts a causal status effect because X, Y, and Z vary in the number of dependents they have: X has two (Y and Z), Y has one (Z), and Z has none.

Although the dependency model thus provides one explanation of the causal status effect, it has not generally fared well in other experimental tests. For example, Rehder and Kim (2006) systematically varied features' number of dependents and found no evidence that they increase in importance as their number of dependents increase. Moreover, the dependency model fails to predict the second major empirical phenomenon described above, coherence effects. One purpose of the present article is to test another prediction of the dependency model, namely, that the size of the causal status effect should increase monotonically with causal strength (see Figure 3A). For example, although after two iterations feature weights are 4, 2, and 1 when $d_{XY} = d_{YZ} = 2$ (yielding a difference of 3 between the weights of X and Z), they are 9, 3, and 1 when the ds $= 3$ (a difference of 8). Our experiments manipulate the strength of the causal links to test whether the causal status effect increases as causal strength increases.

The Generative Model

A second model that has been offered as an account of causal knowledge and classification is the *generative model* (Rehder, 2003a, 2003b; Rehder & Kim, 2006). Building on *causal-model theory* (Sloman & Lagnado, 2005; Waldmann & Holyoak, 1992),

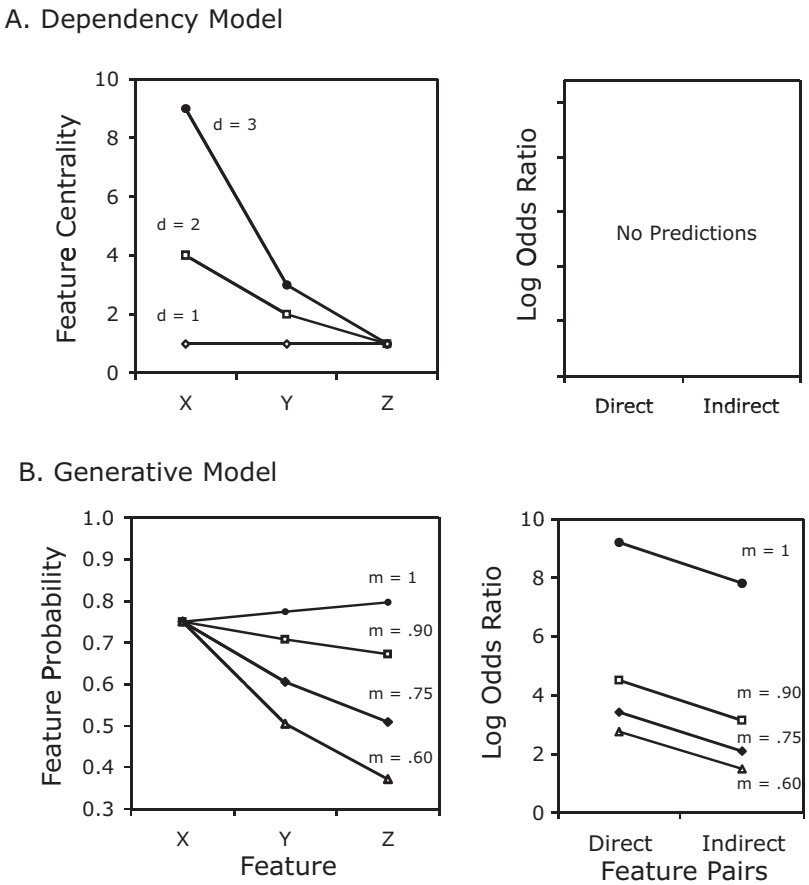

Figure 3. Predictions from two theoretical models: (A) the dependency model and (B) the generative model.

the generative model assumes that interfeature causal relations are represented as probabilistic causal mechanisms and that classifiers consider whether an object is likely to have been produced or generated by those causal mechanisms. Objects that are likely to have been generated by a category's causal model are considered to be good category members, and those unlikely to be generated are poor category members.

Quantitative predictions for the generative model can be derived assuming a particular representation of causal relations first introduced by Cheng (1997) and later applied to a variety of category-based tasks (Rehder, 2003a, 2003b; Rehder, 2009; Rehder & Burnett, 2005; Rehder & Hastie, 2001; Rehder & Kim, 2006, 2009a). For example, for the causal model in Figure 1, assume that the causal mechanism relating feature j and its parent i operates (i.e., produces j) with probability m_{ij} when i is present and that any other potential background causes of j collectively operate with probability b_j. Given other reasonable assumptions (e.g., that the causal mechanisms operate independently; see Cheng & Novick, 2005), then i and the background causes form a "fuzzy-or" network that together produce j in members of category k with probability,

$$p_k(j \mid i) = 1 - (1 - b_j)(1 - m_{ij}) = m_{ij} + b_j - m_{ij}b_j. \quad (2)$$

When i is absent, the causal mechanism has no effect on j, in which case the probability of j is simply

$$p_k(j \mid \bar{i}) = b_j. \quad (3)$$

The probability of the root cause, $p_k(\mathrm{X})$, is a free parameter c_X.

By applying Equations 2 and 3 iteratively, one can derive the equations representing the likelihood of any possible combination of the presence or absence of three features X, Y, and Z as a function of the c, m, and b parameters. These likelihood equations are presented in Table 2. The table also presents the probability of each item for different parameter values. The strengths of the causal links between X and Y (m_{XY}) and between Y and Z (m_{YZ}) vary across .60, .75, .90, and 1.0 while b_Y and b_Z are held fixed at .10. Parameter c_X is fixed at .75, reflecting the assumption that X is a typical feature of category k. The claim of course is that an item's probability determines its degree of category membership. That is, according to the generative model, the effects of causal relations on classification are mediated by the statistical distribution of features that those relations are expected to produce.

Although the generative model's predictions in Table 2 are for whole items in which the state of all features is known, one can derive statistics corresponding to the two sorts of empirical effects on classification we have described, namely, feature weights and feature interactions. First, according to the generative model, the evidence that an individual feature provides for membership in a category is determined by its probability of appearing in category members (i.e., its *category validity*). Whereas research has traditionally assumed that (a classifier's belief about) a feature's category validity is derived from first-hand observation of category members (e.g., Rosch & Mervis, 1975), the generative model assumes that it can also be derived from (the classifier's belief about) the causal relations in which the feature participates. Specifically, for the chain network of Figure 1, the probability of feature j appearing in a member of category k can be derived from Equations 2 and 3,

$$\begin{aligned} p_k(j) &= p_k(j \mid i)p_k(i) + p_k(j \mid \bar{i})p_k(\bar{i}), \\ p_k(j) &= (m_{ij} + b_j - m_{ij}b_j)p_k(i) + b_j p_k(\bar{i}), \\ p_k(j) &= m_{ij}p_k(i) + b_j - m_{ij}b_j p_k(i). \end{aligned} \quad (4)$$

Table 2
Predictions of the Generative Model for Experiment 1 for Different Values of m_{XY} and m_{YZ} Holding c_X, b_Y, and b_Z Constant

Model parameters					
c_X		.750	.750	.750	.750
m_{XY}, m_{YZ}		.600	.750	.900	1.0
b_Y, b_Z		.100	.100	.100	.100
Exemplar likelihoods					
$p_k(111)$	$c_\mathrm{X}(m_{\mathrm{XY}} + b_\mathrm{Y} - m_{\mathrm{XY}}b_\mathrm{Y})(m_{\mathrm{YZ}} + b_\mathrm{Z} - m_{\mathrm{YZ}}b_\mathrm{Z})$	.307	.450	.621	.750
$p_k(011)$	$(1 - c_\mathrm{X})b_\mathrm{Y}(m_{\mathrm{YZ}} + b_\mathrm{Z} - m_{\mathrm{YZ}}b_\mathrm{Z})$	.016	.019	.023	.025
$p_k(101)$	$c_\mathrm{X}[1 - (m_{\mathrm{XY}} + b_\mathrm{Y} - m_{\mathrm{XY}}b_\mathrm{Y})]b_\mathrm{Z}$	.027	.017	.007	0
$p_k(110)$	$c_\mathrm{X}(m_{\mathrm{XY}} + b_\mathrm{Y} - m_{\mathrm{XY}}b_\mathrm{Y})[1 - (m_{\mathrm{YZ}} + b_\mathrm{Z}m_{\mathrm{YZ}}b_\mathrm{Z})]$	.173	.131	.061	0
$p_k(100)$	$c_\mathrm{X}[1 - (m_{\mathrm{XY}} + b_\mathrm{Y} - m_{\mathrm{XY}}b_\mathrm{Y})](1 - b_\mathrm{Z})$	.243	.152	.061	0
$p_k(010)$	$(1 - c_\mathrm{X})b_\mathrm{Y}[1 - (m_{\mathrm{YZ}} + b_\mathrm{Z} - m_{\mathrm{YZ}}b_\mathrm{Z})]$	.009	.006	.002	0
$p_k(001)$	$(1 - c_\mathrm{X})(1 - b_\mathrm{Y})b_\mathrm{Z}$	.023	.023	.023	.023
$p_k(000)$	$(1 - c_\mathrm{X})(1 - b_\mathrm{Y})(1 - b_\mathrm{Z})$	.203	.203	.203	.203
Feature probabilities					
$p_k(\mathrm{X})$		.750	.750	.750	.750
$p_k(\mathrm{Y})$		.505	.606	.708	.775
$p_k(\mathrm{Z})$		.373	.509	.673	.798
Interfeature associations[a]					
log θ_k(X, Y)		2.77	3.43	4.51	9.21
log θ_k(Y, Z)		2.77	3.43	4.51	9.21
log θ_k(X, Z)		1.49	2.11	3.15	7.82

[a] The odds ratio is not defined when there are zero cells, a case that arises in Table 2 when $m = 1.0$. Thus, for this case, the log odds ratio presented in Table 2 are computed assuming that $m = .999$.

where i is the parent of j. For instance, when $c_X = .75$, $m_{XY} = m_{YZ} = .90$, and $b_Y = b_Z = .10$, then $p_k(X) = .750$, $p_k(Y) = .708$, and $p_k(Z) = .673$. For these causal model parameters, the generative model thus predicts that feature X will have a larger categorization weight than Y, which in turn will have a greater weight than Z, that is, a causal status effect should obtain.

The predictions of the generative model regarding how the size of the causal status effect should vary as a function of the m parameters are presented in the left panel of Figure 3B. The figure shows that those probabilities decrease from X to Y to Z for many values of the ms, and the magnitude of that decrease becomes greater as the ms decrease. That is, the generative model predicts that the size of causal status effect should decrease as the strength of the causal links increases; indeed, the causal status effect can even reverse at high levels of causal strength. These predictions hold so long as the bs < 1 (and that the bs stay fixed as the ms vary, an assumption discussed further below). These predictions distinguish the generative model from the dependency model, which predicts that the causal status effect should increase with causal strength (see Figure 3A).

Second, according to the generative model, the importance of feature combinations to classification stem from the fact that causally related features should be correlated. Whereas previous research has assumed that a classifier's belief about feature configurations is derived from first-hand observations (e.g., Hayes-Roth & Hayes-Roth, 1977; Medin & Schaffer, 1978; Reitman & Bower, 1973), the generative model assumes that objects will be considered good category members to the extent they maintain correlations that are expected on the basis of causal relations. A measure of association between two variables i and j is the *log odds ratio*, defined as the natural log of the ratio of the odds that j occurs in the presence of i and the odds that it occurs in the absence of i (Agresti, 2002).

$$\log\ \theta_k(i,j) = \log[odds(j \mid i)/odds(j \mid \bar{i})],$$

$$\log\ \theta_k(i,j) = \log[[p_k(j \mid i)/p_k(\bar{j} \mid i)]/[p_k(j \mid \bar{i})/p_k(\bar{j} \mid \bar{i})]],$$

$$\log\ \theta_k(i,j) = \log[p_k(j \mid i)p_k(\bar{j} \mid \bar{i})/p_k(j \mid \bar{i})p_k(\bar{j} \mid i)]. \quad (5)$$

For instance, for the chain network in Figure 1, Table 2 shows that when $c_X = .75$, $m_{XY} = m_{YZ} = .90$, and $b_Y = b_Z = .10$, then $\log\ \theta_k(X, Y) = \log\ \theta_k(Y, Z) = 4.51$, and $\log\ \theta_k(X, Z) = 3.15$. These values reflect the expected pattern of correlations between features in a causal chain: stronger correlations between the directly related features than between the indirectly related ones. The generative model thus explains the coherence effect in Rehder and Kim (2006), in which a large interaction weight on directly related features and a smaller weight on the indirectly related features (see Figure 2B) were found.

The right panel of Figure 3B presents how the feature correlations predicted by the generative model for a chain network vary as a function of the m parameters.[1] First, it predicts that both direct (between X and Y, and Y and Z) and indirect (X and Z) correlations should be positive for all values of the ms > 0 and increase as the ms increases, reflecting the larger correlations expected when features are more strongly causally related. Second, the indirect correlation should be smaller than the direct ones, but this difference should itself decrease as m increases. These predictions distinguish the generative model from the dependency model, which does not predict coherence effects.

In summary, the dependency and generative model make very different predictions regarding how causal-based classification should be affected by causal link strength and, as we show below, other model parameters. Of course, neither model assumes that classifiers explicitly carry out the computations specified in Equations 1–5. Rather, those computations are intended to approximate the mental processes through which causal knowledge affects intuitive judgments of category membership.

Overview of Experiments

To provide new tests of the models just described, in the present article we go beyond previous studies (e.g., Ahn et al., 2000; Rehder, 2003b; Rehder & Kim, 2006) by manipulating the parameters of categories' causal models. In Experiment 1, we begin by testing the contrasting predictions of the dependency model and the generative model regarding how the causal status and coherence effects vary with causal link strength (see Figure 3). In Experiment 2, we then test how that effect varies as a function of alternative causes of category features (i.e., the generative model's b parameters). In Experiment 3, we then assess how the causal status effect changes when observed features are causally related to unobserved ones. Finally, in Experiment 4, we test how this effect itself interacts with causal link strength. In addition, in all four experiments, we test the generative model's prediction that the causal status effect is mediated by the perceived probability with which features appear in category members (i.e., their subjective category validity) by asking participants to make explicit feature likelihood judgments. In all four experiments, we test the generative model's predictions regarding the coherence effect.

Experiment 1

To test the effect of causal link strength on the causal status and coherence effects, we assigned participants in Experiment 1 to one of the three conditions shown in Figure 4. Whereas all participants were taught three category features (such as the three Myastar features shown in Table 1), in the Chain-100 and Chain-75 conditions, participants were also taught two causal relationships linking X, Y, and Z into a causal chain. Participants were also given explicit information about the strengths of those causal links. For example, participants in the Chain-100 condition who learned about Myastars were told that that each causal link had a strength of 100%: "Whenever a Myastar has a very hot temperature, it will cause that star to have high density with probability 100%." and "Whenever a Myastar has high density, it will cause that star to have a large number of planets with probability 100%." Participants in the Chain-75 condition were told that the causal links operated with probability 75% instead of 100%. Participants were instructed on no causal relationships in the Control condition.

[1] The odds ratio is not defined when there are zero cells, a case that arises in Table 2 when $m = 1.0$. Thus, the log odds ratios presented in Table 2 for this case are computed assuming that $m = .999$. In the subsequent Tables 4 and 6, zero cells also arise when $b = 0$, in which case the presented log odds ratios are computed assuming $b = 0.001$.

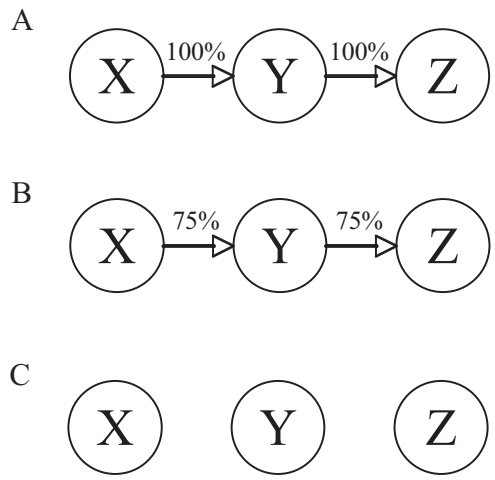

Figure 4. Causal structures tested in Experiment 1: (A) Chain-100 condition, (B) Chain-75 condition, and (C) Control condition.

Recall that the dependency model and the generative model make distinct predictions for this experiment. Because the dependency model predicts that the size of the causal status effect is a monotonic function of causal strength (see Figure 3A), it predicts a stronger causal status effect in the Chain-100 condition versus the Chain-75 condition. In contrast, because the generative model predicts that a causal status effect should be strong at intermediate values of causal strength and zero or negative when those strengths are 100% (see Figure 3B), it predicts a stronger causal status effect in the Chain-75 condition versus the Chain-100 condition. Both models predict no causal status effect in the Control condition.

Experiment 1 augments the classification ratings with an additional dependent variable by asking participants to provide an explicit likelihood rating for each feature. If features indeed exhibit a stronger causal status effect in the Chain-75 condition, the generative model predicts that this occurs because X is viewed as more common among category members than Y, which in turn is more common than Z. Thus, differences in feature importance to classification ratings should be reflected in the feature likelihood ratings. The dependency model makes no predictions regarding feature likelihood ratings.

The generative model also predicts how the coherence effect should be affected by the causal strength manipulation. First, because it predicts that the size of that effect should increase with strength (see Figure 3B), the effect should be larger in the Chain-100 condition versus the Chain-75 condition (and should be zero in the Control condition). Second, because it predicts a larger difference between direct and indirect correlations for probabilistic causal links, the direct and indirect interaction weights should exhibit a larger difference in the Chain-75 condition than the Chain-100 condition. The dependency model does not predict coherence effects.

Recall that an important assumption regarding the generative model's predictions concerns the possibility of alternative causes of features Y and Z. The predictions in Figure 3 assume that alternative causes are held constant over different levels of causal strength, but Experiment 1's Chain-75 participants might partly compensate for their weaker causal links by assuming stronger alternative causes. Quantitative fits of the generative model to the results of Experiment 1 will assess this possibility.

Method

Materials. Six novel categories were tested: two nonliving natural kinds (Myastars, Meteoric Sodium Carbonate), two biological kinds (Kehoe Ants, Lake Victoria Shrimp), and two artifacts (Romanian Rogos, Neptune Personal Computers). Each category had three binary feature dimensions. In previous studies demonstrating coherence effects in causal-based categorization, one value on each dimension was described as characteristic or typical of the category, whereas the other, atypical value was often described as "normal." For example, in Rehder and Kim (2006), participants were told that "Most Myastars have high temperature, whereas some have a normal temperature," "Most Myastars have high density, whereas some have a normal density," and so on. Although the intent was to define Myastars with respect to the superordinate category (all stars), Marsh and Ahn (2006) argued that this use of "normal" might have inflated coherence effects because participants might expect all the normal dimension values to appear together. To address this concern, in the experiments that follow, we tested categories with binary dimensions with opposing values. For example, rather than Myastars having either a hot temperature or a normal temperature, they were described as having a hot temperature or a *low* temperature.

Each causal relationship was described as one characteristic feature causing another accompanied with one or two sentences describing the mechanism responsible for the causal relationship (see Table 1 for examples). In addition, a sentence describing the strength of the relationship was provided (e.g., "Whenever a Myastar has high density, it will cause that star to have a large number of planets with probability N%" [where N was either 75 or 100]). A complete list of the features and causal relationships for all six experimental categories is available from the authors.

Participants. One hundred and eight New York University undergraduates received course credit for participating in this experiment. They were randomly assigned to the Chain-100, Chain-75, and Control conditions and to one of the six categories, subject to the constraint that an equal number of participants appeared in each cell.

Procedure. Experimental sessions were conducted by computer. Participants first studied several screens of information about the category at their own pace and then performed classification and feature likelihood tests. All participants were presented with three initial screens that presented the category's cover story and which features occurred in "most" versus "some" category members. Chain-100 and Chain-75 participants were presented with two additional screens. The first instructed them on the two causal relationships that formed a chain network, including the causal link itself and the causal mechanism responsible for that link (see Table 1 for examples). The second presented a diagram similar to those in Figures 4A and 4B depicting the structure and strength of the causal links (where "X," "Y," and "Z" were replaced with the actual features; e.g., "high density").

When ready, participants took a multiple-choice test that tested them on the knowledge they had just studied. While taking the test, participants were free to return to the information screens they had studied; however, doing this obligated the participant to retake the

test. The only way to pass the test and to proceed to subsequent phases was to take it all the way through without errors and without returning to the initial information screens for help. In the Control condition, the test consisted of six questions querying which features were typical and which were atypical. In the causal conditions, participants answered an additional 10 questions asking which features were causally related, the direction of the links, and information about the causal mechanism.

