Causal Status and Coherence in Causal-Based Categorization

Bob Rehder

ShinWoo Kim

Department of Psychology

New York University

Send all correspondence to:
Bob Rehder
Department of Psychology
6 Washington Place
New York, NY 10003
Phone: (212) 992-9586
Email: bob.rehder@nyu.edu

Abstract

Two models of how interfeature causal knowledge affects classification were evaluated. First, a *causal status effect* was stronger for probabilistic versus deterministic causal links (Expt. 1). Second, a causal status effect was stronger for essentialized categories (Expt. 2) although only for probabilistic causal links (Expt. 3). Third, a causal status effect was absent when information about probabilistic causal links and essential features was omitted (Expts. 4-6). Fourth, the causal status effect was mediated by features' *category validity*, the probability with which they occur in category members. Fifth, Experiments 1-6 exhibited robust *coherence effects* that accounted for an order of magnitude more variance than the causal status effect, indicating that coherence is the most important effect in causal-based categorization. These findings were consistent with a generative model of categorization but inconsistent with an alternative model.

Causal Status and Coherence in Causal-Based Categorization

An important goal in the psychology of concepts is to identify how both empirical information—information that we 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 every type of category-based judgment that has been tested, including learning, induction, and categorization (see Murphy, 2002, for a review). This article is 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, 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; Sloman, Love, & Ahn, 1998; Kim & Ahn, 2002). But in order 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.

Researchers have tested a number of different causal network topologies in this manner, and one network that has proven to be of particular theoretical importance is the three-feature chain in Figure 1A 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 and the robustness of the causal status and coherence effects.

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 1A 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 salience (Sloman et al. 1998) and their *cue validity* (the extent to which they are diagnostic of that category versus another, Rosch & Mervis, 1975). But the claim is that when these factors are controlled, causal status dominates.

Figure 2 presents two empirical studies that provide at least partial support for the causal status effect. In Ahn et al. (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 asked to rate on a 0 to 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 causally related in 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 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 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 the Ahn et al (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 features, Rehder, 2003b, Experiment 1).

These results notwithstanding, a number of 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 secondary aspects of the experimental procedure. For example, Rehder and Kim found evidence that the causal status effect depends on the order in which features are presented on the computer screen, suggesting that the large causal status effect in Ahn et al. may have arisen in part because they presented features in causal order (X first, then Y, then Z). In addition, the studies differ on the number of test items presented on the classification test. Experiments that follow will assess these and a number of other hypotheses regarding what factors promote a causal status effect.

The Coherence Effect

Although the two studies just reviewed provide some support for a causal status effect, one difference between them concerns how feature weights were assessed. Whereas Ahn et al. asked participants to rate the category membership of only three test exemplars (missing-X, missing-Y, and missing-Z), Rehder and Kim 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 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 4.4 indicates, for example, that categorization ratings were about 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 to the direct terms reflects the fact that X and Z should also be correlated, albeit not as strongly due to the indirect relation between them (more about this below). This sensitivity to expected correlations has been found in numerous studies (Rehder, 2003b; Marsh & Ahn, 2006), including those testing other causal network topologies (Rehder & Hastie, 2001; Rehder 2003a; Rehder & Kim, 2006).

As was the case for the causal status effect, questions have been raised regarding to what extent the coherence effect might depend on secondary experimental factors. For example, Marsh and Ahn (2006) noted that the materials used in Rehder's series of experiments usually used "normal" as the atypical dimension value. 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 argued that this use of "normal" values might have inflated coherence effects because participants might expect all the normal values to appear together. To address this concern, the experiments that follow test 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.

Two Theoretical Models

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. Specifically, according to the dependency model, feature *i*'s weight or centrality, $c_{i,}$, can be computed from the iterative equation,

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

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. This equation converges in a finite number of iterations.

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, the weights for X, Y, and Z converge (in two iterations) to 4, 2, and 1. 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 test another prediction of the dependency model, namely, that the size of the causal status effect should increase monotonically with causal strength (Figure 3A). For example, although feature weights are 4, 2, and 1 when $d = 2$ (yielding a difference of 3 between the weights of X and Z), they are 9, 3, and 1 when $d = 3$ (a difference of 8). Our experiments manipulate the strength of the causal links in order to assess how the causal status effect varies as a function of those strengths.

The Generative Model

A second model that has been offered as an account of causal knowledge and classification is known as the *generative model*. Building on *causal-model theory* (Waldmann & Holyoak, 1992; Sloman, 2005), 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, whereas those unlikely to be generated are poor category members.

One important advantage of the generative model is that it is unique in being able to account for the coherence effect. The generative model predicts that a population of category members generated by a causal network should exhibit the expected pattern of correlations between causally related features and thus a likely category member is one that maintains those correlations. Of course, two features will be correlated (will usually be both present or both absent) when they are directly linked by a causal relation.

In addition, the generative model predicts that category members should also exhibit more subtle patterns of correlations when three or more variables are causally related.

The right panel of Figure 3B presents the generative model's predictions regarding the expected correlations between feature pairs for a three-feature causal chain as a function of causal link strength *m*. (We distinguish *m* from the dependency model's parameter *d* because whereas *d* is unbounded, *m* is a probability and thus is constrained to be < 1.) First, it predicts that the correlation between those pairs that are directly connected (XY and YZ) should be positive for all values of $m > 0$. Second, it predicts that the magnitude of the direct correlations should increase monotonically with causal strength, reflecting the fact that stronger causal relationships imply stronger correlations. Third, it predicts that the indirectly connected pair of features (XZ) should also be positively correlated and that correlation should increase monotonically with causal strength. Fourth, when the causal relationships are deterministic ($m = 1$), X, Y, and Z should all be perfectly correlated, and thus the direct and indirect terms should have the same magnitude. But when the causal links are probabilistic ($0 < m < 1$), the indirect correlation should be smaller than the direct ones, reflecting the weaker correlation between variables that are only indirectly causally related. The generative model thus explains the large weights on the direct interaction terms and substantial but smaller weights on the indirect term found in Rehder and Kim (2006) (Figure 2B).

In addition to predicting the coherence effect, the generative model can also explain the causal status effect, although it requires an additional assumption to do so. Rehder (2003b) has shown that, depending on the parameters of the category's causal model, the generative model can predict a number of different patterns of feature weights: decreasing feature weights along a causal chain (a causal status effect), increasing weights, or equal weights. Thus, to explain the causal status effect, Rehder proposed that people often assume that a category's observable features are generated by a deeper but hidden cause associated with the category. For example, one of the novel categories used in Ahn et al. (2000) was a novel disease which was described as causing a symptom X which caused another symptom Y which caused a third symptom Z. But although Ahn et al. assumed that participants' causal model consisted of X → Y → Z, people understand that a disease causes its symptoms, so participants were likely to have assumed the more complex causal model in Figure 1B in which E (the disease) is the essential feature that is the ultimate cause of the symptoms. Under these circumstances, participants were likely to have reasoned backwards from the symptoms to the disease, and, of course, symptoms closer (in a causal sense) to the disease (e.g., X) were taken to be more diagnostic of that disease.

Participants may have also assumed a complex causal model like the one in Figure 1B for the artificial natural kinds and artifacts tested in many studies. 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; Medin & Ortony, 1989). Moreover, essential features 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 be the intentions of the their designers (Bloom, 1998; Keil, 1995; Matan & Carey, 2001; Rips, 1989). Thus, the causal model that people use during categorization may include the background causes that categorizers presume bring rise to a category's observable features.

The left panel of Figure 3B presents the probability that each feature in the causal chain will be generated assuming that the chain is caused by an underlying essential feature (Figure 4) and that there are no other unspecified causes. In fact, those probabilities decrease from X to Y to Z for all values of $m < 1$ and > 0. For example, when all three causal links have a strength of $m = .75$ (and there are no other potential causes of the features), X will appear in category members with probability .75 (because it is generated by E with probability .75 and E is present in all category members). Y in turn will appear with probability ($.75 \times .75 =$) .56, and Z should appear with probability ($.75^3 =$) .42. Thus, because judgments of category membership are known to vary with features' probability given the category (that is, their *category validity*), the generative model predicts a causal status effect whenever causal links are probabilistic. In contrast, there should be no causal status effect when causal links are deterministic ($m = 1$), because each feature is generated with equal probability (namely, 1.0), or when they are absent ($m = 0$). This nonmonotonic increase in the size of the causal status effect distinguishes the generative model from the dependency model that predicts that the size of the effect should increase monotonically with causal strength.