During the classification test, participants rated the category membership of all possible eight exemplars that can be formed from three binary features, each presented twice. The order in which a test item's features were listed was randomized across participants (but was consistent from trial to trial for a particular participant). Underneath the test item the question "Is this an X?" appeared, where X was the name of the category. Responses were entered by positioning a slider on a scale in which the left end was labeled "Sure that it isn't." and the right end was labeled "Sure that it is." The slider could be set to 21 distinct positions, which were scaled into the range 0–100. The order of presentation of the 16 test items was randomized for each participant.

During the feature likelihood rating task that followed the classification test, each of the two features on the three binary dimensions was presented alone on the computer screen, and participants rated what proportion of all category members possessed that feature. Each of the six features was presented twice, for a total of 12 likelihood rating trials. The order of these trials was randomized for each participant. Experimental sessions lasted approximately 40 min.

Results

The average category membership ratings given to the eight test items in each condition are presented in Table 3. To determine the effect of causal network on the importance of features and the interactions between features, we analyzed those ratings by performing a multiple regression for each participant. Three predictors (f_X, f_Y, f_Z) were coded as 1 if the characteristic feature on that dimension was present and –1 if the uncharacteristic feature was present. Recall that the regression weight associated with each f_i represents the influence that dimension i had on category membership ratings. A positive weight indicates that the presence of the characteristic feature increased categorization ratings and the uncharacteristic feature decreased it. Three additional predictors were constructed by computing the multiplicative interactions between each possible dimension pair: f_{XY}, f_{XZ}, and f_{YZ}. These variables are coded –1 if one of the characteristic features is present and the other absent, and 1 if both are present or both absent. Recall that for each interaction term, a positive weight indicates that ratings are sensitive to whether the expected correlation between that pair of dimensions is preserved (cause and effect both present or both absent) or broken (one present and the other absent). The effect of causal strength on features and feature interactions are presented separately in the following two sections.

Feature weights. Initial analyses of the feature weights revealed that there was no effect of which category participants learned, and so regression weights averaged over participants for features X, Y, and Z are presented in the left-hand panels of Figure 5. Figure 5A presents the comparison between the Chain-100 and Control conditions, and Figure 5B presents the comparison between the Chain-75 and Control groups. Recall that the dependency model predicts that the magnitude of the causal status effect should increase monotonically with causal strength, whereas the generative model predicts the opposite. In fact, Figure 5B confirms a decrease in feature weights in the Chain-75 condition from 8.6 to 7.3 to 5.1 for features X, Y, and Z, respectively. In contrast, features weights in the Chain-100 condition (6.2, 7.7, and 6.6) did not decrease.

To assess the differences between feature weights statistically, we conducted a 3 × 3 mixed analysis of variance (ANOVA) in which the between-subjects factor was condition (Chain-100, Chain-75, Control), and the within-subject factor was feature (X, Y, Z). The interaction between condition and feature was significant, $F(4, 210) = 3.07$, $MSE = 21.9$, $p < .05$. In addition, the test of the interaction between condition and the linear trend in feature weights (from X to Y to Z, the appropriate test of a causal status effect) confirmed that the slopes differed in the three conditions, $F(2, 105) = 3.79$, $MSE = 25.3$, $p < .05$. In separate analyses, the linear trend in the Chain-75 condition was significantly different from those in the Chain-100 condition, $F(1, 70) = 5.56$, $MSE = 141.0$, $p < .05$, and the Control condition, $F(1, 70) = 5.81$, $MSE = 147.2$, $p < .05$. In contrast, the linear trend in the Chain-100 and the Control conditions did not differ ($F < 1$), consistent with the presence of a causal status effect in the Chain-75 condition but not the Chain-100 condition. (Note that the small quadratic effect in the Chain-100 condition suggested by Figure 5A did not reach significance, $p > .15$.).

The 3 × 3 ANOVA also revealed a main effect of condition, $F(2, 105) = 9.24$, $MSE = 123.5$, $p < .001$, indicating that feature weights were significantly lower in the two chain conditions as

Table 3
Classification Ratings From Experiment 1

Test Item	Chain-100 condition	Chain-75 condition	Control condition
111	94.2 (2.1)	93.8 (1.8)	93.3 (2.2)
011	24.4 (4.6)	40.6 (4.6)	67.4 (2.8)
101	18.7 (3.7)	38.7 (4.2)	67.2 (2.7)
110	22.3 (4.6)	46.7 (4.2)	65.8 (2.3)
100	13.9 (2.4)	32.6 (3.2)	40.2 (3.4)
010	14.1 (2.2)	25.2 (2.8)	39.0 (3.6)
001	13.4 (2.3)	24.5 (2.8)	40.6 (3.1)
000	47.4 (6.3)	52.4 (6.2)	19.7 (4.6)

Note. Standard errors are shown in parentheses.

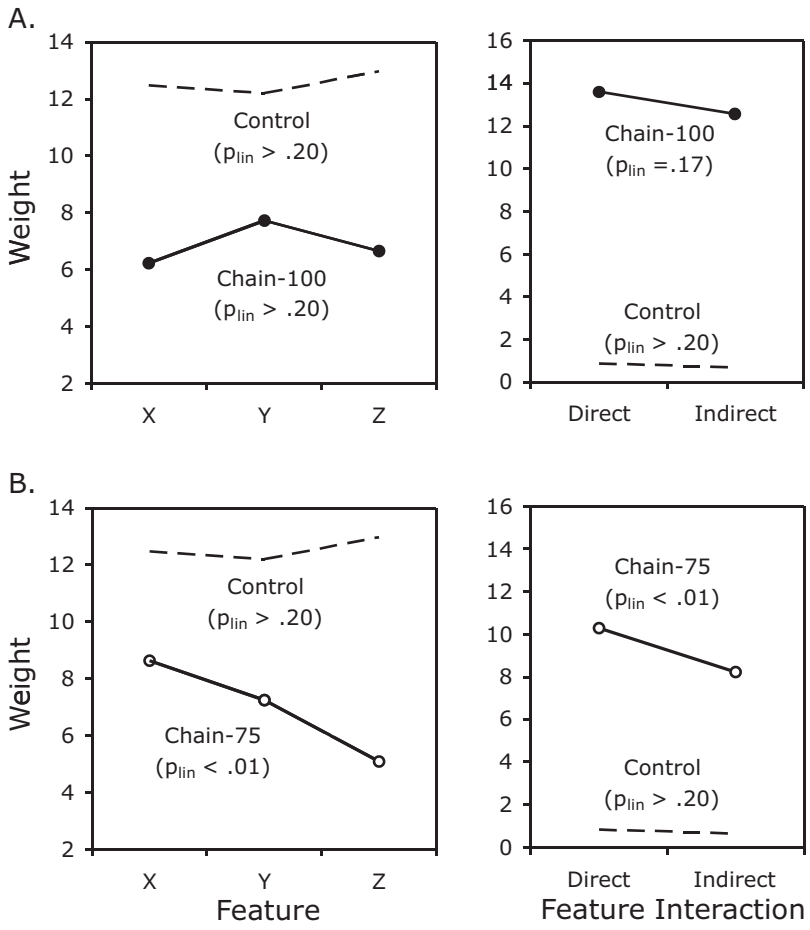

Figure 5. Results of regression analyses of classification ratings from Experiment 1: (A) Chain-100 condition versus Control condition and (B) Chain-75 condition versus Control condition. Left panels: feature weights; right panels: interaction weights; p_{lin} is the significance of the linear trend in each condition.

compared with the Control condition. We defer discussion of this finding until after presenting the remaining results. The main effect of feature did not reach significance ($p > .20$).

Feature interactions. The three two-way interaction terms were aggregated according to whether they were between feature pairs that were linked directly (f_{XY} and f_{YZ}) or indirectly (f_{XZ}). As was the case for feature weights, there was no effect of which category participants learned, and thus the interaction weights averaged over participants are presented in the right-hand panels of Figure 5. Recall that the generative model predicts that the weights on the direct and indirect interaction terms should both increase with causal strength and that the difference between the two should be larger for weaker causal strengths. Consistent with these predictions, the direct and indirect interaction weights were 13.6 and 12.5, respectively, in the Chain-100 condition as compared with 10.3 and 8.2 in the Chain-75 condition. As expected, in the Control condition, the interaction weights were close to 0 (0.8 and 0.7).

A 3 × 2 ANOVA of the interaction weights was conducted with condition (Chain-100, Chain-75, Control) and interaction type (direct vs. indirect) as factors. There was a main effect of condition, $F(2, 105) = 33.98$, $MSE = 84.2$, $p < .0001$, reflecting the larger interaction weights in the two causal conditions. In addition, as predicted by the generative model, the interaction terms were larger in the Chain-100 condition as compared with the Chain-75 condition (13.1 vs. 9.3), $F(1, 105) = 6.24$, $MSE = 84.2$, $p < .05$. There was a main effect of interaction type, $F(1, 105) = 6.79$, $MSE = 9.8$, $p < .05$, indicating that the direct interaction terms were generally greater than the indirect one. Although the interaction between condition and interaction term was not significant ($p > .15$), a separate analysis of the Chain-75 conditions revealed a significant effect of direct versus indirect interaction terms ($p < .01$). In contrast, this difference did not reach significance in the Chain-100 condition ($p = .17$).

Feature likelihood ratings. Feature likelihood ratings for the three characteristic features are presented in Figure 6 and generally exhibit the same pattern as the feature regression weights. In the Chain-75 condition (see Figure 6B), features were rated as less prevalent as one moved down the causal chain, from 76.7% for X, to 73.7% for Y, to 70.2% for Z. In contrast, in the Chain-100 condition (see Figure 6A), likelihood ratings were not significantly different from one another (77.8%, 77.3%, and 77.0%) or from those in the Control condition (77.2%, 74.7%, and 76.7%).

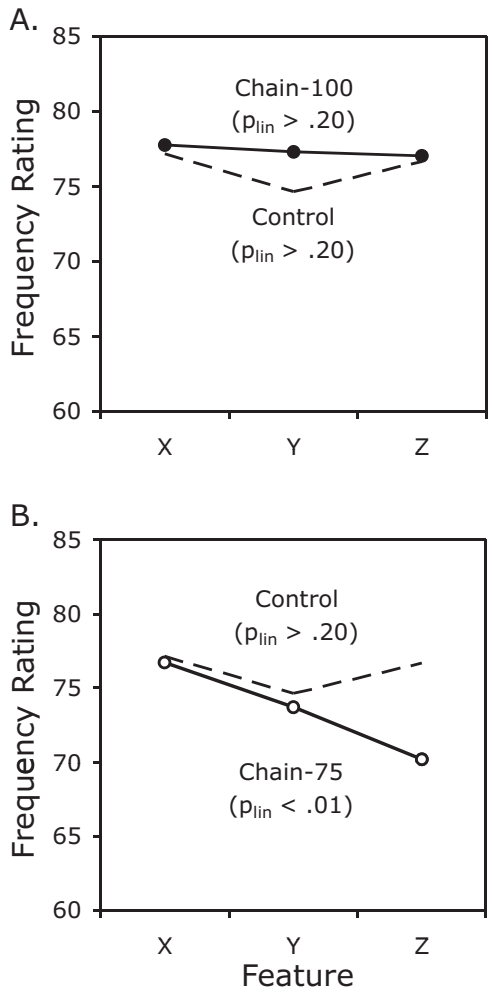

Figure 6. Feature likelihood ratings from Experiment 1: (A) Chain-100 condition versus Control condition and (B) Chain-75 condition versus Control condition; p_{lin} is the significance of the linear trend in each condition.

A 3 × 3 ANOVA of the feature likelihood ratings confirmed an effect of condition on the linear trend, $F(2, 105) = 4.71$, $MSE = 44.4$, $p < .01$, reflecting the different pattern of ratings in the three conditions. The linear effect in the Chain-75 condition was significantly different from both the Chain-100 condition, $F(1, 105) = 6.76$, $MSE = 44.4$, $p < .01$, and the Control condition, $F(1, 105) = 7.36$, $MSE = 44.4$, $p < .01$. The linear effect in the Chain-100 condition did not differ significantly from that in the Control condition ($F < 1$). In the 3 × 3 ANOVA, there was an overall effect of feature, $F(2, 210) = 5.96$, $MSE = 33.2$, $p < .01$, but no main effect of condition, $F(2, 105) = 1.80$, $MSE = 229.1$, $p = .17$.

Selected test items. Both the causal status effect and the coherence effect can be observed directly in the classification ratings of individual test items. First, the causal status effect in the Chain-75 condition is apparent in the ratings for the items missing only feature X and only feature Z presented in Figure 7A: The missing-X item (011) was rated 6.1 points lower than the missing-Z item (110), indicating the relatively greater importance of X. In contrast, 011 was not rated lower than 110 in the Chain-100 condition, reflecting the lack of a causal status effect in that condition.

Second, the strong coherence effect found in both the Chain-100 and Chain-75 conditions is apparent in Figure 7B, which presents the test item classification ratings as a function of their number of characteristic features. In the Control condition, ratings are a simple monotonic function of the number of features. In contrast, in both causal conditions, items with two or one features are rated lower than those with three or zero (i.e., Items 111 and 000). Intuitively, the explanation for these differences is simple. When Control participants are told, for example, that "most" Myastars are very hot, have high density, and have a large number of planets, they expect that most Myastars will have most of those features and that the atypical values exhibited by "some" Myastars (unusually cool temperature, low density, and small number of planets) will be spread randomly among category members. That is, they expect the category to exhibit a normal family resemblance structure in which features are independent (i.e., are uncorrelated within the category). However, when those features are causally related, the Prototype 111 and Item 000 receive the highest ratings. Apparently, rather than expecting a family resemblance structure with uncorrelated features, participants expected the "most" dimension values to cluster together (111) and the "some" values to

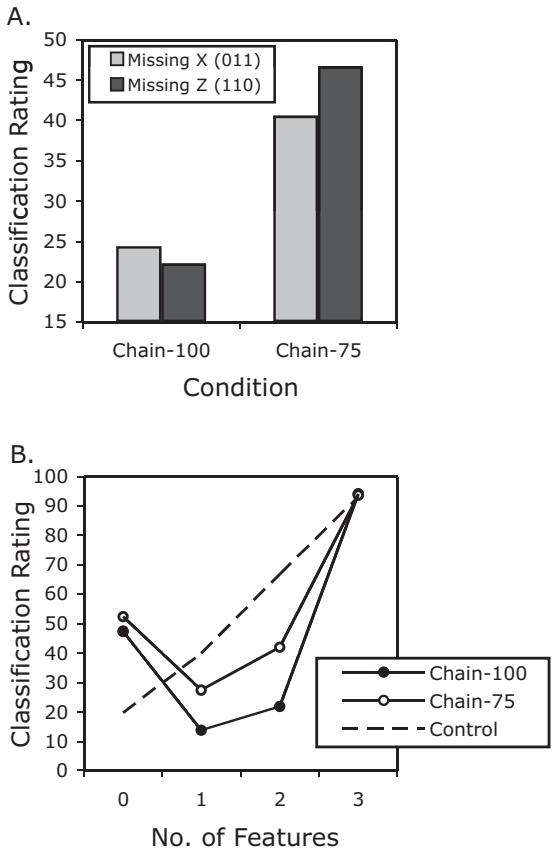

Figure 7. Classification ratings from Experiment 1: (A) for test items missing only feature X and only feature Z and (B) for all test items as a function of their number of characteristic features.

cluster together (000) because that distribution of features is most sensible in light of the causal relations that link them. As a result, the rating of Test Item 000 is an average 30 points higher in the causal conditions than in the Control condition. And, items that are incoherent because they have one or two characteristic features (and thus have a mixture of "most" and "some" values) are rated 29 points lower than in the Control condition.

Individual differences. We also asked whether Experiment 1's classification results are manifested consistently by all participants or whether they arose as a result of averaging over individuals with substantially different response profiles. In fact, cluster analyses revealed subgroups of participants with qualitatively different responses. In both the Chain-100 and Chain-75 conditions, inspection of these subgroups revealed that they varied on the emphasis they placed on coherence and feature typicality information. In both conditions, a minority of participants classified on the basis of coherence alone. These participants produced very low ratings on test items with one or two characteristic features (and that thus violated the causal laws). The ratings on those items with zero or three characteristic features (000 and 111) were not only high but equal, indicating that they ignored feature typicality. In both conditions, another minority of participants appeared to ignore the causal links entirely, generating classification ratings that were indistinguishable from the control group in that they responded on the basis of family resemblance information (i.e., number of typical features) alone. Nevertheless, the modal participants in both conditions performed in a manner consisted with the group averages in Figures 5 and 7, namely, they exhibited sensitivity to both feature typicality information and coherence. In Appendix A, we describe these individual differences in more detail for Experiments 1–4.

Model fitting. To further demonstrate how the generative and dependency models account for the present data, we fit both models quantitatively to the classification ratings. For this purpose, we introduce parameters β_0 and β_1 that map a model's predictions onto the rating scale according to a linear transformation. Thus, for the generative model, the ratings are predicted according to the formula,

$$rating(t_i) = \beta_0 + \beta_1 p_K(t_i;\, c_X,\, m_{XY},\, m_{YZ},\, b_Y,\, b_Z),$$

where t_i is a test item, and expressions for p_K are given in Table 2 for each t_i. Parameters m_{XY} and m_{YZ} were set to 1.0 in the Chain-100 condition and to .75 in the Chain-75 condition, respectively. Parameters b_Y and b_Z were assumed to be equal and so collapsed into a single free b parameter. Because the base rates of features within the category were not given exact numerical values (participants were only told that they appeared in "most" category members), c_X was also free. Parameters c_X and b were constrained to the range [0, 1].

For the dependency model, the ratings are predicted according to the formula,

$$rating(t_i) = \beta_0 + \beta_1 \sum_i c_i f_i,$$

where c_i is the centrality of feature i, and f_i codes the presence or absence of feature i in t_i (1 or 0). Because a three-element causal chain converges in two iterations, analytical expressions for the weight of each feature as a function of the causal strength parameters d_{ij} can be derived: $c_X = d_{XY}d_{YZ}c_{Z,0}$, $c_Y = d_{YZ}c_{Z,0}$, and $c_Z = c_{Z,0}$. Because causal link strengths were equal within each condition, parameters d_{XY} and d_{YZ} were collapsed into a single d parameter; d was a free parameter because it is unclear a priori which of its values correspond to strengths of 100% and 75%. We constrained d to the range [1, 4].[2] Parameter $c_{Z,0}$ (Z's initial centrality parameter) is absorbed by β_1 and so is dropped from the model. For both models, parameter β_0 was constrained to be ≥ 0, and β_1 was constrained to be >0.

The models were fit to each participant's classification ratings by identifying parameters that minimized squared error. For the generative model, the best fitting parameters averaged over participants were $c_X = .67$, $b = 0.23$, $\beta_0 = 15$, and $\beta_1 = 130$ in the Chain-100 condition, and $c_X = .70$, $b = 0.35$, $\beta_0 = 23$, and $\beta_1 = 168$ in the Chain-75 condition. For the dependency model, the averaged parameter values were $d = 1.22$, $\beta_0 = 13.8$, and $\beta_1 = 10.3$ in the Chain-100 condition, and $d = 1.48$, $\beta_0 = 23$, and $\beta_1 = 12.2$ in the Chain-75 condition. The predicted and observed ratings for each test item averaged over participants are presented in Figure 8 for each condition. As is apparent from the figures, the generative model provides an overall good qualitative account of the data from this experiment. Most prominently, it reproduces the large coherence effects by predicting the highest ratings for Test Items 111 and 000 and lower ratings for those test items with a combination of present and absent features. As a consequence, the generative model accounts for 98% and 89% of the variance in the observed ratings in Figure 8 in the Chain-100 and Chain-75 conditions, respectively (R^2s averaged over participants were .87 and .74). One notable result is that the average b parameter was larger in the Chain-75 condition (0.35) than in the Chain-100 condition (0.23), suggesting that the Chain-75 group partly compensated for the weaker causal relations by assuming stronger alternative causes. We return to this issue in the Discussion section.

The dependency model, in contrast, is unable to reproduce the qualitative response pattern in either condition of this experiment. For example, because it is unable to take into account feature interactions, it predicts that Test Item 000 will receive the lowest ratings in both conditions. As a consequence, the dependency model explains only 20% and 35% of the variance in the Chain-100 and Chain-75 conditions, respectively (R^2s averaged over participants were .29 and .41). The better fit of the generative model also holds according to a measure, root mean squared error (RMSE), that corrects for the different number of parameters in the two models.[3] *RMSE* averaged over participants was 11.4 and 19.0 in the Chain-100 and Chain-75 conditions for the generative model as compared with 33.4 and 25.5 for the dependency model.

Figure 8 also reveals some systematic differences between the generative model's predictions and the observed ratings. Specifically, it tends to overpredict test items with zero or one typical feature (e.g., 000, 100, 010, and 001) and to underpredict items with two typical features (011, 101, and 110). For example, al-

[2] Parameter d was constrained to be ≥ 1 because when $d < 1$, effect features can become more central than causes, contradicting the model's main assumption. It was constrained to be <4 because as d becomes large, c_X becomes $>> c_Y$ and c_Z, in which case it becomes redundant with β_1. These constraints affect the average parameter values we report. However, they have virtually no impact on the overall quality of the dependency model's fits, which improved by less than 6% when d was unconstrained.

[3] $RMSE = \text{SQRT}\,[SSE/(n - p)]$, where SSE = sum of squared error for a participant, n = number of data points fit (8), and p = a model's number of parameters (in Experiment 1, $p = 4$ for the generative model, and $p = 3$ for the dependency model).

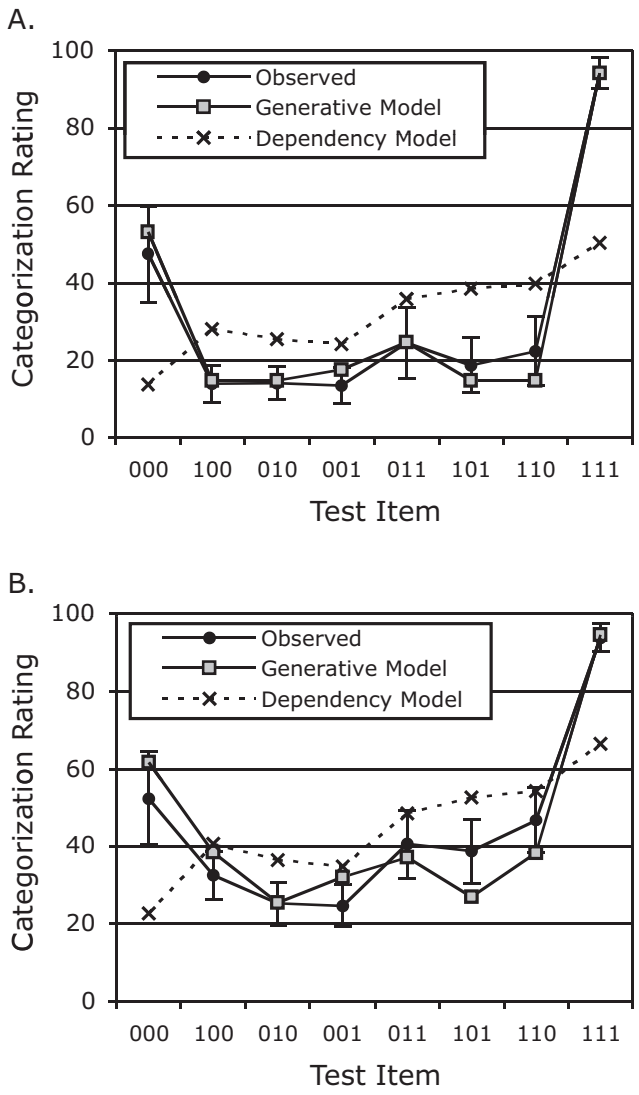

Figure 8. Fits of the generative and dependency models to the classification ratings of Experiment 1: (A) Chain-100 condition and (B) Chain-75 condition. Error bars are 95% confidence intervals.

though for the Chain-100 condition the generative model predicts that Items 100, 010, 101, and 110 should all be equally unacceptable category members (because they include cases in which a cause feature is present but its effect is absent, situations that should never arise if the causal links are deterministic), the items with two features (101 and 110) were rated significantly higher than those with one (100 and 010), 20.5 versus 14.0. A post hoc explanation for these results involves the individual differences presented above. As mentioned, a minority of participants ignored the causal relations and responded like the control participants by basing their category membership ratings solely on a test item's number of features (e.g., these participants assigned an average rating of 63 to the two-feature items). We analyze the effect of these participants on the generative model's fits in more detail in Appendix A.

Finally, two points regarding how parameters β_0 and β_1 map the generative model's predicted probabilities onto the response scale are worthy of note. First, the parameters produced by the generative model fits implied substantially different probabilities for the prototype item (111) in the Chain-100 and Chain-75 conditions (average of .67 and .49, respectively). Nevertheless, Figure 8 shows that the model is able to account for the fact that participants rated those items about equally in the two conditions (94.2 and 93.8), with a larger value of β_1 in the Chain-75 condition (159) than in the Chain-100 condition (130). Furthermore, even though the generative model predicts a probability of zero for Items 110, 101, 100, and 010 in the Chain-100 condition (see Table 2), Figure 8 shows that the generative model accounts for these items' nonzero rating with a value of 15 for parameter β_0. We believe that parameters β_0 and β_1 are capturing known facts about how people use response scales (Poulton, 1989). First, extreme items tend to anchor the ends of a scale (thus, Item 111 receives a very high rating in the Chain-75 condition because it is the most probable item). Furthermore, people tend not to use a scale's most extreme values because they do not like to express the certainty that those extreme values convey (thus, Items 110, 101, 100, and 010 do not receive ratings of exactly 0 in the Chain-100 condition).

Discussion

The generative model predicts that the magnitude of the causal status effect should vary nonmonotonically with causal strength, that is, it should be maximal at intermediate causal strengths and decrease as causal strength increases. Consistent with this prediction, a significant linear effect of feature weights obtained in the Chain-75 condition but not the Chain-100 or Control conditions. This result was corroborated by feature likelihood ratings that exhibited the same pattern. In contrast, these results are inconsistent with the dependency model that predicts that greater causal strengths should lead to larger causal status effect.

One unexpected result from Experiment 1 is that feature weights in both causal conditions were lower than those in the Control condition. However, we interpret this result as stemming from how participants used the response scale. In the Control condition, the only sources of evidence relevant to category membership were the features considered individually, and thus the entire range of the 0–100 scale could be allocated solely as a function of the number of characteristic features possessed by a test item (0, 1, 2, or 3). In contrast, in the Chain-100 and Chain-75 conditions, ratings were affected by additional sources of evidence, namely the two-way interactions between features. Because these participants allocated some of the 0–100 response scale to the interactions, they could allocate less to the features individually. Thus, we suggest that the lower feature regression weights in the causal conditions obtained not because the features were thought by the participants to be less diagnostic but rather because the fixed-size response scale also needed to be used to express the effect of coherence between features.

The large coherence effect found in Experiment 1 was another prediction of the generative model. Despite Marsh and Ahn's (2006) concern that the coherence effect may be primarily due to the wording of feature values used in previous research (in which the uncharacteristic feature value was described as "normal"), the large and positive weights on both the direct and indirect interaction terms in both causal conditions indicated a large coherence effect. This effect was manifested in the Chain-100 and Chain-75

conditions in the lower category membership ratings of the incoherent test items with one or two characteristic features and the higher ratings of the coherent item with zero characteristic features. The generative model also successfully predicted many of the more subtle aspects of the interaction terms. It correctly predicted that the magnitude of the interaction terms should be larger in the Chain-100 condition than in the Chain-75 condition. Furthermore, it correctly predicted that the direct interaction terms should be larger than the indirect term in the Chain-75 condition but not the Chain-100 condition.

To bolster the conclusions from Experiment 1, we conducted a number of follow-up studies. First, one factor that was not controlled in Experiment 1 was the possibility of alternative causes of features Y and Z (corresponding to the generative model parameters b_Y and b_Z). The generative model fits suggested that Chain-75 participants compensated for the weaker causal relations by assuming stronger alternative causes as compared with the Chain-100 condition. Thus, to establish that the effects reported in Experiment 1 hold even when alternative causes are held constant, we conducted a version of Experiment 1 in which participants were told that features Y and Z had no alternative causes within the category (i.e., that $b_Y = b_Z = 0$). Second, Experiment 1 compared deterministic versus probabilistic causal links, but Figure 3 shows that the generative model also predicts a weaker causal status effect for strong probabilistic links as compared with weak ones. Accordingly, we also compared conditions in which the strengths of the causal links were either 90% or 60%. Third, to demonstrate that our results generalize beyond our specific stimuli and experimental procedure, we performed two replications of Ahn et al.'s (2000) Experiment 1 (that found a large causal status effect; Figure 2A). Recall that the causal relations used in that study implied probabilistic relations (e.g., "Sticky feet *tends to* allow roobans to build nests on trees."). On the basis of the results from Experiment 1, we hypothesized that this effect would disappear if deterministic causal links were tested instead. Accordingly, we replicated Ahn et al.'s experiment by comparing a condition using the original "tends to" wording with one that used "always" instead (e.g., "Sticky feet *always* allows roobans to build nests on trees."). Finally, the fits of the generative model implied that the rating scale was being used differently in the Chain-100 and Chain-75 conditions. To address any concern that the differences in the causal status and coherence effects across conditions were an artifact of scale usage, we tested a variant of Experiment 1 in which the causal strength manipulation was within-subjects.

The results of these experiments, reported in Appendices B and C, replicate the key results from Experiment 1. Specifically, in every experiment, the causal status effect was weaker in the condition with stronger causal links; indeed, it was absent entirely when causal links were deterministic. In addition, stronger causal links also always resulted in a stronger coherence effect. These results are consistent with the generative model and inconsistent with the dependency model.

Experiment 2

Experiment 2 conducts another test of the generative and dependency models by manipulating the strength of alternative causes of the category features, that is, parameter *b*. Participants were assigned to one of the two conditions shown in Figure 9. In both conditions, the causal relationships between features X, Y, and Z were described as having a strength of 75%. However, participants were given different information about the possibility of alternative causes of Y and Z. Those in the Background-0 condition were told that Y and Z had no other causes. For example, participants who learned about Myastars learned not only that hot temperature causes high density with probability 75% but also that "There are no other features of Myastars that cause high density. Because of this, when its known cause (very hot temperature) is absent, high density occurs in 0% of all Myastars." In contrast, participants in the Background-50 condition were told that "There are also one or more other features of Myastars that cause high density. Because of this, even when its known cause (very hot temperature) is absent, high density occurs in 50% of all Myastars." Because Experiment 1 has already confirmed that features are weighed equally when not causally related (and because our main interest lies in comparing the Background-0 and Background-50 conditions), this Control condition was omitted in Experiment 2.

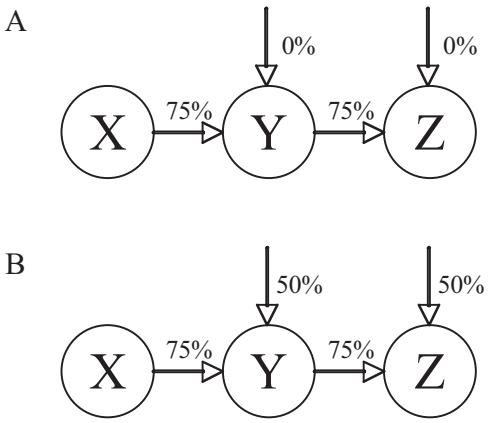

Figure 9. Causal structures tested in Experiment 2: (A) Background-0 condition and (B) Background-50 condition.

To demonstrate how the generative model's predictions vary with *b*, in Table 4 we present the exemplar probability distributions computed using the likelihood equations in Table 2 for causal strengths varying b_Y and b_Z over the values 0, 0.25, 0.50, and 0.75 while holding m_{XY} and m_{YZ} fixed at 0.75. The resulting feature probabilities indicate that the causal status effect should become weaker as potential background causes get stronger; indeed, it should reverse as *b* grows much larger than 0.50. Intuitively, this occurs because as the *b*s approach 1, features Y and Z will be present in all category members. As a consequence, for Experiment 2, the generative model predicts a causal status effect in the Background-0 condition and the absence of one in the Background-50 condition.

In addition, Table 4 shows that the magnitude of the log odd ratios reflecting the strength of association between features should decrease monotonically with the strength of background causes. For example, whereas the log odds ratios between the directly and indirectly related features are 8.01 and 6.60 when the *b* parameters are 0, respectively, they are 1.95 and 0.78 when the *b*s = 0.50. This prediction reflects the intuition that the correlation between a cause and effect will weaken to the extent

that the effect has additional causes. That is, the generative model predicts stronger coherence effects in the Background-0 condition as compared with the Background-50 condition.

The dependency model makes a different prediction for this experiment. Because it specifies that a feature's centrality is a sole function of the strength of the links with its dependents, supplying a feature with additional background causes should have no effect on its centrality. Thus, because centrality should be unaffected by the background cause manipulation, the dependency model predicts an identical causal status effect in the Background-0 and Background-50 conditions.[4] Of course, the dependency model also differs from the generative model in not predicting (a) an effect on feature likelihood judgments or (b) coherence effects.

Method

The materials and procedure were identical to those in Experiment 1, except for the information about background causes. Seventy-two New York University undergraduates received course credit for participating in this experiment. They were randomly assigned to the Background-0 and Background-50 conditions and to one of the six categories, subject to the constraint that an equal number of participants appeared in each cell.

Results

The average category membership ratings given to each of the 8 test items in each condition are presented in Table 5. As in Experiment 1, there were no effects of which category participants learned on any dependent variable, and thus the results are collapsed over this factor. We first present the regression analyses and then the feature likelihood ratings.

Table 5
Classification Ratings From Experiment 2

Test Item	Background-0 condition	Background-50 condition
111	89.4 (3.0)	95.1 (1.2)
011	25.1 (4.4)	52.8 (3.7)
101	25.8 (4.5)	42.1 (4.5)
110	42.0 (4.5)	51.9 (3.4)
100	27.5 (3.8)	36.2 (3.2)
010	19.2 (2.9)	29.4 (3.3)
001	23.2 (4.0)	36.0 (3.3)
000	48.8 (6.0)	44.0 (5.9)

Note. Standard errors are shown in parentheses.

Feature weights. The regression weights averaged over participants for features X, Y, and Z are presented in the left-hand panel of Figure 10. As predicted by the generative model, a larger causal status effect obtained in the Background-0 condition (feature weights of 8.5, 6.3, and 3.2) as compared with the Background-50 condition (7.9, 8.9, and 8.1).

A 2 × 3 ANOVA of the feature weights was conducted with condition (Background-0 and Background-50) and feature (X, Y, Z) as factors. There was a marginal effect of condition, $F(1, 70) = 2.16$, $MSE = 125.1$, $p = .14$; a marginal effect of feature, $F(2, 140) = 3.01$, $MSE = 42.5$, $p = .05$; and a significant Condition × Feature interaction, $F(2, 140) = 3.22$, $MSE = 42.5$, $p < .05$. In addition, there was an overall effect of condition on the linear trend, $F(1, 70) = 4.49$, $MSE = 120.8$, $p < .05$, confirming that the slopes differed in the two conditions. In separate analyses, the linear effect was significantly different than zero in the Background-0 condition ($p < .01$) but not the Background-50 condition ($p > .20$).

Table 4
Predictions of the Generative Model for Experiment 2 for Different Values of b_Y and b_Z Holding c_X, m_{XY}, and m_{YZ} Constant

Model parameters				
c_X	.750	.750	.750	.750
m_{XY}, m_{YZ}	.750	.750	.750	.750
b_Y, b_Z	0	.250	.500	.750
Exemplar likelihoods				
$p_k(111)$	.422	.495	.574	.659
$p_k(011)$	0	.051	.109	.176
$p_k(101)$	0	.035	.047	.035
$p_k(110)$	.141	.114	.082	.044
$p_k(100)$	.188	.105	.047	.012
$p_k(010)$	0	.012	.016	.012
$p_k(001)$	0	.047	.063	.047
$p_k(000)$	.250	.141	.063	.016
Feature probabilities				
$p_k(X)$	.750	.750	.750	.750
$p_k(Y)$	.563	.672	.781	.891
$p_k(Z)$	.422	.628	.793	.917
Interfeature associations[a]				
log $\theta_k(X, Y)$	8.01	2.57	1.95	1.61
log $\theta_k(Y, Z)$	8.01	2.57	1.95	1.61
log $\theta_k(X, Z)$	6.60	1.33	0.78	0.43

[a] The odds ratio is not defined when there are zero cells, a case that arises in Table 4 when $b = 0$. For this case, the log odds ratio presented in Table 4 are computed assuming that $b = 0.001$.

[4] There may be some uncertainty regarding the dependency model's predictions for this experiment that stems from ambiguity regarding how its construct of "dependency strength" should be interpreted. We interpret it as reflecting the propensity of the cause to produce its effect, that is, as a causal power (corresponding to the generative model's m parameter). However, an alternative interpretation is that it corresponds to the well known ΔP rule of causal induction. On one hand, we interpret the work of Cheng and colleagues as showing that when you ask people for the strength of a causal link (a type of dependency relation), they generally respond with an estimate of causal power rather than ΔP (Buehner, Cheng, & Clifford, 2003; Cheng, 1997). However, what people induce in a causal learning task is controversial (e.g., Lober & Shanks, 2000), and even Buehner et al. (2003) found that substantial minority of participants responded according to ΔP. The distinction between causal power and ΔP is important because whereas the Background-0 and Background-50 conditions have the same causal power, ΔP is larger in the former condition (.750) than in the latter (.375; see Table 4). However, whereas the ΔP version of the dependency model thus also predicts a weaker causal status effect in the Background-50 condition, it still differs from the generative model, which predicts that the causal status effect in that condition should be absent entirely—or be slightly negative, $p_k(X) - p_k(Z) = -.043$ (see Table 4). Of course, the models also differ in that only the generative model predicts that the manipulation of background causes will also affect the coherence effect and feature likelihood ratings. We thank an anonymous reviewer for raising this issue.

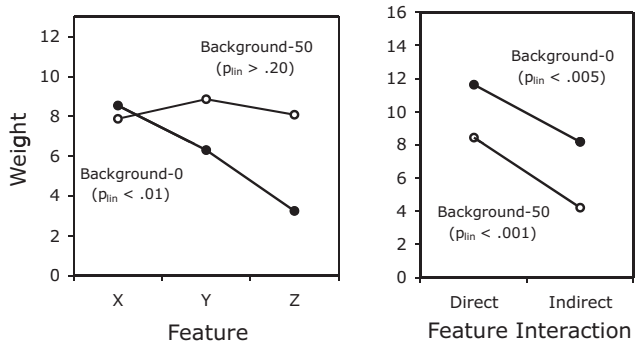

Figure 10. Results of regression analyses of classification ratings from Experiment 2. Left panel: feature weights; right panel: interaction weights; p_{lin} is the significance of the linear trend in each condition.

Feature interactions. The interaction weights are presented in the right-hand panel of Figure 10. The generative model predicts that interaction weights should be larger in the Background-0 than in the Background-50 condition and that in both conditions the direct interaction weights should be larger than the indirect one. Both of these predictions were confirmed. The direct and indirect interaction weights were 11.7 and 8.2, respectively, in the Background-0 condition, and 8.4 and 4.2 in the Background-50 condition.

A 2 × 2 ANOVA of the interaction weights was conducted with condition and interaction type as factors. There was a main effect of condition, $F(1, 70) = 4.19$, $MSE = 110.6$, $p < .05$, reflecting the larger interaction weights in the Background-0 condition, and a main effect of interaction type, $F(1, 70) = 27.42$, $MSE = 95.5$, $p < .0001$, reflecting that the direct interactions were larger than the indirect one. There was no interaction.

Feature likelihood ratings. Feature likelihood ratings for the three characteristic features are presented in Figure 11 and follow the same pattern as the feature regression weights. In the Background-0 condition each subsequent feature was rated as less prevalent, from 78.2% for X, to 73.0% for Y, to 67.0% for Z. Feature likelihood ratings in the Background-50 condition, in contrast, did not differ from one another (76.8%, 75.4%, and 76.0%).

A 2 × 3 ANOVA of the feature likelihood ratings revealed a marginal effect of condition, $F(1, 70) = 3.36$, $MSE = 175.6$, $p = .07$, and a main effect of feature, $F(2, 140) = 10.54$, $MSE = 60.4$, $p < .0001$. Importantly, there was an effect of condition on the linear trend, $F(1, 70) = 15.47$, $MSE = 82.1$, $p < .001$, confirming that the slopes differed in the two conditions. In separate analyses, the linear effect was significantly different than zero in the Background-0 condition ($p < .0001$) but not the Background-50 condition ($p > .20$).

Selected test items. Following Experiment 1, we show how the causal status effect and the coherence effect can be observed directly in the classification ratings of individual test items. First, Figure 12A shows that in the Background-0 condition, the missing-X item was rated 14.5 points lower than the missing-Z item, reflecting a causal status effect. In contrast, the nearly equal ratings received by those items in the Background-50 condition reflect the lack of a causal status effect.