Our experiments will test the generative model's predictions that a causal status effect depends on probabilistic causal links and an essentialized category. They will also test the generative model's prediction that the causal status effect is mediated by the perceived probability with which features appear in category members by asking participants to make explicit feature frequency judgments. On this account, whenever a full causal status effect appears, participants should believe that X appears more frequently in category members than feature Y and that Y appears more frequently than Z.

Overview of Experiments

This article has two parts. The purpose of the first part is to present new tests of the two models just described. Experiment 1 begins by testing the contrasting predictions of the dependency model and the generative model regarding how the causal status effect varies as a function of causal link strength. Experiment 2 then assesses how that effect is mediated by the presence of an underlying category essence. Experiment 3 tests how the effect of essential features interacts with causal link strength. In addition to categorization ratings, all three experiments gather explicit feature frequency ratings. All three experiments test the generative model's predictions regarding the coherence effect.

Having identified, we believe, the primary theoretical variables that are responsible for the causal status effect, in the second part of this article we turn to the question of what secondary experimental factors are also likely to contribute to that effect. In Experiments 4-7 we test the hypotheses that the effect depends on (a) whether features are presented in causal order, (b) whether the atypical dimensions values are described as "normal," (c) whether detailed information about causal mechanism is provided, and (d) the number of items presented during the classification test.

PART 1: TESTING ALTERNATIVE MODELS

Experiment 1

Experiment 1 tests the effect of causal link strength on the causal status and coherence effects. To this end, participants were assigned 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. However, unlike previous studies, participants were given explicit information about the strengths of those causal links. For example, in addition to information about the causal mechanism involved (Table 1) 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 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.

The dependency model and the generative model make distinct predictions for this experiment. Recall that the dependency model predicts that the size of the causal status effect is a monotonic function of causal strength (Figure 3A). Thus, it predicts a stronger causal status effect in the Chain-100 condition versus the Chain-75 condition. In contrast, the generative model predicts that a causal status effect should be present at intermediate values of causal strength and absent when strengths are 100% (Figure 3B). Thus, it predicts a stronger causal status effect in the Chain-75 condition versus the Chain-100 condition. Both models predict the absence of a causal status effect in the Control condition.

Experiment 1 augments the regression analyses of classification ratings with an additional dependent variable by asking participants to provide a frequency rating for each feature. If features indeed exhibit a stronger causal status effect in the Chain-75 condition, the generative model claims this is the case because X is viewed as more common among category members than Y which in turn is more common than Z. Thus, any differences in feature importance detected in classification ratings should be reflected in the feature frequency ratings.

The generative model also predicts how the coherence effect should be affected by the causal strength manipulation. First, the size of that effect (as measured by the magnitude of the interaction weights) should increase with strength. Thus, it should be larger in the Chain-100 condition versus the Chain-75 condition (and be zero in the Control condition). Second, the difference between the direct and indirect terms should be maximal at intermediate values of causal strength and zero when links are deterministic or absent. Thus, that difference should be largest in the Chain-75 condition and zero in the Chain-100 and Control conditions. The dependency model does not predict coherence effects.

Recall that questions have been raised regarding how both the causal status effect and the coherence effect might depend on secondary procedural variables. First, because of evidence that a causal status effect might depend on the features being presented in causal order (Rehder & Kim, 2006), Experiment 1 randomized the presentation order of features over participants (although it was fixed for a given participant). Second, because of the possibility that the coherence effect depends on the uncharacteristic dimension values being described as "normal" (Marsh & Ahn, 2006), Experiment 1 used two opposing values on each dimension (e.g., most Myastars have high density whereas some have *low* density). Thus, any coherence effects that obtain will be free of any potential problem introduced by the perhaps unusual "normal" wording.

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. One value on each dimension was described as characteristic or typical of the category by stating that it occurred in "most" category members. The other value (an opposing value rather than "normal") was described as atypical by stating that it occurred in only "some" category members.

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 assigned in equal numbers to the Chain-100, Chain-75, and Control conditions and to one of the six categories.

Procedure. Experimental sessions were conducted by computer. Participants first studied several screens of information about the category at their own pace and then performed a classification test. The initial screens presented a cover story and which features occurred in "most" versus "some" category members. Chain-100 and Chain-75 participants were then also instructed on the two causal relationships that formed a chain network. These participants also observed a diagram like 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 proceed to subsequent phases was to take it all the way through without errors and without returning to the initial information screens for help.

During the classification test, participants rated the category membership of all possible 8 exemplars that can be formed from three binary features, each presented twice. The order in which the features were listed was randomized across participants (but was consistent from trial to trial for a particular subject). Responses were entered by positioning a slider on a scale where the left end was labeled "Definitely not an X" and the right end was labeled "Definitely an X," where X was the name of the category. 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 frequency rating task that followed the classification test, each of the two features on the three observable binary dimensions were presented alone on the computer screen and participants rated what proportion of all category members possessed that feature. Each of the 6 features was presented twice, for a total of 12 frequency rating trials. The order of these trials was randomized for each participant. Experimental sessions lasted approximately 40 minutes.

Results

The average category membership ratings given to the 8 test items in each condition are presented in Table 2. To determine the effect of causal network on the importance of features and the interactions between features, those ratings were analyzed by performing a multiple regression for each participant. Three predictor variables (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 predictor variables 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 network 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 a causal status effect only for intermediate causal strengths. 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 regression weights statistically, a 3 x 3 mixed ANOVA of those weights was conducted where the between-subject 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, $F(2, 105) = 3.79$, $MSE = 25.3$, $p < .05$. A separate analysis involving only the Chain-75 and Control groups confirmed that the negative linear trend was significantly greater in the Chain-75 condition, indicating the presence of a causal status effect, $F(1, 70) = 5.81$, $MSE = 25.3$, $p < .05$. In contrast, the linear trend did not differ between the Chain-100 and Control conditions, $F < 1$. (In this second analysis the small quadratic effect in the Chain-100 condition suggested by Figure 5A did not reach significance, $p > .15$).

Note that the 3 X 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 compared to the Control condition. We will defer discussion of this finding until after presenting the remaining results.

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 a coherence effect. Specifically it 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 maximal for intermediate 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 to 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 x 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 much 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 to 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 frequency ratings. Feature frequency 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 (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 (Figure 6A) frequency 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%, 76.7%). A 3 x 3 ANOVA of the feature frequency 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, $F < 1$.

Selected test items. Finally, 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, only Y, and only 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. Note, however, that despite the fact that both the regression weights and feature frequency ratings showed that feature X was also more important than Y, the missing-X item was rated *higher* than the missing-Y item (101). This result reflects the fact that in the Chain-75 condition the direct interactions weights were larger than the indirect weights. The missing-X item is given a higher rating because it violates one direct and one indirect interaction whereas the missing-Y item violates two direct interactions.

Note that although a causal status effect obtained in the Chain-75 condition, the pattern of ratings in Figure 7A differs both quantitatively and qualitatively from that found by Ahn et. al. (2000) (Figure 2A). First, Ahn et al. found a difference between the missing-X and missing-Z items that was much larger (35 points on a 100-point scale) than in the current Chain-75 condition (6 points). Second, whereas Ahn et al. found that the missing-Y item was rated higher than the missing-X item, this preference was reversed in Experiment 1. We will address these differences in subsequent experiments.

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 2 or 1 features are rated lower than those with 3 or 0 (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). But 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 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. In contrast, items that are incoherent because they have 1 or 2 characteristic features (and thus have a mixture of "most" and "some" values) are rated 29 points lower than in the Control 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 absent when strengths are 100%. In contrast, the dependency model predicts that greater causal strengths should always lead to larger causal status effect. In fact, a significant linear effect of feature weights for causal strengths of 75% but not 100% indicates that the causal status effect is a nonmonotonic function of causal strength, a result consistent with the generative model but not the dependency model. This result was corroborated by feature frequency ratings that exhibited the same pattern.

Note that a causal status effect arose in the Chain-75 condition despite the fact that features were not always presented on the computer screen in causal order. Although Rehder and Kim (2006) found tentative evidence that a causal status effect might be mediated by presentation order, the results of Experiment 1 indicate that causal presentation order is not a necessary condition for a causal status effect. Experiment 4 will directly test for an effect of presentation order on the causal status effect.

One unexpected result from Experiment 1 was that feature weights in both causal conditions were lower than those in the Control condition. However, we interpret this result as arising 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 coherence effect found in Experiment 1 was another prediction of the generative model of course. Recall that Marsh and Ahn (2006) raised the possibility that the coherence effect may be primarily due to the unusual wording of feature values used in previous research (in which the uncharacteristic feature value was described as "normal"). In fact, the large and positive weights on both the direct and indirect interaction terms in both causal conditions revealed a large coherence effect despite the absence of the "normal" wording. This effect was manifested in the causal conditions in the much lower category membership ratings of the incoherent test items with 1 or 2 characteristic features and the much higher ratings of the coherent item with 0 characteristic features.

The generative model also successfully predicted many of the more subtle aspects of the interaction terms. It predicted that the magnitude of interaction terms should be larger in the Chain-100 condition than in the Chain-75, a difference that in fact obtained. And, it also correctly predicted that the direct interaction terms should be larger than the indirect term in the Chain-75 condition only.

Finally, it is illuminating to assess the relative importance of the causal status and coherence effects in this experiment 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 the full regression model with separate predictors for each feature and each two-way interaction as compared to 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 would have been a simple function of the number of characteristic features displayed by an exemplar (as they in fact were in Experiment 1's Control condition). 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. For the Chain-75 group (the only condition in which a causal status effect obtained), the coherence effect accounted for 97.8% of the variance explained by the full model as compared to 2.2% for the causal status effect. In the Chain-100 condition, the coherence effect accounted for 99.8% of the variance (an unsurprising result given the absence of a causal status effect in that condition). In other words, even when a causal status effect was present, it accounted for more than an order of magnitude less of the variance than the coherence effect. This was true despite the absence of the "normal" wording that Marsh and Ahn (2006) have suggested may be primarily responsible for the coherence effect.

Experiment 2

Recall that, according to the generative model, the two conditions that promote a causal status effect are probabilistic causal links and an essentialized category. Whereas Experiment 1 tested probabilistic causal links, Experiment 2 tests the importance of the category being essentialized by comparing the causal structures shown in Figure 8. As in Experiment 1, each category consisted of three observable features. However, the categories were now "essentialized" by endowing them with a 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[i]. Note that the essential property was never directly observed during the subsequent classification test. That is, participants were told which values each test item had on the same three binary dimensions as in Experiment 1 but not whether it possessed the essential feature.

The three conditions in Figure 8 differ according to the causal knowledge participants learned. In the Essentialized-Chain-80 (Figure 8A) and Unconnected-Chain-80 conditions (Figure 8B), participants were instructed on two causal links forming a causal chain between features X, Y, and Z. In the Essentialized-Chain-80 condition they were also instructed on a third causal relationship linking feature X to the essential feature. For example, in that condition participants who learned Myastars were told that "Ionized helium causes the star to be very hot. Ionized helium participates in nuclear reactions that release more energy than the nuclear reactions of normal hydrogen-based stars, and the star is hotter as a result." Importantly, all causal links were presented as probabilistic by describing them as possessing a strength of 80% (e.g., "Whenever a Myastar has ionized helium, it will cause that star to have a hot temperature with probability 80%.")[ii]. The Control condition served as a comparison group in which no causal links are provided.

According to the generative model, a causal status effect is promoted when observable features are thought to be caused by an underlying category essence and thus it predicts a stronger causal effect in the Essentialized-Chain-80 condition as compared to the Unconnected-Chain-80 condition. Of course, it is not our claim that participants don't sometimes assume the presence of a causally potent essence on their own, that is, even when one is not presented as part of the explicit category description. Indeed, we think that this is just what they did in the previous demonstrations of a causal status effect reviewed earlier (and in Experiment 1's Chain-75 condition). Nevertheless, this assumption may not be made by all participants (or any participant with 100% confidence), and thus providing an explicit essence should magnify the size of the causal status effect. In addition to a causal status effect, the generative model of course also predicts coherence effects in both the Essentialized-Chain-80 and Unconnected-Chain-80 conditions.

The dependency model makes a very 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.

As a consequence, providing a feature with an additional cause (e.g., in the form of an essence) should have no influence on that feature's centrality. For example, when $d = 1$ the dependency model predicts feature centralities of 3, 2, and 1 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 differs from the generative model in not predicting a coherence effect.

Method

Materials. The materials were the same as in Experiment 1, with the exception of the essentialized feature for each category and the additional causal relationship linking it and feature X.

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

Procedure. The procedure for the first part of the experiment was the same as in Experiment 1, except that (a) participants were also instructed on the essentialized feature (and tested on that feature during the multi-choice test) and (b) causal strengths were described as 80% in the two causal conditions. Because the essentialized feature was never presented during the classification test that followed, that test was identical to the one in Experiment 1. After the classification test, participants performed a feature frequency rating task that was identical to the one in Experiment 1.

Results

The average category membership ratings given to each of the 8 test items in each condition are presented in Table 3. Once again, those ratings were analyzed by performing a multiple regression for each participant. As in Experiment 1, there were no effects of which category participants learned on any dependent variable and thus the results are presented collapsed over this factor. We first present the regressions analyses and then the feature frequency ratings.

Feature weights. The regression weights averaged over participants for features X, Y, and Z are presented in the left-hand panels of Figure 9. Recall that the generative model predicts that features should be more diagnostic and exhibit a stronger causal status effect when they are causally related to an underlying category essence. Figure 9 indicates that both of those predictions were confirmed. First, feature weights in the Essentialized-Chain-80 condition (12.8, 10.6, and 8.3 for X, Y, and Z, respectively) were larger than those in the Unconnected-Chain-80 condition (7.3, 6.6, and 5.8). Second, the causal status effect was larger in the Essentialized-Chain-80 condition (a difference between the weights of features X and Z of 4.5) as compared to the Unconnected-Chain-80 condition (a difference of 1.5).

A 3 x 3 ANOVA of the feature weights was conducted with condition (Essentialized-Chain-80, Unconnected-Chain-80, Control) and feature (X, Y, Z) as factors. Following Experiment 1, our analyses focus on the linear effect of features, the appropriate test of a causal status effect. There was an overall effect of condition on the linear trend, $F(2, 105) = 4.52$, $MSE = 29.5$, $p < .05$. In a separate analysis, the linear trend was significantly different in the Essentialized-Chain-80 condition as compared to the Control condition, $F(1, 105) = 9.02$, $MSE = 265.8$, $p < .005$. In contrast, the linear trend in the Unconnected-Chain-80 condition did not differ significantly from the Control condition, $F(1, 105) = 1.86$, $MSE = 29.5$, $p = .18$. A direct comparison of the linear effect in the Essentialized-Chain-80 and Unconnected-Chain-80 conditions did not reach full significance however, $F(1, 105) = 2.69$, $MSE = 29.5$, $p = .10$.

There was also a main effect of condition, $F(2, 105) = 9.23$, $MSE = 95.4$, $p < .001$, reflecting that the average feature weights differed between conditions. As predicted by the generative model, weights in the Essentialized-Chain-80 were significantly greater than those in the Unconnected-Chain-80 condition, $F(1, 105) = 9.02$, $MSE = 95.4$, $p < .005$. Weights in the Essentialized-Chain-80 condition did not differ from those in the Control condition, $p > .20$.

Feature interactions. The interaction weights are presented in the right-hand panels of Figure 9. 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 6.7 and 3.2, respectively, in the Essentialized-Chain-80 condition, and 11.9 and 10.0 in the Unconnected-Chain-80 condition, as compared to 1.4 and 1.5 in the Control condition. Moreover in both causal conditions the direct terms were larger than the indirect term.

A 3 x 2 ANOVA of the interaction weights was conducted with condition and interaction type as factors. There was a main effect of condition, $F(2, 105) = 33.03$, $MSE = 49.9$, $p < .0001$, reflecting the larger interaction weights in the two causal conditions. Unexpectedly, the weights were significantly larger in the Unconnected-Chain-80 condition than in the Essentialized-Chain-80 condition, $F(1, 105) = 25.8$, $MSE = 49.9$, $p < .0001$, a result we will return to in the Discussion. There was also a significant main effect of interaction type, $F(1, 105) = 12.99$, $MSE = 13.1$, $p < .001$, and a significant interaction between interaction type and condition, $F(2, 105) = 4.36$, $MSE = 13.1$, $p < .05$, reflecting the difference in the direct and indirect interaction terms in the causal conditions.