Second, the strong effect of coherence in both causal conditions is apparent in the pattern of ratings shown in Figure 12B. (Figure 12B includes ratings from Experiment 1's Control condition for compar-

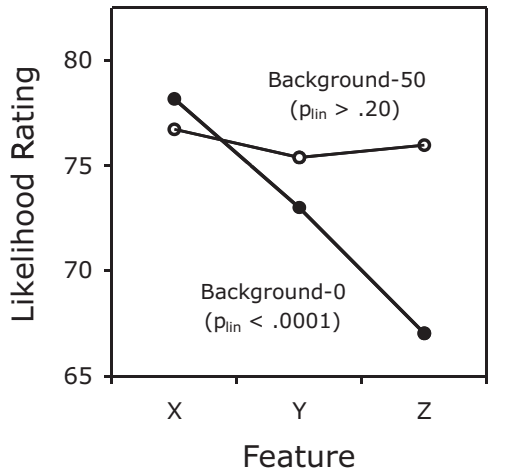

Figure 11. Feature likelihood ratings from Experiment 2; p_{lin} is the significance of the linear trend in each condition.

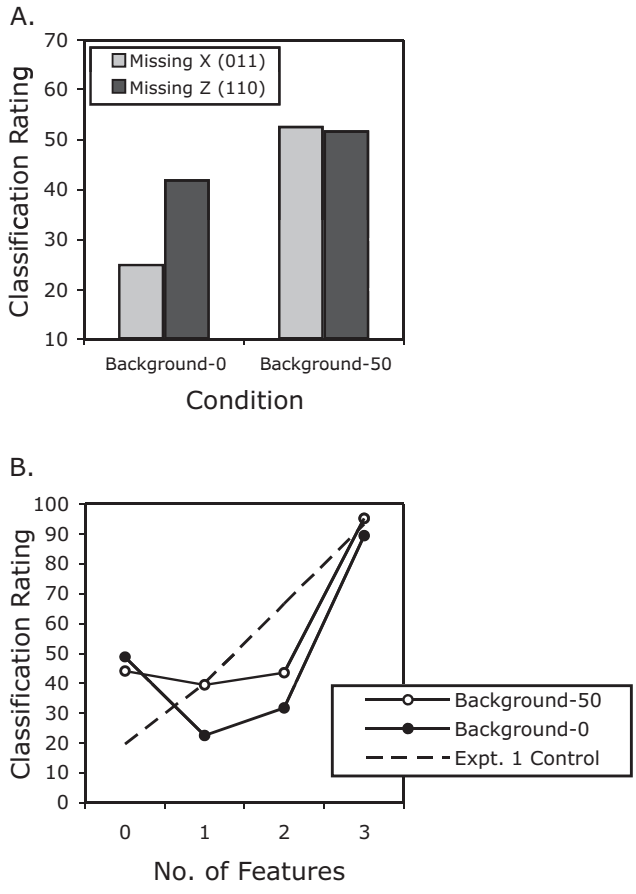

Figure 12. Classification ratings from Experiment 2: (A) for test items missing only feature X and only feature Z and (B) for all test items as a function of their number of characteristic features.

ison.) Whereas in the Control condition ratings are a monotonic function of the number of characteristic features, in the causal conditions, incoherent items with one or two features are rated lower (an average of 19.1 points) relative to the Control condition, and the Coherent Item 000 is rated higher (26.7 points). Apparently, when causal relations link category features, participants no longer expect a family resemblance structure with uncorrelated features. Instead, they expect category members to reflect the correlations that the causal relations generate: The causally linked characteristic features should be more likely to appear together in one category member, and atypical features should be more likely to appear together in another.

Model fitting. Following Experiment 1, we fit the generative and dependency models to the results of Experiment 2. For the generative model, parameters m_{XY} and m_{YZ} were assumed to be 0.75, b_Y and b_Z were set to 0 in the Background-0 condition and 0.50 in the Background-50 condition, and parameter c_X was free. For the dependency model, parameter d was free. All free parameters were constrained as in Experiment 1. For the generative model, the best fitting parameters averaged over participants were $c_X = .79$, $\beta_0 = 19$, and $\beta_1 = 150$ in the Background-0 condition, and $c_X = .76$, $\beta_0 = 33$, and $\beta_1 = 121$ in the Background-50 condition. For the dependency model, the averaged parameter values were $d = 1.78$, $\beta_0 = 18.1$, and $\beta_1 = 9.4$ in the Background-0 condition, and $d = 1.52$, $\beta_0 = 24.7$, and $\beta_1 = 13.5$ in the Background-50 condition.

The predicted and observed ratings for each test item averaged over participants are presented in Figure 13 for each condition. The generative model provided a generally good qualitative account of these data and a fair quantitative account, accounting for 84% and 93% of the variance in the observed ratings in Figure 13 in the Background-0 and Background-50 conditions, respectively (R^2s averaged over participants were .67 and .59; *RMSE*s were 21.0 and 20.8). In contrast, the dependency model only accounted for 25% and 54% of the variance (averaged R^2s were .34 and .48; average *RMSE*s were 30.2 and 22.1). As was the case in Experiment 1, the generative model tends to overpredict test items with few typical features (especially Item 100 in the Background-0 condition) and to underpredict items with two typical features. Furthermore, items for which the generative model predicts a probability of zero (i.e., 011, 101, 010, and 001; see Table 4) in the Background-0 condition did not receive zero ratings, reflecting that people tend to not use the extreme ends of scales.

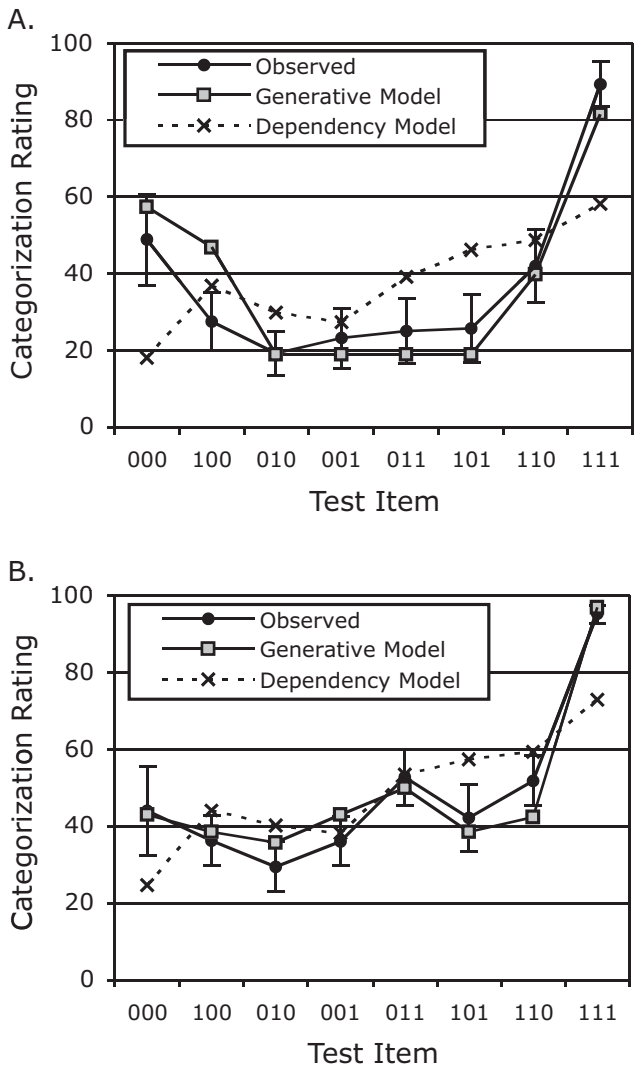

Figure 13. Fits of the generative and dependency models to the classification ratings of Experiment 2: (A) Background-0 condition and (B) Background-50 condition. Error bars are 95% confidence intervals.

Discussion

As predicted by the generative model, the causal status effect was larger in the Background-0 condition as compared with the Background-50 condition, a result that was seen in both classification regression weights and feature likelihood ratings. The dependency model, in contrast, predicted no effect of manipulating the features' background causes.

Experiment 2 once again yielded strong coherence effects. Moreover, the generative model correctly predicted the larger interaction terms in the Background-0 condition as compared with the Background-50 condition, consistent with the idea that interfeature correlations will be stronger when there are no other causes of an effect. Also as predicted, the direct interaction terms were larger than the indirect one in both the Background-0 and the Background-50 condition. Overall, the feature and interaction weights in Experiment 2 were consistent with the predictions of the generative model and inconsistent with those of the dependency model.

To bolster the conclusions from Experiment 2, we conducted two follow-up studies. First, as in Experiment 1, to address any concern that the differences between the Background-0 and Background-50 conditions were an artifact of differences in response scale usage, we tested a variant of Experiment 2 in which the background strength manipulation was within-subjects. Second, Experiment 2 compared background cause strengths of 0% and 50%, but Table 4 shows that the generative model predicts a stronger causal status whenever background causes are weaker, not just absent. Accordingly, we also compared two conditions in which the strengths of the background causes were either 25% or 75%. The results of these experiments, reported in Appendix C,

replicate the key results from Experiment 2, namely, larger causal status effects and coherence effects in those conditions with weaker background causes. The results of Experiment 2 and these follow-on experiments are consistent with the predictions of the generative model and are inconsistent with those of the dependency model.

Experiment 3

In Experiments 1 and 2, we tested the effect of varying the *m* and *b* parameters on the causal status and coherence effects. However, there are good reasons to expect that categorizers often reason with a causal model that is more elaborate than one that includes only observable features. For example, numerous researchers have suggested that people view many kinds as being defined by underlying properties or characteristics (an *essence*) that is shared by all category members and by members of no other categories (Gelman, 2003; Keil, 1995; Medin & Ortony, 1989; Rehder & Kim, 2009a; Rips, 2001) and that are presumed to generate, or cause, perceptual features. Although many artifacts do not appear to have internal causal mechanisms (e.g., pencils and wastebaskets), it has been suggested that the essential properties of artifacts may include the intentions of their designers (Bloom, 1998; Chaigneau, Barsalou, & Sloman, 2004; Keil, 1995; Matan & Carey, 2001; Rips, 2001; cf. Malt, 1994; Malt & Johnson, 1992). Thus, the causal model that people reason with during categorization may include the underlying causes that categorizers presume bring rise to a category's observable features.

In Experiment 3, we test the importance of the category being essentialized by comparing the causal structures shown in Figure 14. As in Experiments 1 and 2, each category consisted of three observable features. However, the categories were now essentialized by endowing them with an additional feature that exhibits two important characteristics of an essence, namely, it appears in all members of the category and in members of no other category. For example, for Myastars, the essential property was "ionized helium," and participants were told that all Myastars possess ionized helium and that no other kind of star does.[5] Although participants were informed of the essential feature during initial category learning, this feature was never observed during the subsequent classification test, that is, test items were presented with values on dimensions X, Y, and Z only. Thus, this test was identical to the one presented in Experiments 1 and 2.

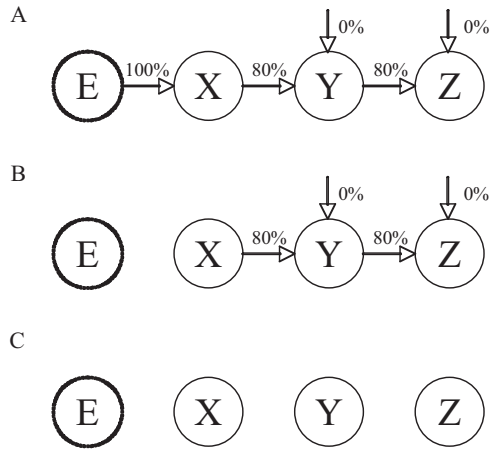

Figure 14. Causal structures tested in Experiment 3: (A) Essentialized-Chain-80 condition, (B) Unconnected-Chain-80 condition, and (C) Control condition.

The three conditions in Figure 14 differ according to the causal knowledge participants learned. In both the Essentialized-Chain-80 (see Figure 14A) and Unconnected-Chain-80 conditions (see Figure 14B), participants were instructed on two causal links relating features X, Y, and Z in a causal chain. Participants were told that these links had a strength of 80% and that there were no other causes of Y and Z. In the Essentialized-Chain-80 condition, they were also instructed on a third causal relationship of strength 100% linking the underlying essential feature E to feature X. To equate the two conditions, we provided no information about alternative causes of X. The Control condition served as a comparison group in which no causal links are provided.

According to the generative model, there are two bases for expecting a larger causal status effect in the Essentialized-Chain-80 condition as compared with the Unconnected-Chain-80 condition. First, the likelihood equations for the essentialized causal model in Figure 14A derived from Equations 2 and 3 are presented in Table 6, assuming that E is always present (that is, $c_E = 1.0$). It also presents the probability of each exemplar when $m_{EX} = 1.0$, $m_{XY} = m_{YZ} = 0.80$, and b_Y and $b_Z = 0$. (These probabilities hold for any value of b_X. Table 6 also shows probabilities when $m_{XY} = m_{YZ} = 1.0$; predictions we make use of in Experiment 4.) For comparison, Table 6 also shows the exemplar probabilities for the unconnected causal model in Figure 14B computed from the likelihood equations in Table 2 for the same parameter values used for the essentialized model (and $c_X = .75$). Finally, Table 6 shows for each condition the probability of each feature appearing in category members, computed from Equation 4.

The larger causal status effect in the Essentialized-Chain-80 condition is indicated by a larger difference between $p_k(X)$ and $p_k(Z)$ than in the Unconnected-Chain-80 condition (0.36 vs. 0.27). This difference arises because whereas $p_k(X)$ is 1.0 in the Essentialized-Chain-80 condition (because E always produces X), it is 0.75 in the Unconnected-Chain-80 condition, and this results in a larger drop in feature probabilities along the causal chain in the former condition. This prediction holds for any value of $c_X < 1$.

[5] Although explicitly defining essential features in this manner controls the knowledge brought to bear during classification, note that these experimentally defined materials may differ in various ways from (people's beliefs about) some real category essences. Although adults' beliefs about essences are sometimes concrete (e.g., DNA in the case of biological kinds for adults), preschool children's knowledge about animals' essential properties is less specific, involving only a commitment to biological mechanisms that operate on their "insides" (Gelman, 2003; Gelman & Wellman, 1991; Johnson & Solomon, 1997). Furthermore, an essential property, not just one that just happens to be present in all category members (and absent in all nonmembers), is one that is present in all category members that *could* exist. However, although the concreteness and noncontingency of people's essentialist beliefs is undoubtedly important under some circumstances, we suggest that a feature that is present in all category members is sufficient to induce a larger causal status effect.

Table 6
Predictions of the Generative Model for the Essentialized Causal Model in Figure 14A for Different Values of m_{XY} and m_{YZ} Holding c_X, b_Y, and b_Z Constant

		Essentialized model		Unconnected model	
Model parameters					
c_E		1.0	1.0		
m_{EX}		1.0	1.0		
c_X				.750	.750
m_{XY}, m_{YZ}		.800	1.0	.800	1.0
b_Y, b_Z		0	0	0	0
Exemplar likelihoods					
$p_k(111)$	$(m_{EX} + b_X - m_{EX}b_X)(m_{XY} + b_Y - m_{XY}b_Y)(m_{YZ} + b_Z - m_{YZ}b_Z)$	.640	1.0	.480	.750
$p_k(011)$	$[1 - (m_{EX} + b_X - m_{EX}b_X)]b_Y(m_{YZ} + b_Z - m_{YZ}b_Z)$	0	0	0	0
$p_k(101)$	$(m_{EX} + b_X - m_{EX}b_X)[1 - (m_{XY} + b_Y - m_{XY}b_Y)]b_Z$	0	0	0	0
$p_k(110)$	$(m_{EX} + b_X - m_{EX}b_X)(m_{XY} + b_Y - m_{XY}b_Y)[1 - (m_{YZ} + b_Z - m_{YZ}b_Z)]$	.160	0	.120	0
$p_k(100)$	$(m_{EX} + b_X - m_{EX}b_X)[1 - (m_{XY} + b_Y - m_{XY}b_Y)](1 - b_Z)$	.200	0	.150	0
$p_k(010)$	$[1 - (m_{EX} + b_X - m_{EX}b_X)]b_Y[1 - (m_{YZ} + b_Z - m_{YZ}b_Z)]$	0	0	0	0
$p_k(001)$	$[1 - (m_{EX} + b_X - m_{EX}b_X)](1 - b_Y)b_Z$	0	0	0	0
$p_k(000)$	$[1 - (m_{EX} + b_X - m_{EX}b_X)](1 - b_Y)(1 - b_Z)$	0	0	.250	.250
Feature probabilities					
$p_k(X)$		1	1.0	.750	.750
$p_k(Y)$		.800	1.0	.600	.750
$p_k(Z)$		.640	1.0	.480	.750
Interfeature associations[a]					
log $\theta_k(X, Y)$		8.29	20.7	8.29	13.8
log $\theta_k(Y, Z)$		8.29	20.7	8.29	13.8
log $\theta_k(X, Z)$		6.90	19.3	6.90	12.4

Note. The likelihood equations for the essentialized model assume $c_E = 1$. Equations for the unconnected model (see Figure 14B) were presented earlier in Table 2.
[a] The odds ratio is not defined when there are zero cells, a situation that arises in Table 6 when $m = 1.0$ or $b = 0$. For these cases, the log odds ratios presented in Table 6 are computed assuming $m = .999$ and $b = .001$.

A second basis for predicting a larger causal status effect in the Essentialized-Chain-80 condition stems from the fact that we expected that participants might construe the classification test as a causal reasoning task. For example, Rehder and Kim (2009a) found that a feature was more diagnostic of category membership when it was caused by an underlying essential feature because participants reasoned backwards from the observed to the essential feature and decided category membership on that basis. This implies a stronger causal status effect in the present Essentialized-Chain-80 condition because only feature X provides direct inferential support for E, contributing further to X's greater classification weight relative to Y and Z. Because of its focus on feature X, this effect of causal reasoning also suggests that coherence effects might be smaller in the Essentialized-Chain-80 condition than the Unconnected-Chain-80 condition.

The dependency model makes a different prediction for this experiment. Recall that the dependency model claims that a feature's centrality is determined by its dependents rather than its causes. Thus, providing feature X with an additional cause in the Essentialized-Chain-80 condition should have no influence on its centrality. For example, when $d_{XY} = d_{YZ} = 2$, the dependency model predicts feature centralities of 4, 2, and 1 after two iterations for features X, Y, and Z, respectively, for both the Essentialized-Chain-80 and Unconnected-Chain-80 conditions (see Figure 3A). Of course, the dependency model also predicts the absence of coherence effects.

Method

Materials. The materials were the same as in Experiments 1 and 2, with the exception of the essentialized feature for each category and the additional causal relationship linking it and feature X. The strengths of the causal links were as shown in Figure 14. Because the causal structure in Figure 14A implies that X should be present in all category members, to allow for this possibility, we introduced a minor change in wording in the description of all features. For example, rather than saying that "Most Myastars have high density, whereas some have low density" as in Experiments 1 and 2, we said "Most Myastars have high density. Any that don't have low density instead."

Procedure. Initial category learning followed the same general procedures as in Experiments 1 and 2. Because the essentialized feature was never presented during the classification and feature likelihood tests that followed, those tests were identical to the one in Experiments 1 and 2. In Experiment 3, participants were provided a diagram of the causal relations during the tests, and the order of those two tests was balanced over participants.

Participants. One hundred and eight New York University undergraduates received course credit for participating in this experiment. They were randomly assigned to the Essentialized-Chain-80, Unconnected-Chain-80, and Control conditions; to one of the six categories; and to one of the test presentation orders,

subject to the constraint that an equal number of participants appeared in each cell.

Results

The average category membership ratings in each condition are presented in Table 7. There were no effects of which category participants learned or of test order, and thus the results are collapsed over these factors.

Feature weights. The regression weights averaged over participants for features X, Y, and Z are presented in the left-hand panels of Figure 15. Recall that the generative model predicts a stronger causal status effect when features are causally related to an underlying category essence. Figure 15 confirms that the causal status effect was larger in the Essentialized-Chain-80 condition (a difference between X and Z feature weights of 19.1) as compared with the Unconnected-Chain-80 condition (a difference of 7.9).

A 3 × 3 ANOVA of the feature weights was conducted with condition (Essentialized-Chain-80, Unconnected-Chain-80, Control) and feature (X, Y, Z) as factors. There was an overall effect of condition on the linear trend, $F(2, 105) = 23.74$, $MSE = 74.7$, $p < .0001$. In separate analyses, the linear trend in the Essentialized-Chain-80 condition was significantly larger than in the Unconnected-Chain-80 condition, $F(1, 105) = 15.07$, $p < .001$, which in turn was larger than in the Control condition, $F(1, 105) = 8.4$, $p < .01$. In the 3 × 3 analysis, there was also a main effect of condition, $F(2, 105) = 13.76$, $MSE = 66.6$, $p < .0001$, reflecting that the weights in the Essentialized-Chain-80 condition (average of 12.1) were significantly greater than those in the Unconnected-Chain-80 condition (average of 8.2), $F(1, 105) = 12.28$, $p < .001$. The magnitude of the weights in the Essentialized-Chain-80 and Control conditions did not differ ($p = .11$).