Feature frequency ratings. Feature frequency ratings for the three characteristic features are presented in Figure 10 and follow the same pattern as the feature regression weights. In the Essentialized-Chain-80 condition features were rated as less prevalent as one moved down the causal chain, from 79.0% for X, to 70.5% for Y, to 65.3% for Z. Feature frequency ratings in the Unconnected-Chain-80 condition, in contrast, did not differ from one another (77.4%, 76.9%, and 75.6%) or from those in the Control condition (76.4%, 76.7%, 78.8%). A 3 x 3 ANOVA of the feature frequency ratings confirmed an effect of condition on the linear trend, $F(2, 105) = 10.5$, $MSE = 78.5$, $p < .0001$, reflecting the different pattern of ratings in the three conditions. The linear effect in the Essentialized-Chain-80 condition was significantly different from both the Unconnected-Chain-80 condition, $F(1, 105) = 16.2$, $MSE = 78.5$, $p < .0001$, and the Control condition, $F(1, 105) = 16.2$, $MSE = 78.5$, $p < .0001$. The linear effect in the Unconnected-Chain-80 condition did not differ significantly from that in the Control condition, $F(1, 105) = 2.03$, $MSE = 78.5$, $p = .16$.

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 11A presents the classification ratings in the Essentialized-Chain-80 condition for the missing-X, missing-Y, and missing-Z items. As expected, the missing-X item was rated 14.4 points lower than the missing-Z item, reflecting a causal status effect. However, despite the fact that both the regression weights and feature frequency ratings showed that feature X was more important than Y, the missing-Y item was not rated significantly higher than the missing-X item. This again reflects the fact that the direct interactions weights were larger than the indirect weights in the Essentialized-Chain-80 condition: whereas the missing-Y item violates two direct interactions, the missing-X item violates one direct and one indirect interaction. As was the case in Experiment 1's Chain-75 condition, the results in Figure 11A differ from those in Ahn et. al. (2000) who found a much larger causal status effect (35 vs. 14 points) and that the missing-Y item was rated significantly higher than the missing-X item. We address these differences in experiments that follow.

Second, the strong effect of coherence in both causal conditions is apparent in the pattern of test item ratings shown in Figure 11B. Whereas in the Control condition ratings are a monotonic function of the number of characteristic features, in the causal conditions incoherent items with 2 or 1 features are rated lower (an average of 19 points) relative to the Control condition and the coherent item 000 is rated higher (17 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.

Discussion

As predicted by the generative model, introducing an explicit essence led to a larger causal status effect, a result indicated by a significant linear effect in the Essentialized-Chain-80 condition but not the Unconnected-Chain-80 condition. Although the difference in the magnitude of the causal status effect in the two conditions did not reach full statistical significance, the significant effect in feature frequency ratings showed that features were deemed less prevalent as one moves down the causal chain in the Essentialized-Chain-80 condition but not the Unconnected-Chain-80 condition. The dependency model, in contrast, predicts that the pattern of feature weights and frequency ratings should be the same in the two conditions.

Another successful prediction of the generative model was that features should be more diagnostic overall in the Essentialized-Chain-80 condition than in the Unconnected-Chain-80 condition, because in the former condition features can be used to infer the underlying category essence (and the presence of that essence is decisive for establishing category membership). Of course, the generative model also predicted that weights in the Essentialized-Chain-80 condition should have been larger than in the Control condition, a result that did not obtain. However, as was the case in Experiment 1, we interpret this finding as indicating that participants used the response scale differently. Whereas the responses of the Control participants were solely based on the number of characteristic features, those of Essentialized-Chain-80 participants were also influenced by interaction between features, leaving less of the response scale available to the Essentialized-80 participants to express their sensitivity to the features considered individually.

Although the greater causal status effect in the Essentialized-Chain-80 condition was predicted, the absence of a significant causal status effect in the Unconnected-Chain-80 condition was somewhat of a surprise, especially given the results from Experiment 1. The Unconnected-Chain-80 condition is very similar to the first experiment's Chain-75 condition, the differences being (a) causal strengths were 80% instead of 75% and (b) the presence of an explicit essence, albeit one that is not causally related to the other features. It is conceivable that the 5% increase in causal strengths may be responsible for reducing the causal status effect, because that increase moves the causal links in the direction of being more deterministic (a situation for which the generative model predicts no causal status effect). In addition, the presence of an explicit essential feature to which the causal chain was not connected may have led participants to assume that the chain was unlikely to be related to any *other* essential property of the category (and of course the generative model claims that essential properties to which the causal chain is causally connected is a necessary condition for the causal status effect).

Another important result of Experiment 2 was 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. However, we again think that this result is due to 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.

As in Experiment 1, we compared the relative variance explained by the causal status effect versus the coherence effect. In the Essentialized-Chain-80 condition, the coherence effect accounted for 90.9% of the variance induced by causal knowledge as compared to 9.1% for the causal status effect. That is, the coherence effect again accounted for an order of magnitude more variance (and this is true despite the absence of the "normal" wording that Marsh and Ahn, 2006, thought might inflate the importance of the coherence effect). In the Unconnected-Chain-80 condition, in which there was no significant causal status effect, the coherence effect accounted for 99.7% of the variance.

Experiment 3

As predicted by the generative model, a causal status effect was promoted in Experiment 2 by causally relating a category's observable features to an explicit essence. Note, however, that this prediction only holds when causal links are probabilistic because each subsequent feature in the causal chain is generated with decreasing certainty. When causal strengths are 80% (as they were in Experiment 2) and there are no other causes, the probability of features X, Y, and Z should be .80, .64, and .51, respectively. In contrast, they should each be present with probability 1.0 when the strengths are 100%. In other words, the generative model predicts that the causal status effect induced by the presence of an explicit essence in Experiment 2 should disappear when causal links are deterministic.

To test this prediction, Experiment 3 consisted of Essentialized-Chain-100 and Unconnected-Chain-100 conditions that were identical to the Essentialized-Chain-80 and Unconnected-Chain-80 conditions of Experiment 2 except that the strengths of the causal links were described as 100%. Because Experiment 2 has already confirmed that feature weights in a Control condition with an essential feature but no causal links do not differ from one another (and because our main interest lies in the comparison between the two causal condition), we did not replicate this control condition in Experiment 3. We predicted that neither causal condition would yield a causal status effect. The dependency model, in contrast, predicts that the stronger links should lead to a larger causal status effect in the Essentialized-Chain-100 and Unconnected-Chain-100 conditions as compared to their counterparts in Experiment 2.

Method

The materials and procedure were identical to those in Experiment 2, except for the causal link strengths of 100%. Seventy-two New York University undergraduates received course credit for participating in this experiment. They were randomly assigned in equal numbers to the Essentialized-Chain-100 and Unconnected-Chain-100 conditions and to one of the six categories.

Results

The average category membership ratings given to each of the 8 test items in each condition are presented in Table 4. There were again no effects of which category participants learned and thus the results are presented collapsed over this factor.

Feature weights. The regression weights averaged over participants for features X, Y, and Z are presented in the left-hand panels of Figure 12. Because the Control condition of Experiment 2 was identical to the two causal conditions in all respects except for the absence of causal links, it is included in Figure 12 for comparison. In fact, neither of the two causal conditions yielded a causal status effect as in both the weight on feature X was *lower* than the one on feature Z (albeit not significantly so). A 2 x 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, $F(1, 70) = 7.34$, $MSE = 128.2$, $p < .01$ reflecting the fact that the weights were larger in the Essentialized-Chain-100 condition than the Unconnected-Chain-100 condition. In cross-experiment comparisons, the linear effect in feature weights in the two causal conditions were not statistically different from that in Experiment 2's Control condition (both *F*s < 1).

Feature interactions. The interaction weights are presented in the right-hand panels of Figure 12. A 2 x 2 ANOVA of the interaction weights revealed a marginally significant main effect of condition, $F(1, 70) = 3.90$, $MSE = 148.3$, $p = .06$, reflecting the larger interaction weights in the Unconnected-Chain-100 condition as compared to the Essentialized-Chain-100 condition. There was also a significant effect of interaction type, $F(1, 70) = 5.57$, $MSE = 10.2$, $p < .05$, reflecting the fact that the direct interaction terms were larger than the indirect one. Although in this analysis the interaction was not significant, $F < 1$, the terms differed significantly in the Unconnected-Chain-100 condition but not the Essentialized-Chain-100 condition.