Feature interactions. The interaction weights are presented in the right-hand panels of Figure 15. The generative model predicts that interaction weights should be larger in the causal conditions than in the Control condition and that directly linked feature pairs should have larger weights than the indirectly linked pair. Both of these predictions were confirmed. The direct and indirect interaction weights were 7.1 and 3.1, respectively, in the Essentialized-Chain-80 condition, and 10.9 and 7.9 in the Unconnected-Chain-80 condition, as compared with 1.1 and 1.0 in the Control condition. Moreover, in both causal conditions, the direct terms were larger than the indirect term. A 3 × 2 ANOVA of the interaction weights was conducted with condition and interaction type as factors. There was a main effect of condition, $F(2, 105) = 33.49$, $MSE = 38.0$, $p < .0001$, reflecting the larger interaction weights in the two causal conditions. In addition, the weights were significantly larger in the Unconnected-Chain-80 condition than in the Essentialized-Chain-80 condition, $F(1, 105) = 66.95$, $p < .0001$. There was also a significant main effect of interaction type, $F(1, 105) = 14.70$, $MSE = 21.2$, $p < .001$, and a significant interaction between interaction type and condition, $F(2, 105) = 3.66$, $MSE = 21.2$, $p < .05$, reflecting the difference in the direct and indirect interaction terms in the causal conditions.

Feature likelihood ratings. Feature likelihood ratings (see Figure 16) mirror the feature regression weights. Recall that the generative model predicts that the essentialized causal model in Figure 14A would result in feature X being viewed as highly prevalent in category members (because it is always produced by E) and that feature likelihoods would then decrease along the causal chain (see Table 6). This prediction was confirmed by likelihoods of 95.0% for X, 79.7% for Y, and 68.2% for Z in the Essentialized-Chain-80 condition. Also as predicted, likelihood ratings decreased less prominently in the Unconnected-Chain-80 condition (83.7%, 78.4%, and 70.9%). A 3 × 3 ANOVA of the feature likelihood ratings confirmed an effect of condition on the linear trend, $F(2, 105) = 29.5$, $MSE = 110.8$, $p < .0001$, reflecting the different pattern of ratings in the three conditions. The linear effect in the Essentialized-Chain-80 condition was significantly larger than that in the Unconnected-Chain-80 condition, $F(1, 105) = 16.26$, $p < .0001$, which in turn was larger than in the Control condition, $F(1, 105) = 13.29$, $p < .001$. In the 3 × 3 ANOVA, there was also a main effect of condition, $F(2, 105) = 3.98$, $MSE = 214.8$, $p < .05$.

Selected test items. As in Experiments 1 and 2, the causal status effect and the coherence effect can be observed directly in the classification ratings of individual test items. In the Essentialized-Chain-80 condition, the missing-X item was rated 45.7 points lower than the missing-Z item, reflecting a causal status effect (see Figure 17A). In contrast, the difference between those items was smaller (17.4) in the Unconnected-Chain-80 condition, reflecting a smaller causal status effect in that condition.

Second, the effect of coherence in both causal conditions is apparent in the pattern of test item ratings shown in Figure 17B. As was the case in Experiments 1 and 2, items with one or two characteristic features were rated lower in the causal conditions relative to the Control condition, because items with a mixture of

Table 7
Classification Ratings From Experiment 3

Test Item	Essentialized-Chain-80 condition	Unconnected-Chain-80 condition	Control condition
111	90.6 (3.1)	92.8 (1.6)	92.7 (2.1)
011	13.8 (3.6)	22.6 (4.0)	67.2 (2.1)
101	35.8 (5.0)	26.9 (4.3)	66.5 (1.9)
110	59.4 (4.8)	40.1 (4.3)	63.2 (2.3)
100	39.7 (4.1)	30.0 (4.3)	39.2 (3.1)
010	9.0 (2.4)	15.0 (2.3)	38.0 (3.3)
001	9.0 (2.2)	15.8 (2.6)	38.8 (2.7)
000	11.1 (3.2)	37.0 (5.4)	23.1 (3.7)

Note. Standard errors are shown in parentheses.

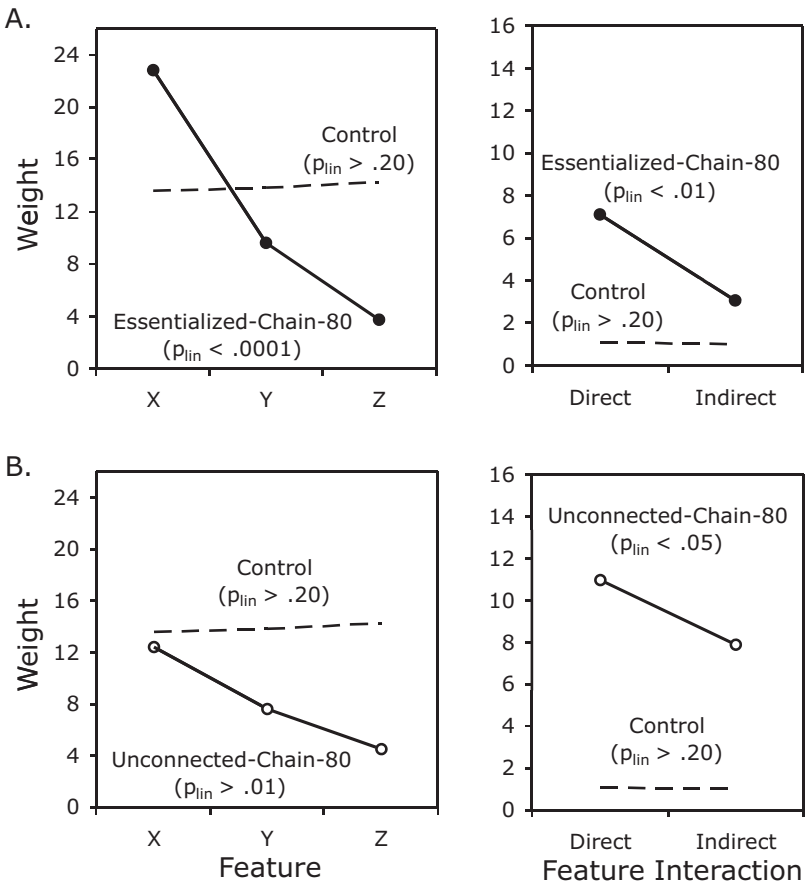

Figure 15. Results of regression analyses of classification ratings from Experiment 3: (A) Essentialized-Chain-80 condition versus Control condition and (B) Unconnected-Chain-80 condition versus Control condition. Left panels: feature weights; right panels: interaction weights; p_{lin} is the significance of the linear trend in each condition.

characteristic and uncharacteristic features are considered incoherent. Furthermore, Test Item 000 was rated higher in the Unconnected-Chain-80 than in the Control condition, reflecting that item's coherence in light of the causal laws. Unlike the previous experiments, Test Item 000 received a very low rating in the Essentialized-Chain-80 condition, a result that obtained because 000 is incoherent in that condition (because E always produces X, X should be present in all category members; see Table 6).

Model fitting. Following Experiments 1 and 2, we fit the generative and dependency models to the results of Experiment 3. For the dependency model, parameter d was free. Fits of the generative model to the Unconnected-Chain-80 condition used the likelihood equations in Table 2, assuming $m_{\text{XY}} = m_{\text{YZ}} = 0.80$, $b_{\text{Y}} = b_{\text{Z}} = 0$; c_{X} was free. To take into account the possible presence of causal reasoning in the Essentialized-Chain-80 condition, we fit that condition's ratings with a mixed model derived from (a) the equations for p_k in Table 6 plus (b) a factor that assigns all classification weight to feature X (because only it implies E). That is,

$$rating(t_i) = \beta_0 + \beta_1[wp_K(t_i; m_{EX}, m_{XY}, m_{YZ}, b_Y, b_Z) + (1 - w)f_{i,X}],$$

where t_i is a test item, $m_{\text{EX}} = 1$, $m_{\text{XY}} = m_{\text{YZ}} = .80$, and $b_{\text{Y}} = b_{\text{Z}} = 0$, and $f_{i,X} = 1$ when feature X is present in t_i and 0 when it is absent. Parameter w was constrained to be between 0 and 1 and assigns relative weight to the two submodels.

For the generative model, the best fitting parameters averaged over participants were $w = .75$, $\beta_0 = 10$, and $\beta_1 = 108$ in the Essentialized-Chain-80 condition, and $c_{\text{X}} = .85$, $\beta_0 = 18$, and $\beta_1 = 139$ in the Unconnected-Chain-80 condition. The predicted and observed ratings for each test item averaged over participants are presented in Figure 18 for each condition. The generative model achieves both a good qualitative and quantitative fit to these data, accounting for 95% and 93% of the variance in the observed ratings in Figure 18 in the Essentialized-Chain-80 and Unconnected-Chain-80 conditions, respectively (R^2s averaged over participants were 0.80 and 0.71; average *RMSE*s were 14.6 and 19.1). As in the previous experiments, the generative model tends to overpredict test items with one typical feature and to underpredict items with two. For example, in the Unconnected-Chain-80 condition, the classification ratings of Test Items 101 and 011 with two typical features were significantly higher (average of 24.8) than Items 010 and 001 with one (15.4) when, according to the generative model, the probability of those items should be the same (namely, zero; see Table

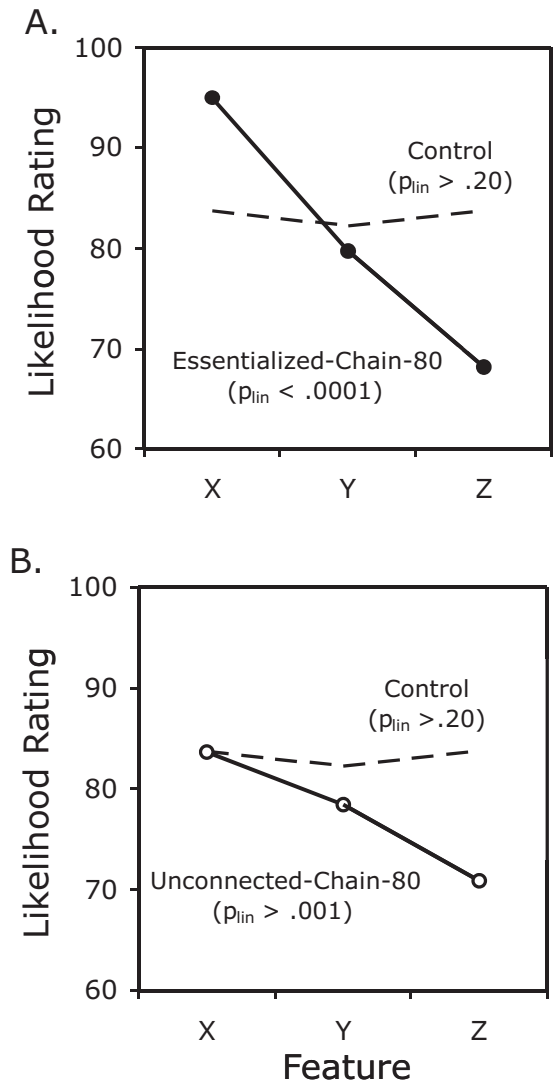

Figure 16. Feature likelihood ratings from Experiment 3: (A) Essentialized-Chain-80 condition versus Control condition and (B) Unconnected-Chain-80 condition versus Control condition; p_{lin} is the significance of the linear trend in each condition.

6). That the value of parameter w was <1 confirms the presence of causal reasoning in this experiment in which feature X receives an especially large classification weight (because only it implies the presence of the underlying essential feature).

The averaged parameter values for the dependency model were $d = 2.79$, $\beta_0 = 4.5$, and $\beta_1 = 7.2$ in the Essentialized-Chain-80 condition, and $d = 1.99$, $\beta_0 = 12.3$, and $\beta_1 = 10.5$ in the Unconnected-Chain-80 condition—fits that accounted for 80% and 42% of the variance, respectively (average R^2s of .66 and .40; average *RMSE*s of 21.1 and 28.0). As in Experiments 1 and 2, Figure 18B shows that the dependency model is unable to account for the qualitative pattern of responses in the Unconnected-Chain-80 condition. Because of its ability to account for the large weight on feature X, it achieved a relatively better fit to the Essentialized-Chain-80 condition, capturing some important qualitative trends in the data (e.g., the relatively high rating received by Item 100). However, because of it inability to account for coherence effects, it overpredicts especially incoherent items (e.g., 010 and 101) and underpredicts especially coherent ones (e.g., 111). As a consequence, the quality of this fit lags behind that of the generative model (average *RMSE*s of 21.1 vs. 14.6).

Discussion

As predicted by the generative model, introducing an explicit essential feature led to a larger causal status effect as indicated by a stronger linear effect of features in the Essentialized-Chain-80 condition as compared with the Unconnected-Chain-80 condition. The effect was observed in both feature regression weights and likelihood ratings. The dependency model, in contrast, predicts identical feature weights in the two conditions.

We suggest there were two reasons for the especially large causal status effect in the Essentialized-Chain-80 condition. First, because X was deterministically caused by E, it should be present in all category members, which in turn should yield a larger causal status effect. That feature X was rated as appearing in almost all category members (95%) provides direct support for this prediction. Second, because X but not Y or Z are directly linked to E, only it provides direct

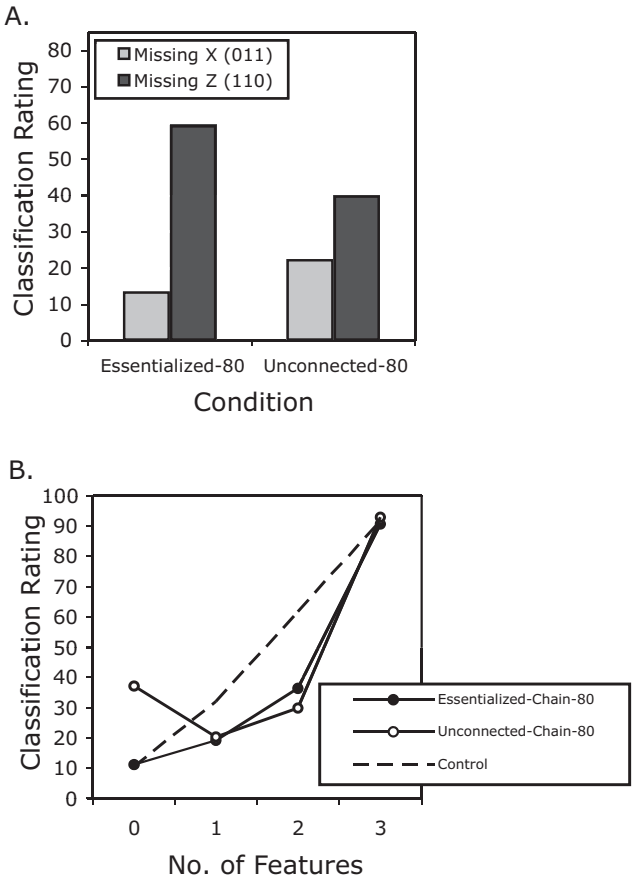

Figure 17. Classification ratings from Experiment 3: (A) for test items missing only feature X and only feature Z and (B) for all test items as a function of their number of characteristic features.

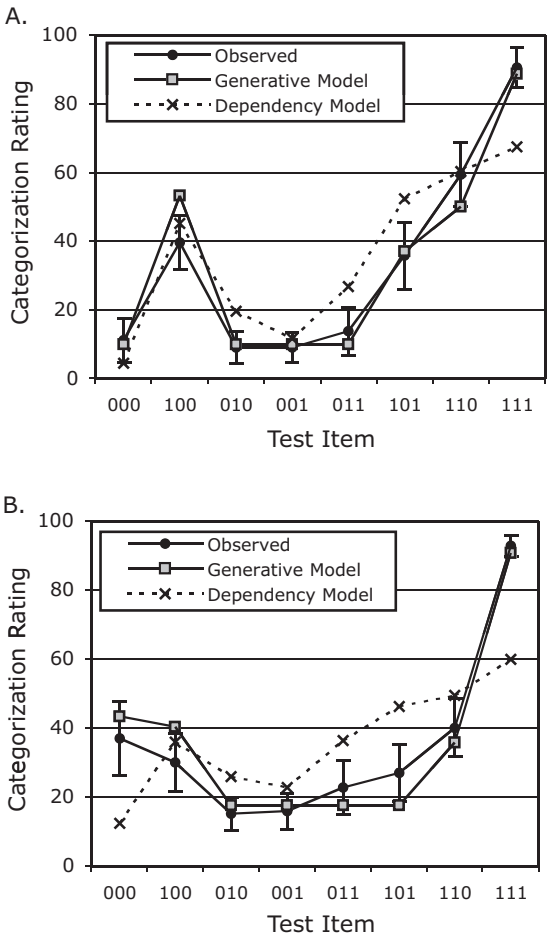

Figure 18. Fits of the generative and dependency models to the classification ratings of Experiment 3: (A) Essentialized-Chain-80 condition and (B) Unconnected-Chain-80 condition. Error bars are 95% confidence intervals.

inferential support for E. Model fitting confirmed that this effect of causal reasoning conferred yet additional importance to feature X.

Experiment 3 again confirmed the presence of a coherence effect in both causal conditions. The generative model successfully predicted the large interaction weights in those conditions and the fact that the direct interaction terms were larger than the indirect term. One unexpected result was the larger interaction weights in the Unconnected-Chain-80 condition than in the Essentialized-Chain-80 condition. We think this occurred for two reasons. The first concerns how participants used the response scale. Because feature weights were larger in the Essentialized-Chain-80 condition than in the Unconnected-Chain-80 condition, this led to less of the response scale being available to express the coherence effect in the Essentialized-Chain-80 condition. The second concerns the presence of causal reasoning in which only X matters to category membership, an effect that reduces the importance of confirming the X $\rightarrow$ Y and Y $\rightarrow$ Z causal links.

As we did in Experiments 1 and 2, to address any concern that the reported differences between the Essentialized-Chain-80 and Unconnected-Chain-80 conditions were an artifact of differences in response scale usage, we tested a variant of Experiment 3 in which the manipulation of alternative causes was within-subjects. The results of this experiment, reported in Appendix C, replicate Experiment 3's key result, namely, a larger causal status effect when a cause feature is itself caused by an essential one.

Experiment 4

Experiment 3 demonstrated how a causal status effect is promoted by causally relating a category's observable features to an essential feature. However, according to the generative model, the magnitude of this effect depends on the strengths of the X $\rightarrow$ Y and Y $\rightarrow$ Z causal links. When those links are probabilistic, then each subsequent feature should be generated with decreasing probability. However, when they are deterministic, then Y should be at least as prevalent as X and Z at least as prevalent as Y. Thus, if feature X is present in all category members (e.g., because it is always caused by an essential feature), then Y and Z should be too. That is, the large causal status effect found in Experiment 3's Essentialized-Chain-80 condition should largely disappear when causal links are deterministic.

To test this prediction, in Experiment 4 we tested Essentialized-Chain-100 and Unconnected-Chain-100 conditions that were identical to the Essentialized-Chain-80 and Unconnected-Chain-80 conditions of Experiment 3, except that the strengths of all causal links were 100%. Because Experiment 3 has already confirmed that feature weights in a condition with an essential feature but no causal links do not differ from one another, we did not replicate this Control condition in Experiment 4. The quantitative predictions of the generative model for the cases when the *m*s = 1.0 are presented in Table 6. The table confirms the lack of a causal status effect in both the Essentialized-Chain-100 and Unconnected-Chain-100 conditions—that is, $p_k(X) = p_k(Y) = p_k(Z)$. It also indicates that all three features should be weighed more in the Essentialized-Chain-100 condition (p_ks of 1.0 vs. 0.75 in the Unconnected-Chain-100 condition). Recall that Experiment 3 also demonstrated that participants exhibited an effect of causal reasoning (in which X is conferred additional weight because only it directly implies E), suggesting that one might expect a small causal status effect in the Essentialized-Chain-100 condition. Nevertheless, this effect should be much smaller than in Experiment 3's Essentialized-Chain-80 condition.

The dependency model makes a different set of predictions for this experiment. Because the dependency model predicts that stronger links result in a larger causal status effect, it predicts a larger causal status effect in the Essentialized-Chain-100 and Unconnected-Chain-100 conditions as compared with the corresponding conditions in Experiment 3. It also predicts the absence of coherence effects.