Feature frequency ratings. Feature frequency ratings for the three characteristic features are presented in Figure 13 and, like the feature regression weights, indicate that a causal status effect was present in neither of the two causal conditions. A 2 x 3 ANOVA revealed no linear effect of the feature, *F*

< 1. There was a main effect of condition, $F(1, 70) = 16.6$, $MSE = 205.6$, $p < .0001$, indicating that the average ratings were higher in the Essentialized-Chain-100 condition (87.5) as compared to the Unconnected-Chain-100 condition (79.5). In addition, a cross experiment comparison revealed that the frequency ratings were larger in the Essentialized-Chain-100 condition as compared to Experiment 2's Control condition (77.3), $p < .0001$. In the 2 x 3 ANOVA the interaction did not approach significance, $F < 1$.

Selected test items. The importance of coherence in this experiment is apparent in the pattern of test item classification ratings. The average ratings for items with 0, 1, 2, and 3 characteristic features were 36.2, 17.1, 36.0, and 95.3 in the Essentialized-Chain-100 condition; 57.3, 15.0, 23.9, and 93.1 in the Unconnected-Chain-100 condition, and 23.1, 38.7, 65.6, and 92.7 in the Control condition. That is, just as in Experiments 1 and 2, the item with 0 characteristic features is rated higher, and those with 1 or 2 features are rated lower, than the corresponding ratings in the Control condition. Once again, when causal relations link a category's characteristic features, participants expect those features to cluster together in one category member and the atypical values to cluster together in another. Items with a mixture of the two types of features are considered incoherent and receive a lower rating as a result.

Discussion

As predicted by the generative model, the increase in the magnitude of the causal status effect found in Experiment 2 that resulted from causally relating features X, Y, and Z to an essential feature was absent in Experiment 3. Instead, as in Experiment 1's Chain-100 condition, the presence of deterministic causal links eliminated any sign of a causal status effect. The dependency model, in contrast, predicted that the causal status effect should be larger in this experiment than in the corresponding conditions of Experiment 2 (in which the causal links were weaker) when in fact it was smaller (i.e., zero).

Recall that in Experiment 2 features were more diagnostic overall when they were causally related to an underlying essential feature, that is, in the Essentialized-Chain-80 condition as compared to the Unconnected-Chain-80 condition. Similarly, in Experiment 3 features were more diagnostic in the Essentialized-Chain-100 condition as compared to the Unconnected-Chain-100 condition, an effect predicted by the generative model because in the former condition those features could be used to infer the underlying essence. Although the Essentialized-Chain-100 condition's feature regression weights were not larger than those in the Control conditions, as we have mentioned this result was most likely due to participants using some of the rating scale to express the coherence effect. Feature frequency ratings, in contrast, were much larger in the Essentialized-Chain-100 condition than in both the Unconnected-Chain-100 and Control conditions. Note that the dependency model of course incorrectly predicted that the pattern of features weights should be same in the two causal conditions.

Finally, as in Experiments 1 and 2 the present experiment yielded strong coherence effects. The results were not in perfect accord with the predictions of the generative model, because whereas it predicts that the direct and indirect interaction terms should be weighed equally when causal links are deterministic, the direct terms were significantly larger than the indirect one, especially in the Unconnected-Chain-100 condition. Overall, though, the generative model did a good job of predicting the overall results from Experiment 3.

PART 2: TESTING ADDITIONAL SOURCES OF THE CAUSAL STATUS EFFECT

Experiments 4-6

Taken together, the results of Experiments 1-3 provide strong support for the generative model. As predicted, Experiment 1 found a causal status effect when causal links were probabilistic but not deterministic. Also as predicted, the causal status effect was strengthened when a category's features were related to an underlying category essence when causal links were probabilistic (Experiment 2) but not when they were deterministic (Experiment 3). In all three experiments the pattern of feature regression weights was corroborated by feature frequency ratings that showed the same pattern. Finally, all six causal conditions in Experiments 1-3 yielded large coherence effects. Note that whereas the generative model predicted all of these effects, the dependency model predicted none of them.

But while these results support the generative model, a number of questions regarding the causal status effect remain. Although a causal status effect obtained in our first two experiments it nevertheless differed both quantitatively and qualitatively from some previous studies. For example, whereas Ahn et al. (2000) found that an exemplar missing only feature X was rated 35 points lower than one missing only Z, that difference was only 6 points in Experiment 1's Chain-75 condition and 14 points in Experiment 2's

Essentialized-Chain-80 condition. And, whereas the test item missing only feature X was rated lower than the one missing Y in Ahn et al., the missing-X item was never significantly lower than the missing-Y item in Experiments 1 and 2. These results are doubly surprising, because information about the two factors that we claim promote a causal status effect—an essentialized category and probabilistic causal links—were not explicitly provided in Ahn et al. as it was in Experiments 1 and 2.

One possibility of course is that a larger causal status effect arose in Ahn et al. (2000) because that study instructed participants on different materials. Even without explicit instruction, it is conceivable that the artificial categories tested in Ahn et al. were viewed as more essentialized, and the causal links as more probabilistic, than those in Experiments 1 and 2, resulting in a larger causal status effect. However, this explanation is undermined by the fact that one of the categories tested in Ahn et al. was Romanian Rogos (a type of automobile), an artificial category that was originally taken from Rehder (1997) and which was also one of the six categories tested here. The fact that Romanian Rogos yielded a large causal status effect in Ahn et al. (just like their other categories) but only a small one in Experiments 1 and 2 (just like our other categories) suggests that the differences between studies cannot be attributed to the specific categories tested[iii].

The goal of this article is to provide empirical support for the claim that the generative model is a complete account of the effect of causal knowledge on categorization. However, this claim is potentially undermined by the fact that the causal status effect does not consistently covary with those factors we claim are responsible for the effect: A causal status effect can be small in the presence of explicitly instructed essential features and probabilistic causal links (Experiments 1 and 2) and large in their absence (Ahn et al., 2000). These findings raise the possibility that those variables may not be the only, or even the most important, factors responsible for the causal status effect. Thus, to address this potential criticism we felt it was incumbent on us to also identify the reasons for the much larger causal status effects that have been observed in other studies. To this end, Experiments 4-7 assess a number of variables for their contribution to the causal status effect in the same conditions tested by Ahn et al., namely, in the absence of explicit information about essential features and probabilistic causal links. Because of their similar methods (and, as it turns out, their similar results), we first describe Experiments

4-6 and then present their results jointly. Experiment 7 then follows.

Experiment 4

Experiment 4 tested how the causal status effect might be influenced by the order in which features are presented during the classification test. Recall that Rehder and Kim (2006) found tentative evidence that the causal status effect depends on whether the first dimension in the causal chain (X) appeared first in the list of a test item's features. One reason this sensitivity to presentation order might arise is that the causal status effect depends on cognitive load, and dimensions presented in causal (canonical) order might reduce that load. A sensitivity to load might arise because computing (explicitly or implicitly) the relative weights of features when causal links are probabilistic is computationally more intensive than assuming the links are deterministic (and that all feature weights are thus equal). Indeed, the fact that features were presented in canonical order in Ahn et al. (2000) but in random order in Experiments 1 and 2 might explain the differences in the size of the causal status effect.

To test for an effect of feature presentation order, in Experiment 4's Chain-Random condition, participants were first instructed on the causal chain linking X, Y, and Z and then presented with test items whose features were presented in a random order (but fixed for each participant), just as they were in Experiments 1-3. In contrast, in the Chain-Canonical condition the test item's features were always presented in causal order, namely dimension X then Y then Z. Following Ahn et al. (2000), in neither condition was information about essential attributes or the strength of the causal links provided. Performance of the Chain-Random and Chain-Canonical groups was compared to Control-Random and Control-Canonical conditions, respectively, that were identical except for the omission of the causal links.

Experiment 5

Experiment 5 tested how the causal status effect is affected by how the values on the binary feature dimensions are described. Recall that in many previous studies of causal status and coherence effects, the nontypical values on each dimension were described as "normal" (e.g., "most Myastars have a high density whereas some have normal density"). In others (including Experiments 1-4), the nontypical value was described as opposite of the characteristic value (e.g., high vs. *low* density). This difference corresponds to the familiar distinction between *additive* versus *substitutive* features (Tversky & Gati, 1978). As mentioned, Marsh and Ahn (2006) argued that use of additive features may inflate the coherence effect (and reduce the causal status effect). For example, if participants learn that "most" Myastars are very hot, have high density, and have a large number of planets, and that "some" have a normal temperature, normal density, and a normal number of planets, coherence effects might emerge simply because participants expected the normal values to all appear together.