Method

The materials and procedure were identical to those in Experiment 3, except for the causal link strengths of 100%. Forty-eight New York University undergraduates received course credit for participating in this experiment. They were randomly assigned to the Essentialized-Chain-100 and Unconnected-Chain-100 conditions, to one of the six categories, and to one of the task presentation orders subject to the constraint that an equal number of participants appeared in each cell.

Results

The average category membership ratings in each condition are presented in Table 8. There were again no effects of which

Table 8
Classification Ratings From Experiment 4

Test Item	Essentialized-Chain-100 condition	Unconnected-Chain-100 condition
111	94.2 (2.3)	95.7 (1.8)
011	13.3 (4.1)	17.5 (5.2)
101	15.0 (4.3)	14.1 (4.7)
110	16.8 (4.1)	17.1 (5.6)
100	8.6 (2.6)	8.6 (2.8)
010	10.2 (3.1)	7.2 (2.2)
001	8.8 (3.0)	8.5 (2.8)
000	15.8 (5.4)	46.1 (8.2)

Note. Standard errors are shown in parentheses.

category participants learned or of test order, and thus the results are presented collapsed over these factors.

Feature weights. The regression weights averaged over participants are presented in the left-hand panel of Figure 19. In fact, neither of the two causal conditions yielded a causal status effect. A 2 × 3 ANOVA with condition (Essentialized-Chain-100 and Unconnected-Chain-100) and feature (X, Y, Z) as factors yielded only a main effect of condition, $F(1, 46) = 4.40$, $MSE = 89.8$, $p < .05$, reflecting the fact that the weights were larger in the Essentialized-Chain-100 condition than in the Unconnected-Chain-100 condition. In this analysis, there was no linear effect of feature and no interaction (both Fs < 1).

Feature interactions. The interaction weights are presented in the right-hand panel of Figure 19. A 2 × 2 ANOVA of the interaction weights revealed a marginally significant main effect of condition, $F(1, 46) = 3.44$, $MSE = 96.2$, $p = .07$, reflecting the larger interaction weights in the Unconnected-Chain-100 condition as compared with the Essentialized-Chain-100 condition. There was no effect of interaction type ($F < 1$), reflecting the fact that the indirect interaction term was not smaller than the direct ones.

Feature likelihood ratings. Like the feature regression weight, the feature likelihood ratings presented in Figure 20 indicate the absence of a causal status effect in both conditions. As predicted, features were rated as more frequent in the Essentialized-Chain-100 condition (98.1) as compared with the Unconnected-Chain-100 condition (89.7). A 2 × 3 ANOVA revealed a main effect of condition, $F(1, 46) = 16.5$, $MSE = 181.7$, $p < .001$; no effect of feature ($F < 1$); and no interaction ($F < 1$).

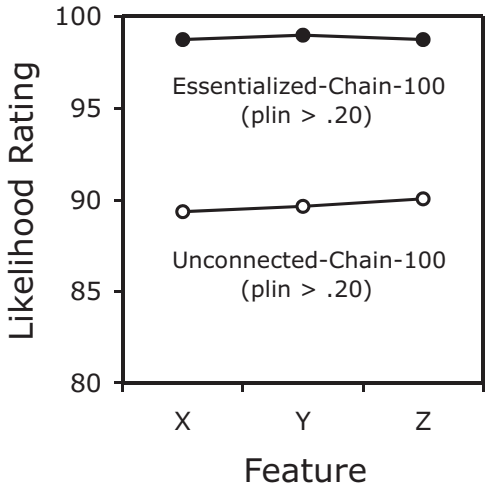

Figure 20. Feature likelihood ratings from Experiment 4; p_{lin} is the significance of the linear trend in each condition.

Selected test items. The importance of coherence in this experiment is apparent in the test item classification ratings. In the Essentialized-Chain-100 condition, the average ratings for items with zero, one, two, and three characteristic features were 15.8, 9.2, 15.0, and 94.2, respectively, a pattern that reflects that all items except for the Prototype 111 are incoherent in that condition (see Table 6). In contrast, those ratings were 46.2, 8.1, 16.3, and 95.8 in the Unconnected-Chain-100 condition, reflecting that Test Items 111 and 000 are the only two coherent items in that condition. Once again, when causal relations link a category's characteristic features, participants expect category members to reflect the correlations that causal relations generate.

Model fitting. The generative and dependency models were fit in the same manner as Experiment 3 except that $m_{XY} = m_{YZ} =$

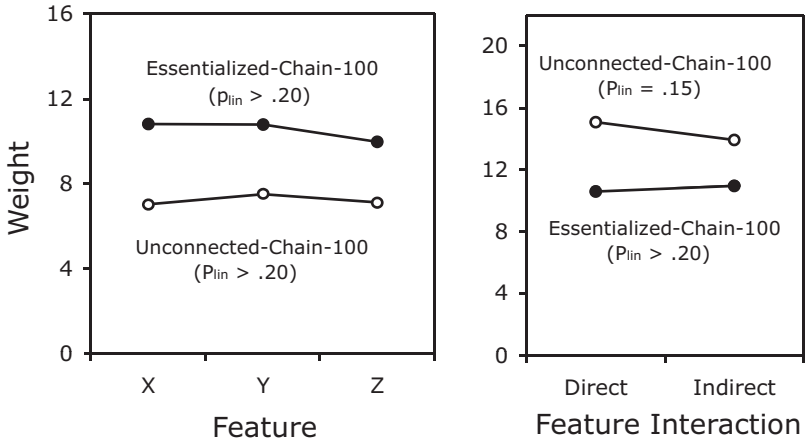

Figure 19. Results of regression analyses of classification ratings from Experiment 4. Left panel: feature weights; right panel: interaction weights; p_{lin} is the significance of the linear trend in each condition.

1 for the generative model fits. The best fitting parameters averaged over participants for the generative model were $w = .93$, $\beta_0 = 11$, and $\beta_1 = 84$ in the Essentialized-Chain-100 condition, and $c_X = .74$, $\beta_0 = 11$, and $\beta_1 = 128$ in the Unconnected-Chain-100 condition. The predicted and observed ratings for each test item averaged over participants are presented in Figure 21. The generative model provides both good qualitative and quantitative accounts of these data, accounting for 98% and 97% of the variance in the observed ratings in Figure 21 in the Essentialized-Chain-100 and Unconnected-Chain-100 conditions, respectively (R^2s averaged over participants were .85 and .86; average *RMSE*s were 10.7 and 9.3). In the Unconnected-Chain-100 condition, the generative model again tends to overpredict test items with one typical feature and to underpredict items with two. This reflects that the classification ratings of test items with two typical features were significantly higher (average of 16.2) than those with one (8.1) when, according to the generative model, those items should have the same (zero) rating. As in Experiment 1, a post hoc explanation for these failed predictions involves a subgroup of participants who ignored the causal relations and responded like the control participants, thus inflating the ratings of the two-feature items relative to the one-feature items (see Appendix A). In the Essentialized-Chain-100 condition, a value of w closer to 1 suggested that participants were less likely to treat the classification task as a causal reasoning task (in which only X is relevant to category membership) as compared with Experiment 3's Essentialized-Chain-80 condition ($w = .93$ vs. .75).

For the dependency model, the averaged parameter values were $d = 1.15$, $\beta_0 = 5.1$, and $\beta_1 = 12.5$ in the Essentialized-Chain-100 condition, and $d = 1.20$, $\beta_0 = 9.8$, and $\beta_1 = 11.3$ in the Unconnected-Chain-100 condition. As is apparent from Figure 21, the dependency model is unable to account for the qualitative pattern of responses in either condition. As a result, it accounted for only 39% and 18% of the group-level variance (average R^2s were .37 and .28; average *RMSE*s were 29.5 and 35.7).

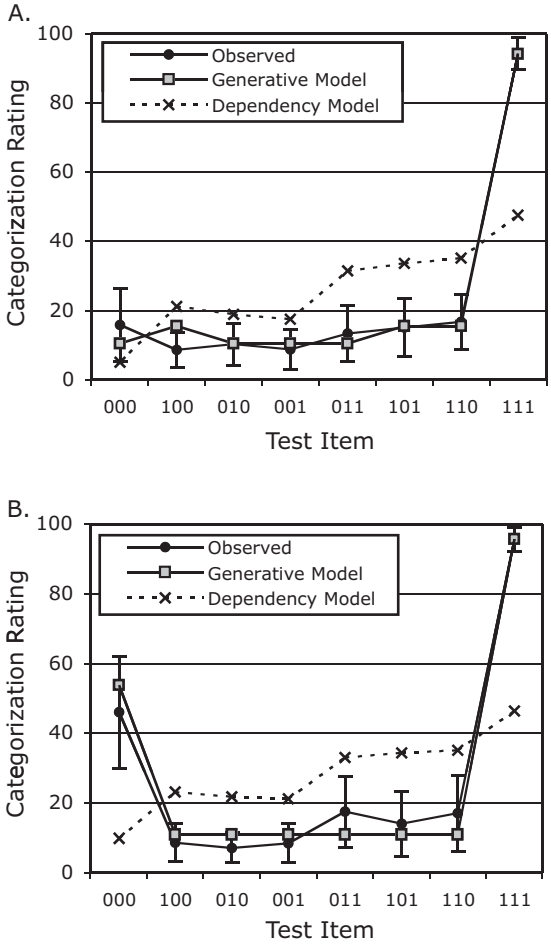

Figure 21. Fits of the generative and dependency models to the classification ratings of Experiment 4: (A) Essentialized-Chain-100 condition and (B) Unconnected-Chain-100 condition. Error bars are 95% confidence intervals.

Discussion

As predicted by the generative model, the increase in the magnitude of the causal status effect found with "essentialized" categories in Experiment 3 was much smaller than in Experiment 4; indeed, the causal status effect was absent entirely in the present experiment's Essentialized-Chain-100 condition. This result was obtained for both feature regression weights and likelihood ratings. The causal status effect was also absent in the Unconnected-Chain-100 condition, a result that replicates Experiment 1's Chain-100 condition that tested the same causal structure. The dependency model, in contrast, predicted that the causal status effect should be larger in both conditions of this experiment than in the corresponding conditions of Experiment 3 (in which the causal links were weaker) when in fact it was smaller (i.e., zero). Another successful prediction of the generative model is that feature regression weights and likelihood ratings were larger in the Essentialized-Chain-100 condition than the Unconnected-Chain-100 condition, consistent with the prediction that features X, Y, and Z were more likely to be "generated" in the former condition.

Another difference between the essentialized conditions of Experiments 3 and 4 is that whereas the former experiment demonstrated a substantial effect of causal reasoning in which X is especially diagnostic of category membership (because only it implies E), this effect was smaller in the current experiment. One possible reason for this difference is that the deterministic links used in Experiment 3 emphasized the importance of coherence among the observable features X, Y, and Z and thus highlighted that the prototype (111) is the only acceptable category member in light of causal laws.

As in Experiments 1–3, the present experiment yielded strong coherence effects. One unexpected result is the larger interaction weights in the Unconnected-Chain-80 condition than in the Essentialized-Chain-80 condition; however, again we think that the larger feature weights in the Essentialized-Chain-80 condition led to less of the response scale being available to express the coherence effect. Overall, the generative model provides a good account of the results from Experiment 4.

General Discussion

In this article, we have addressed the changes in classification performance brought about by causal knowledge that links cate-

gory features. In the first two sections, we review those conditions that promote a causal status effect and the implications those results have for alternative models. In the second section, we discuss how the causal status effect is mediated by changes in features' subjective category validity. We close with discussion of the coherence effect and limitations and extensions of our model.

Causal Status and Causal Models

One contribution of this research is that it has demonstrated the value of the causal model approach to explaining how causal knowledge affects the importance of individual features to classification. Whereas previous research has generally been limited to testing different topologies of interfeature causal links, in the present study we tested how the causal status effect varies as function of the parameters of a category's causal model. In Experiment 1, we manipulated the strength of the causal relationships and found a causal status effect for probabilistic but not deterministic causal links. In addition, when categories were essentialized, a causal status effect was found when the links were probabilistic (Experiment 3) but not when they were deterministic (Experiment 4). Finally, in Experiment 2, we manipulated another important parameter—the probability of background causes—and found, as predicted, a stronger causal status effect when such causes were absent versus present.

Why should these parameters influence the size of the causal status effect? Stated intuitively, what we think is going on is this. When confronted with a causal network of features, classifiers will often adopt a "generative" perspective, that is, they will think about the likelihood of each successive event in the chain. This process may be equivalent to a kind of mental simulation in which they repeatedly "run" (consciously or unconsciously) a causal chain (Kahneman & Tversky, 1982). Feature probabilities are then estimated by averaging over runs. Of course, in a run, the likelihood of each subsequent feature in the chain increases as a function of the strength of chain's causal links. When links are deterministic, then all features will be present whenever the chain's root cause is; in such cases, no causal status effect appears. However, when causal links are probabilistic, each subsequent feature in the causal chain is generated with lower probability, and thus a causal status effect arises. However, working against this effect is classifiers' beliefs about the strength of alternative background causes. Background causes will raise the probability of each feature in the causal chain in each simulation run and, if sufficiently strong, will cancel out (and possibly even reverse) the causal status effect.

The dependency model, in contrast, is based on a competing intuition that features are important in people's conceptual representations to the extent they are responsible for other features. For example, DNA is more important than the color of an animal's fur because so much depends on DNA; hormones are more important than the size of its eyes for the same reason. However, despite the plausibility of this intuition, it is does not conform to our participant's category membership judgments. Whereas the dependency model predicts that causal features should be weighed more heavily as a function of how heavily their effects depend on them, we found instead that the causal status effect was weaker for deterministic causal links and stronger for probabilistic ones. Furthermore, whereas the dependency model predicts that features' weights should be unaffected by the introduction of additional *causes*, we found instead a weaker causal status effect when background causes were present.

Our demonstration that the causal status effect depends on the strength of the causal links sheds light on the dramatic differences in the size of that effect across studies. As mentioned, whereas Ahn et al. (2000) observed a difference of 35 points in the rating of the item missing only X versus the one missing only Z, that difference was an order of magnitude smaller in Rehder and Kim (2006). However, even though these studies tested some of the same materials, they systematically differed in the wording of the causal links. For example, both studies tested Romanian Rogos (a type of automobile) with features "butane-laden fuel" and "hot engine temperature"; however, whereas Ahn et al. described the causal relationship between the two as "Butane-laden fuel tends to cause hot engine temperatures," Rehder and Kim omitted the phrase "tends to." This difference might have led Rehder and Kim's participants to interpret their causal links as more deterministic, conditions under which the generative model predicts no causal status effect. In support of this interpretation, as part of another experiment testing Rehder and Kim's materials, we asked participants to judge how often a cause produced its effect. The average response was 91% (the modal response was 100%), and as we have seen, the generative model predicts a very weak or absent causal status effect with such strong causal links.

Given that the causal status effect is also reduced by the presence of background causes, it is interesting to speculate about classifiers' default assumptions regarding such causes. For example, is a statement about Myastars that "A hot temperature causes the star to have high density," interpreted by default to mean that hot temperature was the *only* cause of high density? In fact, there is evidence suggesting that a starting assumption of no alternative causes may be the default in many learning and reasoning situations. For example, Lu, Yuille, Liljeholm, Cheng, and Holyoak (2008) have proposed a model that explains certain causal learning results by assuming that people initially assume that absence of background causes. Furthermore, Rehder and Milovanovic (2007) found that participants overestimated the base rate of a cause event, apparently to avoid the need to assume alternative causes of an effect. Of course, if alternative causes are assumed to be absent, an enhanced causal status effect will result.

The present results add to others supporting the generative model over the dependency model as an account of how causal knowledge affects feature importance. As mentioned, Rehder and Kim (2006) systematically manipulated features' number of dependents and found that their importance increased with their number of causes (as predicted by the generative model) but not with their number of dependents (as predicted by the dependency model; also see Rehder, 2003a). Our results also speak against any model that represents causal links as a symmetric relation. For example, Rehder and Murphy (2003) proposed the Knowledge-Resonance (KRES) recurrent connectionist model that represents relations between category features as excitatory links and by so doing accounts for a number of known effects of knowledge on category learning. However, because the excitatory links are symmetric, features X and Z in the three-element causal chain in Figure 1 would have equivalent representations, and thus there would be no basis to predict that one feature would be more important than the other. In other words, the causal status effect is an example of how people can reason causally, and, unlike KRES,

the generative model explains this effect by taking into the account the asymmetries inherent in such relations (Rehder, 2003a; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005; Waldmann & Holyoak, 1992; Waldmann, Holyoak, & Fratianne, 1995).

Causal Status and Hidden Features

Besides the importance of causal link strength and background causes, our third main finding concerns the effect of hidden features on the causal status effect. The generative model generally predicts that when categories are "essentialized," directly caused features (X) should be generated with greater probability that in turn leads to a sharper decrease in probabilities along the causal chain—and in fact essentialized categories led to a large causal status effect in Experiment 3. In contrast, when causal links are deterministic, indirectly caused features (Y and Z) should be at least as probable as X—and in fact in Experiment 4 we found no causal status effect. Because the dependency model predicts that a feature's weight should increase with its number of effects, it is unable to account for how introducing an essential feature as a *cause* should have any effect an feature importance.

Other studies provide support for the view that the causal status effect is promoted by hidden features. For example, Rehder (2003b) also found that essentialized categories led to a larger causal status effect—the current results replicate that finding and also extend them by showing how that effect is mediated by causal link strength. Furthermore, Ahn and colleagues have found that expert clinicians both view mental disorders as less essentialized than laypersons (Ahn, Flanagan, Marsh, & Sanislow, 2006) and exhibit only a weak causal status effect (Ahn, Levin, & Marsh, 2005).

Although we focused on the power of an essential feature to generate a larger causal status effect, note that an underlying feature would produce that effect even if it was only highly diagnostic of, but not truly essential to, category membership. This prediction is important because the question of whether real-world categories are essentialized is a controversial one. Although good evidence exists for the importance of underlying properties to category membership (Gelman, 2003; Keil, 1995; Rips, 2001), Hampton (1995) has demonstrated that even when biological categories' so-called essential properties are unambiguously present (or absent), characteristic features continue to exert an influence on judgments of category membership (also see Braisby, Franks, & Hampton, 1996; Kalish, 1995; Malt, 1994; Malt & Johnson, 1992). Our own suspicion is that although the unobserved properties of many categories are distinctly important to category membership, few may be truly essential (see Rehder, 2007, for discussion). However, according to the generative model, all that is required for an enhanced causal status effect is that the unobserved property be highly diagnostic of category membership.

Recall that the generative model implies two distinct mental processes through which an essentialized category might influence categorization judgments. The first is that the essential feature will change participants' beliefs about the likelihood of features within a category, and, as we have seen, feature likelihood ratings in Experiments 3 and 4 provided direct evidence for such a change. However, these effects may be augmented by causal reasoning in which classifiers reason explicitly from observed features to the essential one. In other words, when categories are essentialized, classification involves an act of causal reasoning in which one reasons from evidence (effects) to the ultimate cause, in the same way that one reasons diagnostically from a disease's symptoms to the disease itself. On this account, directly caused features should provide greater evidence for category membership (because only they directly imply the essential features) than indirectly caused ones, and model fitting results in Experiment 3 (and to a lesser extent Experiment 4) confirmed an especially large classification weight on feature X.

Evidence for causal reasoning in support of classification has been found in other studies. For example, Hampton, Estes, and Simmons (2007) found that whether a transformed animal was judged to have changed category membership often depended on what participants could infer about underlying causal processes and structures. Furthermore, Rehder and Kim (2009a) found direct evidence that classification can involve a kind of diagnostic reasoning from observable features to underlying ones that are decisive for determining category membership. There is also reason to believe that classifiers reason not only backwards, or *diagnostically*, but also forwards, or *prospectively*, from features to the underlying properties they cause. For example, a physician may suspect the presence of HIV given the certain forms of sarcoma, lymphoma, and pneumonia that HIV is known to produce. However, the case for HIV is made stronger still by the presence of one or more of its known *causes*, such as blood transfusions, sharing of intravenous needles, or unsafe sex (for related evidence, see Chaigneau et al., 2004; Oppenheimer & Tenenbaum, 2009; Rehder, 2007).

Category Validity Mediates the Causal Status Effect

Another main finding concerns how the change to features' categorization importance brought about by causal knowledge is mediated by their subjective category validity. According to the generative model, causal knowledge changes the subjective likelihood with which a feature is generated by a category's causal model, and any feature that occurs with greater probability among category members (i.e., has greater category validity) should provide greater evidence in favor of category membership (Rosch & Mervis, 1975). Consistent with this prediction, in every condition in which a causal status effect obtained, participants also rated feature X as more frequent than Y and Y as more frequent than Z. Conversely, a causal status effect was never observed when feature likelihood ratings were not significantly different from one another.