To test this conjecture, Experiment 5 compared a Chain-Substitutive condition in which dimensions were described as having two opposing values (just as in Experiments 1-4) with a Chain-Additive condition in which the atypical value was described as normal. Features were always presented in canonical order. Performance of the Chain-Substitutive and Chain-Additive groups was compared to Control-Substitutive and Control-Additive conditions, respectively, that were identical except for the absence of causal links.

Experiment 6

Experiment 6 investigates whether the causal status effect is affected by whether the causal relationships are accompanied by information about the mechanism by which the cause produces the effect (Table 1). Conceivably, information about the causal mechanism may have led participants to treat the causal links as more plausible, and thus stronger. Or, because participants were required to answer several multiple choice questions about the causal mechanisms, the links may have been treated as stronger because they were repeated so frequently. On either account, the causal links might be treated as more probabilistic when causal mechanism information is omitted, leading (as shown in Experiment 2) to a larger causal status effect. Indeed, the fact that causal mechanism information was omitted in Sloman et al. (1998) might explain that study's large causal status effect (29 points on a 100-point scale).

To test for an effect of causal mechanism information, Experiment 6 compared a Chain-Mechanism condition that provided mechanism information with a Chain-No-Mechanism condition that omitted it. In the Chain-No-Mechanism condition the tutorial provided no information about the causal relationships, including the causal mechanisms. Instead, the causal links were displayed as a diagram at the top of the screen during each trial of the classification test. In contrast, the Chain-Mechanism condition was just like Experiments 1-5 in that the tutorial presented the causal links and mechanisms and then tested participants on that information during the multiple-choice test. To make the conditions the same except for the causal mechanism information, the diagram was also presented on each classification trial in the Chain-Mechanism condition. Both conditions used substitutive feature dimensions (e.g., high vs. low density) and presented features in canonical order.

Results and Discussion of Experiments 4-6

The classification ratings of individual test items in each condition of Experiments 4-6 are presented in the Appendix. The essential measures from those experiments, including the average feature regression weights, the weights on the direct and indirect interaction terms, the average feature frequency ratings, and the classification ratings for the missing-X, missing-Y, and missing-Z test items, are summarized in Table 5. For purposes of comparison, Table 5 also includes the results from two previous studies (Ahn et al., 2000 and Rehder & Kim, 2006), Experiments 1-3, and the upcoming Experiment 7. The results from Experiments 4-6 are straightforward: Not only did none of the manipulations enhance the causal status effect, a significant causal status effect was absent in all six of the causal conditions. This was true regardless of whether one assesses a causal status effect by the regression weights, the frequency ratings, or the classification ratings of the items missing single features.

Although the absence of a causal status effect in Experiments 4-6 is surprising given previous studies that found a large causal status effect, it is understandable in light of the results from our earlier experiments: Whereas a causal status effect was promoted by providing explicit information about probabilistic causal links and essential category features in Experiment 1 and 2, neither of those factors was present in Experiments 4-6. One potential explanation for this finding is that the results in the six causal conditions were quite similar to those in Experiment 1's Chain-100 condition that also failed to exhibit a causal status effect, suggesting that participants in this experiment assumed that the causal links were deterministic, or nearly so. We'll return to this interpretation in the General Discussion.

The absence of a causal status effect in Experiments 4-6 contrasts with the presence of large coherence effects. In every causal condition, the weights on the interaction terms were each statistically greater than the weights in their corresponding control conditions. Moreover, the coherence effect accounted for more than 99% of the variance induced by causal knowledge in each of these experiments.

This large effect of coherence can be seen in Figure 14 that presents the test item ratings as a function of the number of characteristic features, collapsed over the 6 causal conditions and 5 control conditions of Experiments 4-6. Whereas ratings increase monotonically with the number of characteristic features in the Control conditions, in the causal conditions the ratings of incoherent items with 1 or 2 characteristic features are lower than the item with 0 (i.e., 000). As in Experiments 1-3, when causal relations link a category's characteristic features, participants expect those features to cluster together in one category member and the atypical values to cluster together in another. This change in the expected distribution of typical and atypical features is by far the largest influence of interfeature causal relations on category membership judgments.

Although participants' category membership judgments were similar across Experiments 4-6, the results of some of the individual experiments are worthy of discussion. First, Experiment 5 tested the conjecture that a causal status effect might be enhanced and coherence effects reduced when substitutive rather additive features are used. In fact, the interaction terms were much larger in the Chain-Substitutive condition (average weight of 12.8) than in the Chain-Additive condition in which atypical values were described as "normal" (average of 7.8). Moreover, there was a greater sign of causal status effect in the Chain-Additive condition (e.g., a marginally significant difference of 1.6 in the regression weights of features X and Z) than in the Chain-Substitutive condition. These results are exactly the opposite of those predicted by Marsh and Ahn (2006) who argued that use of "normal" feature values artificially inflates the coherence effect and diminishes the causal status effect.

Second, Experiment 6 provided some weak evidence that omitting information about causal mechanisms enhances the causal status effect. On the one hand, whereas the Chain-Mechanism condition showed no sign of a causal status effect, in the Chain-No-Mechanism condition, both the feature regression weights and frequency ratings reflected a monotonic decrease in the importance from feature X, to Y, to Z. Moreover, the missing-X item was rated lower (38.9) than the missing-Y item (40.2) which was rated lower than the missing-Z item (44.2). Nevertheless, of these effects only the decrease in feature frequency ratings reached marginal significance.

In summary, although there was evidence that some of the manipulations tested in Experiments 4-

6 might contribute (very modestly) to the causal status effect, clearly none of these factors is the key variable explaining the large causal status effect found in some studies. Experiment 7 tests one final possibility.

Experiment 7

Experiment 7 investigates one final variable that might contribute to the causal status effect, namely, the number of items presented during the classification test. On the one hand, those studies that found a small causal status effect have asked participants to produce a classification rating for every test item formed by systematically varying the values on each binary dimension (e.g., Rehder, 2003; Rehder & Kim, 2006; and the present Experiments 1-6). This resulted, in Experiments 1-6 for example, in asking participants about 8 distinct test item (each of which were presented twice). In contrast, participants in studies with the largest causal status effects classified only those items missing one of the characteristic features, that is, a missing-X item, a missing-Y item, and a missing-Z item (e.g., Ahn et al. 2000; Sloman et al. 1998). Conceivably, with only three test items to rate, participants may have felt free to use a greater portion of the response scale to express any (otherwise small) difference between those items. Or, those items might have been seen as equally good category members, but the perceived need to rate the items differently led participants to search for some basis for distinguishing them. The fact that items differed only on which feature was missing may have triggered a comparison of the relative importance of those features that wouldn't have been present otherwise. In other words, a causal status effect may have arisen in these studies because of the demands of the task.

To test this hypothesis, Experiment 7 tested a Chain-No-Mechanism-3Q condition that was identical to the Chain-No-Mechanism condition of Experiment 6 except that participants rated the category membership of only the items missing X, missing Y, and missing Z. Moreover, participants previewed the three test items without making a categorization judgment so that the experimenter could point out which characteristic feature was missing in each. As in Experiment 6's Chain-No-Mechanism condition, features were always presented in canonical order, no causal mechanism information was provided, and a diagram of the causal relationships was presented on each classification and feature frequency trial. We hypothesized that a large causal status effect would obtain in the Chain-No-

Mechanism-3Q condition as compared to the Control condition (and Experiment 6's Chain-No-Mechanism condition).

Method

Materials. The materials were the same as in Experiments 4-6.

Participants. Seventy-two New York University undergraduates received course credit for participating. They were assigned in equal numbers to the Chain-No-Mechanism-3Q and Control conditions and to one of the six categories.

Procedure. For all participants, no causal relationships were presented during the initial tutorial. Instead, in the Chain-No-Mechanism-3Q condition causal relationships were presented as a diagram at the top of the screen on each classification trial. However, in this experiment only three classification trials were presented. Participants first observed the missing-X, missing-Y, and missing-Z trials (in that order) without producing a rating. For each item the experimenter noted that it possessed 2 out of the 3 characteristic features and pointed out which feature was missing. The test items were then presented a second time (in random order) so that participants could enter their rating. After the classification test, participants performed a feature frequency rating task that was identical to the one in Experiments 1-6 except that each trial was presented once rather than twice.

Results

There were again no effects of which category participants learned. Results are presented in Table 5.

Test item ratings. The average classification ratings for the missing-X, missing-Y, and missing-Z items are presented in Figure 15 for the Chain-No-Mechanism-3Q and Control conditions. Because the Chain-No-Mechanism condition of Experiment 6 was identical to the Chain-No-Mechanism-3Q condition except for the 8 versus 3 distinct test items, results from that condition are also included in Figure 15 for comparison. The results are straightforward: A much larger causal status effects arose in the Chain-No-Mechanism-3Q condition than in the other two conditions. Whereas the difference between the missing-X and missing-Z items was 0 and 5 points in the Control condition and the Chain-No-Mechanism conditions, respectively, that difference rose to 25 points in the Chain-No-Mechanism-3Q condition.

Moreover, for the first time the missing-Y item was rated significantly higher than the missing-X item.

A 3 x 2 ANOVA yielded a main effect of chain versus control, $F(1, 70) = 19.36$, $MSE = 1136$, $p < .0001$, a main effect of feature, $F(2, 140) = 10.33$, $MSE = 306$, $p < .0001$, and an interaction between condition and the linear effect of feature, $F(1, 70) = 16.34$, $MSE = 344$, $p < .0001$, confirming the different pattern of ratings in the two condition. We also conducted a cross-experiment comparison of the Chain-No-Mechanism-3Q and Chain-No-Mechanism condition, and this 3 x 2 ANOVA also yielded an interaction between condition and the linear effect of feature, $F(1, 70) = 9.79$, $MSE = 355$, $p < .01$, confirming the larger causal status effect that arises when participants rate 3 versus 8 test items.

Feature frequency ratings. Table 5 indicates that feature frequency ratings exhibited a decrease in the Chain-No-Mechanism-3Q condition, from 78.7 to 74.2 to 71.2. In contrast, no such decrease was observed in the Control condition. A 3 x 2 ANOVA yielded a main effect of feature, $F(2, 140) = 5.58$, $MSE = 67$, $p < .01$, and a marginally significant interaction between condition and the linear effect of feature, $F(1, 70) = 3.14$, $MSE = 102$, $p = .08$. An analysis of the Chain-No-Mechanism-3Q condition alone revealed a significant linear trend, $p < .05$. Interestingly, whereas the test item ratings were significantly different in the Chain-No-Mechanism-3Q condition and the Chain-No-Mechanism condition of Experiment 6 (Figure 15), a cross experiment analysis revealed that the linear trends in the feature frequency ratings in those conditions were not significantly different, $F < 1$.

Discussion

The results of Experiment 7 confirm that the size of the causal status effect is larger when participants rate 3 versus 8 test items. Whereas the difference between the missing-X and missing-Z items was 5 points in Experiment 6's Chain-No-Mechanism condition, that difference increased to 25 points in the Chain-No-Mechanism-3Q condition. Although this difference is still not as large as the 35 points in Ahn et al. (2000), comparison of Figure 15 with Figure 2A indicates a generally similar pattern of results over the two studies.

Note that, in contrast to the classification test, the same items were presented during the feature frequency tests of Experiments 6 and 7, and thus unsurprisingly there were no differences in the feature frequency ratings in the Chain-No-Mechanism and Chain-No-Mechanism-3Q conditions. This finding suggests that differences in the classification results between these two conditions do not reflect differences in participants' underlying beliefs about the diagnosticity of features. Rather, the large causal status effect arose in this experiment either because participants felt free to use a larger portion of the response scale to express otherwise small differences or because task demands led to a consideration of the relative importance of features that wouldn't have been present otherwise. In summary, the results of Experiments 4-7 indicate that the primary factor responsible for the large differences in the size of the causal status effect across studies is the number items presented during the classification test.

General Discussion

This article has addressed the changes in classification performance brought about by causal knowledge that links category features. In the first section below 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' perceived category validity. The coherence effect is discussed in the final section.

Variables that Promote a Causal Status Effect

One contribution of this research is that it has identified with greater precision than before those conditions in which a causal status effect is (and isn't) likely to appear. We found that both probabilistic causal links and essentialized categories promote a causal status effect. The importance of probabilistic causal links was demonstrated in Experiment 1 that found a causal status effect in the Chain-75 condition but not the Chain-100 condition. In addition, the causal status effect found in Experiment 2's Essentialized-Chain-80 condition disappeared in Experiment 3's Essentialized-Chain-100 condition that was identical except for deterministic causal links. The importance of essentialized categories was demonstrated in Experiment 2 that found a larger causal status effect in the Essentialized-Chain-80 condition than in the Unconnected-Chain-80 condition that was identical except that the chain of observable features was not causally related to the essence.

Why should probabilistic causal links and essentialized categories promote a causal status effect? Stated intuitively, what we think is going on is this. When confronted with a causal chain of features, classifiers will often adopt a "generative" perspective, that is, they will think about the likelihood of each successive event in the chain. Conceivably, this process may be equivalent to a kind of mental simulation in which they repeatedly "run" (consciously or unconsciously) a causal chain and examine the likely outcomes. When causal links are deterministic, in each simulation run events in the chain are either all present or all absent: if X is assumed present then Y and Z will also be present and if X is assumed absent then Y and Z will also be absent (assuming the absence of alternative causes of Y and Z). Thus, because $P(X) = P(Y) = P(Z)$, each feature dimension is weighed equally in categorization judgments. But when links are probabilistic, each successive event is less likely: Because Y is only sometimes produced by X, it will be less prevalent than X; because Z is only sometimes produced by Y, it will be less prevalent than Y. Thus, because $P(X) > P(Y) > P(Z)$, each successive feature dimension is assigned a lower weight.

Classifiers sometimes spontaneously adopt a generative perspective on their own, of course. For example, we think they did in Experiment 1's Chain-75 condition (and in the other cases of a causal status effect in the literature). But this perspective may become more likely when the head of the causal chain is an invariant property of the category because classifiers are naturally likely to focus their attention on that property which, because it is responsible for producing the rest of the features, triggers the kind of mental simulations we've described. Alternatively, the essential feature might induce participants to engage in a form of *diagnostic reasoning* in which they reason backwards from observable features to the essence, and features that are causally closer (such as X) are more diagnostic of that essence than more distant features (Z). Either account predicts a stronger causal status effect for a category with an essentialized feature that causally generates the observable features.

The dependency model, in contrast, is based on competing intuitions regarding which features are important in people's conceptual representations, namely, those that 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. But despite the plausibility of this intuition, it is does not conform to our subject's category membership judgments. For example, 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, however, that the causal status effect was weaker (indeed, absent) for deterministic causal links and stronger for probabilistic ones (Experiment 1). And, whereas the dependency model predicts that a feature's weight should increase with its number of effects, we found (in Experiment 2) that introducing an essential feature as a *cause* produced a stronger causal status effect. In other words, neither of our main findings regarding the effect of causal knowledge on feature weights can be accounted for by the intuition underlying the dependency model.

The importance of the two factors that promote a causal status effect—probabilistic causal links and essential features—was further reinforced by Experiments 4-6 that provided information about neither of those factors. These experiments tested whether the causal status effect would be affected by (a) presenting features in causal order, (b) describing the atypical value on each binary dimension as "normal," and (c) omitting causal mechanism information. In fact, not only did none of these manipulations enhance the causal status effect, a fully significant causal status effect was absent in all six of the causal conditions. Indeed, we suspect that the most likely explanation of these results is that, without information to the contrary, participants treated the causal links as deterministic, or nearly so—conditions under which the generative model predicts that a causal status effects should be small or nonexistent. Although we did not directly assess participants' beliefs about causal strengths in these experiments, we can cite two additional sources of evidence in favor of this interpretation. First, the results in the chain conditions of Experiments 4-6 were very similar to those in the Chain-100 condition of Experiment 1 that specified explicit causal strengths of 100%. In particular, in five of the six causal conditions in those experiments there was no difference between the magnitude of the direct and indirect interaction terms, a result consistent with the predictions of the generative model for deterministic links. Second, as part of another experiment testing the same materials we asked participants for explicit causal strength judgments and found an average link strength of 91%; the modal answer was 100%. Under these conditions, the generative model predicts only a weak causal status effect (and a small difference between the direct and indirect interaction terms). Note that near-deterministic causal links explain the small causal status effects found in many of our other studies testing the same materials (e.g. Rehder & Kim, 2006; Rehder, 2003b).