Other studies have shown that a feature's influence on categorization judgments correlates with its subjective category validity. For example, Sloman et al. (1998) conducted a factor analysis that showed that category features vary along three dimensions. The first two were identified as diagnosticity (or *cue validity*, assessed with questions like "Of all things that grow on trees, what percentage are apples?") and perceptual salience (assessed with questions like "How prominent in your conception of apples is that it grows on trees?"). Measures that loaded on the third factor included ones related to a construct labeled *conceptual centrality* or *mutability* (assessed with questions like "How good an example of an apple would you consider an apple that does not ever grow on trees?") and category validity (e.g., "What percentage of apples grow on trees?"). Moreover, although Sloman et al. (1998, Study

5) found that features' judged mutability dissociated from their true category validity (the probability that they appeared in observed category members), they tracked participants' subjective *judgments* of category validity, further supporting the conclusion that mutability and subjective category validity are based on the same underlying construct. Because categorization ratings in the present Experiments 1–4 were likely to reflect neither diagnosticity (because they were generated with respect to a single category) nor salience (because the features used in six novel category were unlikely to vary systematically in their salience), we think our measure of features' categorization importance, or weight, was also assessing conceptual centrality. Furthermore, like Sloman et al., we found that that measure did not dissociate from subjective category validity.

Some studies have claimed to show just such a dissociation, however. For example, in Ahn et al. (2000, Experiment 2), participants first observed exemplars with three features that appeared with equal frequency and then rated the likelihood of each feature. They then learned causal relations forming a causal chain and rated the goodness of missing-X, missing-Y, and missing-Z test items. Whereas features' likelihood ratings did not differ, the missing-X item was rated lower than the missing-Y item, which was lower than the missing-Z item—a result the authors interpreted as demonstrating a dissociation between category validity and categorization importance. This conclusion is unwarranted, however, because the likelihood ratings were gathered *before* the presentation of the causal relations. Clearly, one can only assess whether perceived category validity mediates the relationship between causal knowledge and features' categorization importance by assessing category validity after the causal knowledge has been taught. Both Sloman et al. (1998, Study 5) and the present experiments gathered likelihood ratings after the causal relationships were learned and found no dissociation with centrality measures.

The Importance of Coherence in Causal-Based Categorization

This article's last major finding concerns coherence effects. The coherence effect was highly robust, appearing in every condition tested. This effect was so pronounced that it led participants to generate higher categorization ratings for items with fewer typical features. For example, in the causal conditions of Experiment 1 and 2, an item with zero typical features (000) was rated higher than items with one typical feature (and sometimes higher than those with two) because 000 was a sensible combination of features in light of causal laws. In contrast, items with a combination of typical and atypical features were inconsistent with those laws. This does not mean that participants ignored typicality altogether because the item with three typical features (111) was always rated much higher than 000. However, it does mean that causal knowledge changed participants' beliefs about the likely distribution of features so that coherent items (111 and 000) became the most likely category members.

Our participants' category membership judgments also exhibited sensitivity to the more subtle pattern of interfeature correlations one expects from causal relationships. Those correlations should get stronger as (a) the causal links get stronger and (b) alternative causes get weaker, predictions that were confirmed in Experiments 1 and 2. Furthermore, except when causal links are deterministic (and there are no other causes of Y and Z), directly related features (X and Y, and Y and Z) should be more strongly correlated than the indirectly related ones (X and Z), predictions that were confirmed in Experiments 1–3. These results support the claim that the likelihood that a test item was generated by a category's causal model forms the basis for classifiers' judgments of category membership. In contrast, the dependency model fails to predict coherence effects.

Note that robust coherence effects obtained in the present experiments despite Marsh and Ahn's (2006) suggestion that coherence effects are inflated when atypical dimension values are described as "normal." In their study, Marsh and Ahn compared an Unambiguous condition in which the uncharacteristic value on each binary dimension was the opposite of the characteristic value (e.g., *low* density vs. high density) with an Ambiguous condition (intended to be a replication of Rehder, 2003b, Experiment 2) in which uncharacteristic values were described as "normal" (e.g., *normal* density). They found that the Unambiguous condition yielded a larger causal status effect and a smaller coherence effect, a result they interpreted as demonstrating that the "normal" wording exaggerates coherence effects. However, this conclusion is unwarranted because the two conditions also differed on another dimension, namely, only the Unambiguous participants were given information about which features were typical of the category. In the absence of such typicality information, it is unsurprising that ratings in the Ambiguous condition were dominated by coherence. Rehder and Kim (2008) instead compared two conditions that were identical except for the "normal" wording and found exactly the opposite results: The "normal" wording produced a *smaller* coherence effect and a *larger* causal status effect. These results combined with the present experiments provide decisive evidence that coherence effects are not a "methodological artifact" of stimulus materials tested (Marsh & Ahn, 2006, p. 561).

The importance of coherence to classification has been found in other studies. For example, Wisniewski (1995) found that certain artifacts were better examples of the category "captures animals" when they possessed certain combinations of features (e.g., "contains peanuts" and "caught a squirrel") but not others ("contains acorns" and "caught an elephant"; also see Murphy & Wisniewski, 1989). Similarly, Rehder and Ross (2001) showed that artifacts were better examples of a category of pollution cleaning devices when their features cohered (e.g., "has a metal pole with a sharpened end" and "works to gather discarded paper") versus when they did not ("has a magnet" and "removes mosquitoes"). Malt and Smith (1984) found that judgments of typicality in natural categories were sensitive to whether items obeyed or violated theoretically expected correlations (also see Ahn, Marsh, Luhmann, & Lee, 2002).

Coherence affects other types of category-related judgment. Rehder and Hastie (2004) found that a novel property was more likely to be generalized to a category when displayed by a coherent versus incoherent exemplar. Patalano and Ross (2007) found that the generalization strength of a novel property from some category members to another varied as a function of the category's overall coherence (and found the reverse pattern when the generalization was made to a noncategory member). Finally, numerous studies have demonstrated how theoretical knowledge that links category features alters how categories are learned, both when learning is supervised (Murphy & Allopenna, 1994; Rehder & Ross, 2001;

Waldmann et al., 1995; Wattenmaker, Dewey, Murphy, & Medin, 1986) and unsupervised (Ahn & Medin, 1992; Kaplan & Murphy, 1999; Medin, Wattenmaker, & Hampson, 1987).

It is illuminating to assess the relative importance of the causal status and coherence effects in this study by comparing the proportion of the variance in categorization ratings induced by causal knowledge that can be attributed to each effect. In this calculation, the total variance induced by causal knowledge was taken to be the additional variance explained by a regression model with separate predictors for each feature and each two-way interaction as compared with a model with only one predictor representing the total number of characteristic features in a test item. The single predictor model is used as a baseline because each feature was described as occurring in "most" category members and thus in the absence of causal knowledge classification ratings are a simple function of the number of characteristic features displayed by an exemplar. The variance attributable to the causal status effect is the additional variance explained by the separate predictors for each feature in the full model, whereas the variance attributable to the coherence effect is the additional variance explained by the interaction terms. In fact, in the present experiments, the coherence effect accounted for >90% of the variance in seven of the eight causal conditions tested in this study.[6] Moreover, coherence has dominated participants' categorization judgments in every other study in which it has been assessed: 60% in Rehder and Hastie (2001, Experiment 2), 80% in Rehder (2003a), 82% in Rehder (2003b, Experiment 1), and 70% in Rehder and Kim (2006). Moreover, despite Marsh and Ahn's (2006) claim that "individual features' causal status, rather than feature combinations, was the predominant determinant" of categorization performance in their Unambiguous condition (p. 566), coherence accounted for 64% of the variance even in that condition as well. Indeed, their suggestion that "a strong case for the role of inter-feature links has yet to be made" (p. 566) is puzzling given that such a "strong case" was present in their own data. Instead, these analyses indicate that the most important factor in causal-based classification is whether an object displays a configuration of features that make sense in light of those laws.

Limitations and Extensions

In this final section, we consider potential limitations of the generative model as an account of causal-based classification. One systematic deviation between its predictions and the observed ratings was that it tended to overpredict items with few typical features and underpredict those with two. Said differently, it tended to put insufficient weight on an item's number of characteristic features in predicting category membership. Our analysis of individual differences suggested that this partly occurred because participants were given two sources of information for making their judgments (causal laws and the typicality information provided by telling participants that features appeared in "most" category members). A substantial minority chose to base them on only one of those sources (and those that chose only typicality contributed to the generative model's incorrect predictions). It is possible that this behavior is an artifact of an experimental situation in which participants consider fewer sources of information to minimize cognitive effort. Alternatively, it may be that real-world acts of classification also tend to be based on one criterion and that which is chosen varies depending on the conditions under which the judgments is made (e.g., Smith & Sloman, 1994; cf. Luhmann, Ahn, & Palmeri, 2006).

There are other conditions under which people's classification judgments might diverge from the generative model. First, as mentioned, our claim is that the generative model's likelihood equations are intended to approximate the intuitive mental processes that assign category membership. Accordingly, one might expect observed and predicted ratings to diverge as the size of category's causal network grows (and thus so do the complexity of the likelihood equations). Second, the generative model's predictions are based on a veridical representation of probability, but the distortions inherent in subjective probabilities are well documented (e.g., Kahneman & Tversky, 1979). Such distortions are likely to affect causal-based classifications as well.

Finally, we also mention a few obvious extensions to the generative model. First, whereas the goal of the present experiments was to examine the effects of causal knowledge in isolation, people also *observe* category members first hand, and of course it is important to specify how causal knowledge and observations are integrated in category representations. With its roots in a theory of mental representation that grew out of the causal learning literature (e.g., Cheng, 1997), the generative model makes clear predictions regarding how causal models should be updated in light of data. Although some work has been conducted in this direction (e.g., Rehder & Hastie, 2001; Rehder & Milovanovic, 2007; Waldman et al., 1995), more is needed. Nevertheless, even the present experiments provide some evidence that participants' beliefs about the causal laws were affected by the small amount of "empirical" category information that was provided. Recall that each feature was described as occurred in "most" category members, and one might expect that the identical wording might have biased participants to expect that features had the same category validity. Consistent with this interpretation, model fitting in Experiment 1 revealed that participants assumed stronger alternative causes in the Chain-75 condition with weaker interfeature causal links than the Chain-100 condition with stronger ones. On this account, Chain-75 participants compensated for the weaker causes by assuming stronger alternative causes to make the features' base rates more equal.

Second, although these studies have asked participants to judge category membership with respect to a single category, real-world acts of classification usually involve choosing which of several categories an object belongs to. In other work, we have applied the generative model to such situations by assuming that the evidence an object provides for a category is given by the familiar ratio rule (the likelihood that an object was generated by a category divided the sum of the evidence that it was generated by all categories, adjusted by the categories' prior probabilities, i.e., Bayes' rule; Rehder & Kim, 2009a).

[6] Specifically, coherence explained >99% and 98% of the variance in Experiment 1's Chain-100 and Chain-75 conditions, 96% and >99% in Experiment 2's Background-0 and Background-50 conditions, 38% and 91% in Experiment 3's Essentialized-Chain-80 and Unconnected-Chain-80 conditions, and >99% in both conditions of Experiment 4. The lower percentage in the Experiment 3's Essentialized-Chain-80 occurred because of the especially large weight placed on feature X in that condition.

We close with a methodological point. This study is the first to provide detailed model fits of classification ratings as a function of categories' causal models whose parameters (i.e., causal strengths) have been specified. However, because there are well-known uncertainties regarding how people use rating scales (Poulton, 1989), more sensitive theoretical tests will require more sensitive measures. For example, to avoid ratings scales, Rehder and Kim (2009b) presented participants with two-alternative forced choice trials. Furthermore, Hayes and Rehder (reported in Rehder, 2010) used 2AFC and then analyzed choices using *logistic* regression to extract the resulting feature weights and interactions.

Summary

In this article, we have presented several main findings. First, stronger causal relations resulted in a smaller causal status effect and a larger coherence effect. Second, weaker alternative causes increased the size of both effects. Third, an essentialized category resulted in a larger causal status effect, albeit only for probabilistic causal links. Fourth, the causal status effect was mediated by features' subjective category validity. Each of these findings is consistent with a generative model of classification and is inconsistent with a dependency model.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

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., Flanagan, E., Marsh, J. K., & Sanislow, C. (2006). Beliefs about essences and the reality of mental disorders. *Psychological Science, 17,* 759–766.

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

Ahn, W., Levin, S., & Marsh, J. K. (2005). Determinants of feature centrality in clinicians' concepts of mental disorders. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 391–396). Mahwah, NJ: Erlbaum.

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

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

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

Braisby, N., Franks, B., & Hampton, J. (1996). Essentialism, word use, and concepts. *Cognition, 59,* 247–274.

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

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

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

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

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

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the nonobvious. *Cognition, 38,* 213–244.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language, 34,* 686–708.

Hampton, J. A., Estes, Z., & Simmons, S. (2007). Metamorphosis: Essence, appearance, and behavior in the categorization of natural kinds. *Memory & Cognition, 35,* 1785–1800.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of examples. *Journal of Verbal Learning and Verbal Behavior, 16,* 321–338.

Johnson, S. C., & Solomon, G. E. A. (1997). Why dogs have puppies and cats have kittens: The role of birth in young children's understanding of biological origins. *Child Development, 68,* 404–419.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York, NY: Cambridge University Press.

Kalish, C. W. (1995). Essentialism and graded category membership in animal and artifact categories. *Memory & Cognition, 23,* 335–349.

Kaplan, A. S., & Murphy, G. L. (1999). The acquisition of category structure in unsupervised learning. *Memory & Cognition, 27,* 699–712.

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

Kim, N. S., & Ahn, N. S. (2002). Clinical psychologists' theory-based representation of mental disorders affect their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131,* 451–476.

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.

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review, 107,* 195–212.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115,* 955–984.

Luhmann, C. C., Ahn, W., & Palmeri, T. J. (2006). Theory-based categorization under speeded conditions. *Memory & Cognition, 34,* 1102–1111.

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

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

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

Marsh, J., & Ahn, W. (2006). The role of causal status versus inter-feature links in feature weighting. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 561–566). Mahwah, NJ: Erlbaum.

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

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–196). Cambridge, MA: Cambridge University Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

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

Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 904–919.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberchien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications* (pp. 23–45). Chichester, England: Ellis Horwood.

Oppenheimer, D. M., Tenenbaum, J. B., & Krynski, T. (2009). *Categorization as causal explanation: Discounting and augmenting of concept-irrelevant features in categorization.* Manuscript submitted for publication.

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

Poulton, E. C. (1989). *Bias in quantifying judgments.* Hillsdale, NJ: Erlbaum.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science, 27,* 709–748.

Rehder, B. (2003b). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1141–1159.

Rehder, B. (2007). Essentialism as a generative theory of classification. In A. Gopnik & L. Schultz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 190–207). Oxford, England: Oxford University Press.

Rehder, B. (2009). Causal-based property generalization. *Cognitive Science, 33,* 301–343.

Rehder, B. (2010). Causal-based classification: A review. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 39–116). Amsterdam, the Netherlands: Elsevier.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of object categories. *Cognitive Psychology, 50,* 264–314.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130,* 323–360.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition, 91,* 113–153.

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

Rehder, B., & Kim, S. (2008). The role of coherence in causal-based categorization. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 285–290). Mahwah, NJ: Erlbaum.

Rehder, B., & Kim, S. (2009a). Classification as diagnostic reasoning. *Memory & Cognition, 37,* 715–729.

Rehder, B., & Kim, S. (2009b). *Causal status and coherence in causal-based classification.* Poster presented at the 50th Annual Meeting of the Psychonomic Society, Washington, DC.

Rehder, B., & Milovanovic, G. (2007). Bias toward sufficiency and completeness in causal explanations. In D. MacNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (p. 1843). Mahwah, NJ: Erlbaum.

Rehder, B., & Murphy, G. L. (2003). A Knowledge-Resonance (KRES) model of category learning. *Psychonomic Bulletin & Review, 10,* 759–784.

Rehder, B., & Ross, B. H. (2001). Abstract coherent concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1261–1275.

Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology, 4,* 194–206.

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

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

Sloman, S. A., & Lagnado, D. A. (2005). Do we "do"? *Cognitive Science, 29,* 5–39.

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

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22,* 377–386.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 216–227.

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

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

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology, 18,* 158–194.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 449–468.

(*Appendices follow*)

Appendix A

Analysis of Individual Differences

A notable result in Experiments 1–4 is the presence of clustering in the classification ratings reflecting subgroups of participants that responded in qualitatively different ways. The presence of subgroups was revealed by submitting the ratings for the eight test items produced by each participant in each experimental condition to a cluster analysis program (SAS PROC CLUSTER METHOD = EML). In Experiment 1, this procedure identified four subgroups in the Chain-100 condition and three subgroups in the Chain-75 condition. The classification ratings of these groups are presented in Panels A and B of Figure A1, collapsed according to the number of characteristic features present in each test item (0, 1, 2, or 3). Each panel includes performance in Experiment 1's Control condition for comparison.

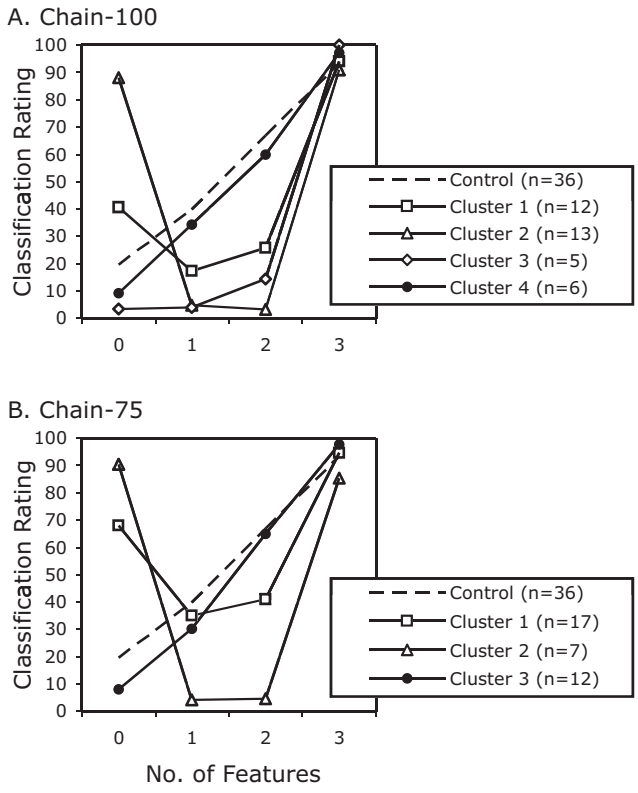

Figure A1. Individual differences in Experiment 1: (A) Chain-100 condition and (B) Chain-75 condition. Performance of control group is included in each panel for comparison.

The subgroups have straightforward interpretations. In the Chain-100 condition (see Panel A of Figure A1), one subgroup of 13 participants produced very high ratings for test items with zero or three features and very low ones for those with one or two features. These participants are exhibiting a large sensitivity to coherence (as indicated by the low ratings for all test items that violate any causal law) and are ignoring the information about which features are characteristic (as indicated by Test Items 111 and 000 receiving the same ratings). A subgroup of six participants produced ratings that increased linearly with the number of characteristic features. These participants are exhibiting no sensitivity to coherence; indeed, the fact that these participants were indistinguishable from the control group suggests that they ignored the causal relationships. Finally, the two remaining subgroups of 12 and 5 participants exhibited sensitivity to both coherence and the number characteristic features. The subgroups in the Chain-75 condition (see Panel B of Figure A1) have a similar interpretation. The subgroups of 7 and 12 participants are basing their ratings solely on coherence and the number of characteristic features, respectively, whereas the majority subgroup of 17 participants is taking both factors into account. Cluster analyses revealed the presence of similar subgroups in Experiments 2–4. Overall, in these experiments, 14% of participants responded on the basis of number of features alone, and 21% responded on the basis of coherence alone.

As mentioned, the presence of the subgroups that ignored the causal relationships contributed to the discrepancies between the generative model's predicted ratings and the observed ones. For example, in Experiment 1's Chain-100 condition, the generative model predicts a likelihood of 0 for two two-feature items (110 and 101) and for two single-feature items (100 and 010), when in fact the former received significantly higher ratings than the latter. However, if the Chain-100 participants who ignored the causal relationships (and gave high ratings to the two-feature items as a result) are removed from the analyses, then the ratings of Items 110, 101, 100, and 010 fall to 14.0, 11.3, 9.5, and 10.3, respectively, and the ratings of the two-features items are no longer significantly higher than the one-feature items. As a result, average root mean squared errors (RMSEs) for the generative model's fits improved (9.1 from 11.4). Removal of subgroups that ignored the causal relationships also led to substantial improvements in other conditions, namely, Experiment 1's Chain-75 condition (average *RMSE*s of 10.3 from 19.0) and Experiment 4's Unconnected-Chain-100 condition (5.3 from 9.3). Note that the effects reported in Experiments 1 and 4 remain unchanged if the participants who ignored the causal links are removed from the analyses.

(*Appendices continue*)

Appendix B

Follow-Ups to Experiment 1

We report the results from a number of follow-up experiments to Experiment 1. Data from these experiments are available in the online supplemental materials.

First, we conducted a version of Experiment 1 in which participants were explicitly told that features Y and Z had no alternative causes (i.e., that $b_Y = b_Z = 0$). (In this replication, we also introduced two minor procedural changes, which were to give participants a diagram of the causal links during the classification and feature likelihood rating tasks and to balance the order of those two tasks over participants.) The essential results from this follow-up experiment are presented in Table B1, which also includes those of Experiment 1 for comparison. The table shows that each of the substantative results from Experiment 1 was replicated in the follow-up. According to all measures (feature regression weights, individual test item ratings, and feature likelihood ratings), the causal status effect was larger in the Chain-75 condition than in the Chain-100 condition, and there was no sign of a causal status effect in the latter condition. Furthermore, coherence effects were larger, and the difference between the direct and indirect interaction terms was smaller, in the Chain-100 condition as compared with the Chain-75 condition. Thus, the key results from Experiment 1 do not depend on differences in the perceived strength of alternative causes between the Chain-100 and Chain-75 conditions.

Second, we conducted a variant of Experiment 1 in which causal strengths were 90% and 60% rather than 100% and 75% (no information about alternative causes was provided, participants were provided a diagram of the causal relationships, and the order of the classification and feature likelihood tasks was balanced). Table B1 confirms that a significant causal status effect was present on all measures in both conditions but was significantly larger for causal strengths of 60% versus 90%. The interaction weights were marginally larger in the 90% condition. (The predicted smaller direct/indirect difference in the 90% condition did not obtain.) Thus, the key predictions of the generative model regarding how causal status and coherence effects vary with causal link strength obtain not only when comparing deterministic versus probabilistic causal links (Experiment 1) but when comparing strong versus weak probabilistic links.

Third, we replicated Ahn, Kim, Lassa line, and Dennis (2000, Study 1) by comparing a condition using the original "tends to" wording with one that used "always" instead (e.g., "Sticky feet *always* allows roobans to build nests on trees."). (In that study, causal links were presented on the computer screen on each classification trial, and participants were sequentially presented with four categories, two with interfeature causal links and two without. One minor procedural change in our replication was that participants first previewed all three test items before rating them.) As in the original study, only three test items were presented—ones missing only feature X, only Y, and only Z. Table B1 confirms that the larger causal status effect found in the probabilistic ("tends to") condition (a difference between the missing-X and missing-Z items of 24.2 points, comparable with Ahn et al., 2000) was reduced to a nonsignificant 5.5 points in the deterministic ("always") condition. Because presenting only three test items prohibits an evaluation of coherence effects, we also replicated these conditions while presenting all eight test items (and then ran Experiment 1's regression analyses on the resulting data). Table B1 indicates that a significant causal status effect obtained in the probabilistic condition but not the deterministic condition according to both the feature regression weights and the ratings of the missing-X and missing-Z test items. Moreover, the interaction weights were larger in the deterministic condition as compared with the probabilistic condition. (The predicted smaller direct/indirect difference in the former condition did not obtain.) Thus, the results from Experiment 1 generalize to other materials and experimental procedures.

The generative and dependency models were fit to the data from these additional experiments just as they were to Experiment 1. The average R^2s and root mean squared errors achieved by these fits are presented in Table B1. In every condition, the fits of the generative model are superior to those of the dependency model.

(*Appendices continue*)

Table B1
Summary of Key Results From Experiment 1 and Four Follow-Up Experiments

	Experiment 1			No alternative causes			Probabilistic links			Ahn et al. (2000) Replication 3 test items			Ahn et al. (2000) Replication 8 test items		
	Chain-100	Chain-75	Difference	Chain-100	Chain-75	Difference	Chain-90	Chain-60	Difference	Deterministic	Probabilistic	Difference	Deterministic	Probabilistic	Difference
m_{XY}	1.0	.75		1.0	.75		.90	.60		"always"	"tends to"		"always"	"tends to"	
b_Y				0	0										
m_{YZ}	1.0	.75		1.0	.75		.90	.60		"always"	"tends to"		"always"	"tends to"	
b_Z				0	0										
Feature weights															
X	6.2	8.6		4.9	11.2		9.3	10.4					12.6	17.2	
Y	7.7	7.3		4.8	8.8		7.9	7.8					13.7	14.0	
Z	6.6	5.1		4.7	3.8		6.3	4.3					13.8	14.0	
X − Z	*−0.4*	*3.5***	*−4.0**	*0.2*	*7.3***	*−7.1***	*3.0**	*6.1***	*−3.2**				*−1.2*	*3.2*	*4.4*†
Test item ratings															
011 [Missing X]	24.4	40.6		12.6	12.6		39.5	38.3		33.4	28.5		35.5	42.5	
110 [Missing Z]	22.3	46.7		13.3	25.7		45.4	40.4		38.9	52.7		32.0	53.4	
110 − 011	*−2.1*	*6.1*	*−8.2*	*0.8*	*12.8***	*−12.0***	*5.9*†	*14.3**	*−8.7**	*5.5*	*24.2***	*18.7***	*−3.5*	*10.9**	*14.4**
Likelihood ratings															
X	77.8	76.7		77.0	77.4		84.5	71.2							
Y	77.3	73.7		76.5	73.4		85.4	60.1							
Z	77.0	70.2		76.6	70.1		79.3	53.7							
X–Z	*0.7*	*6.5***	*−5.8***	*0.4*	*7.3***	*−6.9***	*5.2**	*17.6***	*−12.4***						
Interaction weights															
Direct	13.6	10.3		16.6	12.3		10.0	7.9					7.4	4.6	
Indirect	12.5	8.2		16.6	8.6		7.8	5.9					4.8	3.1	
M(Direct, Indirect)	*13.0***	*9.2***	*3.8***	*16.6***	*10.4***	*4.2***	*8.9***	*6.9***	*2.0*†				*6.1***	*3.8***	*2.3***
Direct − Indirect	*1.1*	*2.1**	*−1.0*	*0*	*3.8***	*−3.8***	*2.2**	*2.0**	*0.2*				*2.6**	*1.5**	*1.1*
Model fits															
Dependency model															
Average R^2	.29	.41		.18	.38		.44	.47					.61	.77	
Average *RMSE*	33.4	25.5		38.4	30.0		25.2	21.2					22.6	16.7	
Generative model															
Average R^2	.87	.74		.92	.70		.75	.77					.77	.89	
Average *RMSE*	11.4	19.0		7.4	20.0		16.9	16.2					17.3	14.0	

Note. Presented for each condition are (a) the feature regression weights, (b) the ratings of test items missing only feature X (011) and only feature Z (110), (c) the feature likelihood ratings, and (d) the interaction regression weights. Three measures of the causal status effect are also presented: the difference between (a) the regression weights on features X and Z, (b) the missing-X and missing-Z test items, and (c) the likelihood ratings for features X and Z. The coherence effect is reflected in the magnitude of the interaction weights. Direct = interaction weight on directly connected feature pairs (X and Y, and Y and Z); Indirect = interaction weight on indirectly connected feature pair (X and Z); RMSE = root mean squared error. Italicized quantities are tested statistically against 0.
† $p < .15$. * $p < .05$. ** $p < .01$.

(*Appendices continue*)

Appendix C

Within-Subject Replications of Experiments 1–3

We report replications of Experiments 1–3 in which the essential manipulations are carried out within-subjects. The same general procedure and materials were used as in Experiments 1–3 except for the additional features and causal links needed to instantiate the causal networks in Figure C1. (These materials are available from the authors.) In each experiment, for half of the participants, two category features were assigned the roles of W and X, and the other two were assigned the roles of Y and Z, and this assignment was reversed for the other half of the participants.

The ensuing classification test presented all 16 items that can be formed on four binary dimensions. Classification ratings were analyzed via multiple-regression with four predictors corresponding to features W, X, Y, and Z and two two-way interaction terms corresponding to the two pairs of causally related features (W and X, and Y and Z). The feature likelihood test presented the two features on each of the four dimensions. A diagram of the causal relations was provided to the participants during the tests, and the presentation order of the tests was balanced over participants. Data from these experiments are available in the online supplemental materials.

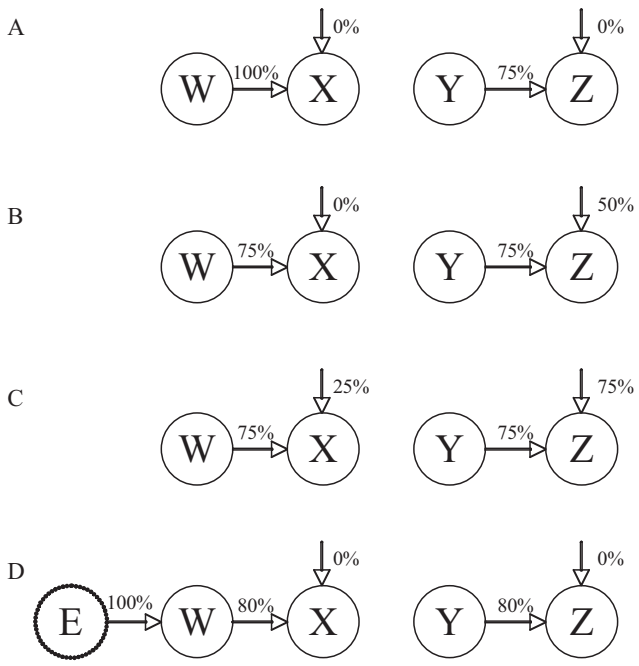

Figure C1. Causal networks taught to participants in within-subject replications of Experiments 1–3.

Within-Subject Replication of Experiment 1

Participants were taught a single category with four features arranged as shown in Panel A of Figure C1. Feature W was described as causing X with probability 100%, and Y was described as causing Z with probability 75%. X and Z were described as having no alternative causes.

Feature regression weights averaged over participants are presented in Table C1 and show that whereas the cause feature in the 75% relation was weighed more than its effect (4.2 vs. −0.2; i.e., a causal status effect), the cause and effect in the 100% relation were weighed about equally (3.7 and 4.0). A 2 × 2 analysis of variance (ANOVA) of these feature weights with causal strength (100% vs. 75%) and role (cause vs. effect) yielded marginal effects of strength, $F(1, 23) = 3.56$, $MSE = 25.1$, $p = .07$; and role, $F(1, 23) = 4.13$, $MSE = 24.6$, $p = .05$; and a significant interaction between the two, $F(1, 23) = 4.35$, $MSE = 30.3$, $p < .05$. Pairwise comparisons confirmed that the weights of features Y and Z differed, $t(23) = 2.65$, $p < .05$, but those of W and X did not ($t < 1$). This effect on feature classification weights can be observed in the ratings of individual test items (see Table C1). Whereas the item missing only feature Y (1101) was rated lower than the one missing only Z (1110), the items missing W (0111) and X (1011) were rated about equally.

Explicit feature likelihood ratings exhibited the same pattern as the classification ratings. Whereas features Z was rated as less prevalent than feature Y (66.5 vs. 75.8), W and X were rated as equally prevalent (78.9 and 79.1). A 2 × 2 ANOVA of these feature weights with causal strength (100% vs. 75%) and role (cause vs. effect) yielded main effects of strength, $F(1, 23) = 23.28$, $MSE = 63.9$, $p < .0001$; role, $F(1, 23) = 13.99$, $MSE = 35.4$, $p < .01$; and an interaction between the two, $F(1, 23) = 11.40$, $MSE = 47.5$, $p < .01$. Pairwise comparisons confirmed that the likelihood ratings for features Y and Z differed, $t(23) = 3.59$, $p < .01$, but those of W and X did not ($t < 1$).

Finally, feature interaction weights presented in Table C1 show the same effect observed in Experiment 1, namely, a larger interaction weight for the 100% relation (22.6) as compared with the 75% one (10.9), $t(23) = 6.16$, $p < .0001$.

Within-Subject Replication 1 of Experiment 2

Participants were taught a category with features arranged as shown in Panel B of Figure C1. Feature W was described as causing X, and Y was described as causing Z, with probability 75%. Whereas X was described as having no alternative causes, Z was described as appearing in 50% of category members in which Y was absent.

(*Appendices continue*)

Table C1
Summary of Key Results From Within-Subject Variants of Experiments 1–3 Described in Appendix C

	Experiment 1 Replication	Experiment 2 Replication 1	Experiment 2 Replication 2	Experiment 3 Replication
m_{EW}				1.0
m_{WX}	1	.75	.75	.80
b_X	0	0	.25	0
m_{YZ}	.75	.75	.75	.80
b_Z	0	.50	.75	0
Feature weights				
W	3.7	8.1	5.7	17.6
X	4.0	3.0	5.5	5.2
Y	4.2	6.0	5.0	9.8
Z	−0.2	7.2	9.9	3.4
W − X	*−0.2*	*5.1**	*0.2*	*12.4***
Y − Z	*4.4**	*−1.3*	*−4.9***	*6.5***
Test item ratings				
0111 [Missing W]	15.0	32.7	54.8	21.1
1011 [Missing X]	20.2	48.8	58.5	65.2
1101 [Missing Y]	31.9	50.2	64.1	36.9
1110 [Missing Z]	62.3	51.0	54.5	64.4
1011–0111	*−5.2*	*−16.0**	*−3.8*	*44.2***
1110–1101	*−30.4***	*−0.8*	*9.6**	*27.5***
Likelihood ratings				
W	78.9	75.2	73.9	99.2
X	79.1	69.4	72.6	81.9
Y	75.8	76.8	73.8	85.7
Z	66.5	78.5	79.8	74.3
W − X	*−0.2*	*5.8**	*1.3*	*17.3***
Y − Z	*9.3***	*−1.6*	*−6.0***	*11.5***
Interaction weights				
WX	*22.6***	*13.5***	*9.6***	*6.2***
YZ	*10.9***	*8.3***	*6.5***	*6.1***
WX − YZ	*11.7***	*5.2***	*3.1**	*0.1*

Note. Presented for each condition are (a) the feature regression weights, (b) the ratings of test items missing only a single feature (W, X, Y, or Z), (c) the feature likelihood ratings, and (d) the interaction regression weights. Three measures of the causal status effect between features W and X and between features Y and Z are presented: the difference between (a) the feature regression weights, (b) the test items missing a single feature, and (c) feature likelihood ratings. The coherence effect is reflected in the magnitude of the interaction weights. Italicized quantities are tested statistically against 0.
$^{\dagger} p < .15$. $^{*} p < .05$. $^{**} p < .01$.

Classification regression weights are presented in Table C1 and show that whereas the cause feature was weighed more than its effect (i.e., a causal status effect) in the absence of background causes (8.1 vs. 3.0), the cause and effect features were weighed about equally for background strengths of 50% (6.0 vs. 7.2). A 2 × 2 ANOVA of these feature weights with background strength (0% vs. 50%) and role (cause vs. effect) yielded main effects of neither background strength nor role ($ps > .15$) but a significant interaction between the two, $F(1, 23) = 5.44$, $MSE = 44.1$, $p < .05$. Pairwise comparisons confirmed that the weights of features W and X differed, $t(23) = 2.33$, $p < .05$, but those of Y and Z did not ($t < 1$). This effect on feature classification weights can be observed in the ratings of individual test items shown in Table C1. Whereas the item missing only feature W (0111) was rated lower than the one missing only X (1011), the items missing Y (1101) and Z (1110) were statistically equal.

Explicit feature likelihood ratings exhibited the same pattern as the classification ratings. Whereas feature X was rated as less prevalent than its cause W (69.4 vs. 75.2), Y and Z were rated about equally (76.8 and 78.5). A 2 × 2 ANOVA of these ratings with background strength (0% vs. 50%) and role (cause vs. effect) yielded a main effect of strength, $F(1, 23) = 12.07$, $MSE = 57.0$, $p < .01$; no effect of role, $F(1, 23) = 1.64$, $MSE = 64.9$, $p < .20$; and an interaction between the two, $F(1, 23) = 8.39$, $MSE = 38.8$, $p < .01$. Pairwise comparisons showed that the likelihood ratings for features W and X differed, $t(23) = 2.34$, $p < .01$, but those of Y and Z did not ($t < 1$).

Finally, feature interaction weights showed the same pattern observed in Experiment 2: a larger interaction weight on the causal link with the weaker alternative cause (13.5 vs. 8.3), $t(23) = 2.86$, $p < .01$.

Within-Subject Replication 2 of Experiment 2

Participants were taught a category with the same causal information as in Panel B of Figure C1 except that the strengths of

(*Appendices continue*)

alternative causes for features X and Z were 25% and 75% (see Panel C of Figure C1). Classification regression weights presented in Table C1 show that whereas the cause and effect features were weighed about equally for background strengths of 25% (5.7 and 5.5), the effect was weighed more than its cause for background strengths of 75% (9.9 vs. 5.0)—that is, a *negative* causal status effect obtained. A 2 × 2 ANOVA of these feature weights with background strength (25% vs. 75%) and role (cause vs. effect) yielded main effects of background strength, $F(1, 47) = 5.55$, $MSE = 29.2$, $p < .05$; of role, $F(1, 47) = 8.37$, $MSE = 32.5$, $p < .01$; and a significant interaction, $F(1, 23) = 9.18$, $MSE = 34.0$, $p < .01$. Pairwise comparisons revealed that the weights of features Y and Z differed, $t(47) = 3.58$, $p < .001$, but those of W and X did not ($t < 1$). This effect on feature classification weights can be observed in the ratings of individual test items shown in Table C1. Whereas the item missing only feature Y (1101) was rated lower than the one missing only Z (1110), the missing W (0111) and X (1011) items were statistically equal.

Explicit feature likelihood ratings exhibited the same pattern as the classification ratings. Whereas feature Y was rated as less prevalent than feature Z (73.8 vs. 79.8), W and X were rated about equally (73.9 and 72.6). A 2 × 2 ANOVA with background strength (25% vs. 75%) and role (cause vs. effect) yielded a main effect of strength, $F(1, 47) = 4.98$, $MSE = 120.1$, $p < .05$; no effect of role, $F(1, 47) = 1.77$, $MSE = 151.7$, $p = .19$; and an interaction, $F(1, 47) = 4.83$, $MSE = 134.2$, $p < .05$. Pairwise comparisons showed that the likelihood ratings for features Y and Z differed, $t(47) = 2.73$ $p < .01$, but those of W and X did not ($t < 1$).

As in Experiment 3, the causal link with the weaker alternative cause had the larger interaction weight (9.6 vs. 6.5), $t(23) = 2.46$, $p < .05$.

Within-Subject Replication of Experiment 3

Participants were taught a category with features arranged as shown in Panel D of Figure C1. As in Experiment 3, feature E was a pseudo-essential feature. E was described as causing W with probability 100%. In addition, feature W was described as causing X, and Y was described as causing Z, with probability 80%. X and Z were described as having no alternative causes.

Feature regression weights averaged over participants are presented in Table C1 and show that a larger causal status effect obtained between features W and X (feature weight difference of 17.6 – 5.2 = 12.4) than between Y and Z (9.8 – 3.4 = 6.5). A 2 × 2 ANOVA of these feature weights with link (essentialized vs. not) and role (cause vs. effect) yielded main effects of link, $F(1, 23) = 11.08$, $MSE = 49.6$, $p < .01$; role, $F(1, 23) = 31.67$, $MSE = 67.5$, $p < .0001$; and a significant interaction, $F(1, 23) = 5.83$, $MSE = 35.9$, $p < .05$. This effect on feature classification weights can be observed in the ratings of individual test items shown in Table C1. The difference in classification between missing-W item (0111) and missing-X items (1011 and 44.2) was larger then the difference between missing-Y (1101) and missing-Z (1110 and 27.5) items.

Explicit feature likelihood ratings (see Table C1) exhibited the same pattern. The difference in likelihood ratings between features W and X (99.2 – 81.9 = 17.3) was greater than the difference between Y and Z (85.7 – 74.3 = 11.5). A 2 × 2 ANOVA of these feature ratings yielded main effects of link type, $F(1, 23) = 53.46$, $MSE = 49.9$, $p < .0001$; role, $F(1, 23) = 145.84$, $MSE = 34.0$, $p < .0001$; and an interaction, $F(1, 23) = 12.12$, $MSE = 16.8$, $p < .01$. These results confirm the conclusion from Experiment 3 that the presence of an essential feature results in a larger causal status effect.

Finally, feature interaction weights are presented in Table C1 and show that both interaction weights (6.2 and 6.1) are significantly greater than 0.

Received February 14, 2008
Revision received January 21, 2010
Accepted March 24, 2010 ■