Besides probabilistic causal links and essential features, the only other experimental manipulation that promoted a causal status effect was the number of test items presented during the classification test.

Although our claim is that the generative model provides a complete explanation of the causal status effect, large differences in its size across studies raised the possibility that it might be influenced by other theoretically important factors, ones not accounted for by the generative model. Experiment 7 assessed the effect of only presenting missing-X, missing-Y, and missing-Z test items in the same conditions as Experiment 6's Chain-No-Mechanism condition and found that presenting only three test items yielded a difference between the missing-X and missing-Z test items of 25 points as compared to 5 points in Experiment 6. One interpretation of this finding is that presenting a small number of test items that normally have almost identical classification ratings amplified the difference between them. Another is that the small number of items introduced task demands that led participants to search for some basis to distinguish those items. In other words, the large differences in the size of the causal status effect across studies can be attributed to a procedural variable of secondary theoretical importance.

Finally, there was one other effect on feature weights that is worth noting, namely, that in both Experiments 2 and 3 features were weighed more heavily overall when observable features were causally related to the essential feature. This result is also readily explicable in terms of the generative perspective we suggest participants adopted: The likelihood with which each event (feature) in the chain occurs is greater when the head of the chain is itself always present. Or, as we have mentioned, participants might be engaged in a process of causal reasoning in which they reasoned backward from X, Y, and Z to the essence. On this view, X, Y, and Z are more diagnostic overall because they also provided evidence for an additional feature (one that happened to be defining of category membership). But on either account, these results corroborate other demonstrating the importance of essential causes to a feature's categorization weight. For example, using a two-alternative forced-choice procedure in which participants were required to choose between two categories, Rehder and Kim (2007) found that features had greater categorization weight when they were described as generated by an underlying essential feature (also see Rehder, 2003b, Experiment 3). And, we have shown that a cause increases a feature's categorization weight even when it is not an essential feature. For example, Rehder and Kim (2006) systematically manipulated a category feature's number of (nonessential) causes and found that, as predicted by the generative model, its influence on categorization increased when it had many rather than few causes (also see Rehder & Hastie, 2001; Rehder, 2003a).

In summary then, the present results add to others supporting the generative model as an account of how causal knowledge affects feature importance. Besides three-feature causal chains, the generative model has been shown to account for classification ratings when features exhibit a four-feature causal chain (Rehder, 2003b), a common cause structure (one feature causes many, Rehder, 2003a), a common effect structure (one feature caused by many, Rehder, 2003a), and three different network topologies linking five features (Rehder & Kim, 2006). In contrast, the present results provide another counterexample to the predictions of the dependency model.

Category Validity Mediates the Causal Status Effect

Our second main finding concerns how the change to features' categorization importance brought about by causal knowledge is mediated by their perceived category validity. According to the generative model, causal knowledge changes the perceived likelihood with which a feature is produced ("generated") by a category's causal model. And, all else being equal, 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 frequency ratings were not significantly different from one another.

Other studies have shown that a feature’s influence on categorization judgments correlates with its perceived 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 perceptual salience and diagnosticity (or *cue validity*, assessed with questions like "Of all things that grow on trees, what percentage are apples?"). Measures that loaded on the third factor included ones related to a construct referred to as *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, Study 5 of Sloman et al. (1998) attempted to dissociate mutability and category validity but found that the two judgments tracked one another. Because categorization ratings in Experiments 1-7 were generated with respect to a single category (and because the artificial features used in our 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. And, like Sloman et al., we found that that measure did not dissociate from category validity (i.e., frequency).

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 frequency 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' frequency 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 frequency 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 1-7 gathered frequency ratings after the causal relationships were learned and found no dissociation with centrality measures.

The Dominance of Coherence in Causal-Based Categorization

This article's third major finding concerns coherence effects. As compared to the causal status effect that appeared in only 3 of the 11 causal conditions, the coherence effect was highly robust, appearing in 10 out of 10 of the conditions in which it could be assessed (it could not be assessed in Experiment 7). These results indicate that the most pronounced effect of a category's causal knowledge is to get classifiers to attend to whether a potential category member's particular configuration of features makes sense in light of that knowledge—in particular, it induces in them an expectation that a set of causally related features will tend to be all present or all absent. In the present experiments, this effect of coherence was so pronounced that it led participants to generate higher categorization ratings for items that had fewer typical features. For example, in every causal condition an item with zero typical features

(000) was rated higher than items with one typical feature (and sometimes higher than those with two typical features) because 000 is a sensible combination of features in light of the causal laws. In contrast, test items with a combination of present and absent features are inconsistent with those laws. Note that this does not mean that our participants ignored the typicality of features considered individually, because the prototypical item with all three features (111) was always rated much higher than 000. But it does mean that causal knowledge changed participants' beliefs about the likely distribution of features so that items with a consistent set of features (111 and 000) became the most likely category members and those with a mixture of present and absent values became unlikely.

Our participants also exhibited sensitivity to the more subtle pattern of interfeature correlations one expects to be generated by causal relationships. Whereas X, Y and Z should all be perfectly correlated when causal links are deterministic (and there are no other causes of Y and Z), those correlations should be less than 1 when causal links are probabilistic and X and Y, and Y and Z, should be more strongly correlated than X and Z. The results of Experiments 1-6 were almost entirely consistent with these predictions. First, the direct and indirect interaction terms in Experiments 1's Chain-100 condition were equal and larger than those in the Chain-75 condition in which the indirect interaction term was smaller than the direct ones. Second, direct terms were larger than indirect terms in both probabilistic conditions of Experiment 2 (although that difference was marginal in the Uncconected-Chain-80 condition). Third, those terms were not different in 6 of the 8 causal conditions of Experiments 3-6, conditions where the absence of a causal status effect suggests that causal links were construed as deterministic. These results strongly support the claim that participants estimate the likelihood that a test exemplar was generated by a category's causal model as part of rendering category membership judgments. The dependency model, of course, predicted none of these effects.

Note that robust coherence effects obtained in the present experiments despite Marsh and Ahn's (2006) claim 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, Expt. 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 a second dimension, namely, participants in the Unambiguous but not the Ambiguous condition were told that category members "tended to have" the characteristic value on each dimension. It is unsurprising that the judgments of the Ambiguous participants were dominated by coherence given that they were provided with no information about which features were typical of the category. In contrast, the present Experiment 5 compared two conditions that were identical except for the "normal" wording and found results exactly the opposite of the Marsh and Ahn claim: The "normal" wording produced a much *smaller* coherence effect and a larger (albeit nonsignificant) causal status effect. Note that these results of Experiment 5 have been replicated with four-element causal chains as well (Rehder & Kim, 2008).

Why might substitutive dimensions lead to stronger coherence effects as compared to the "normal" wording? One possibility is that such dimensions might lead participants to infer the existence of additional causal links. For example, if you are told that some Myastars have either high or low temperature and either high or low density, and that high temperature causes high density, you might take this to mean that *low* temperature also causes *low* density. On this account, a star with high temperature and low density violates two causal links rather than one. Or, you might treat temperature and density as continuous variables, in which case high temperature and low density is a more egregious violation of the causal link than high temperature and *normal* density (which implies that density is at some intermediate point on the density scale). But whatever the reason, it is clear that substitutive dimensions highlight, rather than diminish, the importance of the correlational structure among features.

The importance of coherence to classification has been documented by a number of 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 Wisniewski & Murphy, 1989). Similarly, Rehder and Ross (2001) showed that artifacts were considered 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”), and worse examples when their features were incoherent (“has a magnet” and “removes mosquitoes”). Coherence also affects other types of category-related judgments. Rehder and Hastie (2004) found that participants’ willingness to generalize a novel property displayed by an exemplar to an entire category varied as a function of the exemplar’s coherence. Maximally coherent exemplars that satisfied all of category’s causal laws supported the strongest generalizations. 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 non-category 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 et al., 1986) and unsupervised (Ahn & Medin, 1992; Kaplan & Murphy, 1999; Medin, Wattenmaker, & Hampson, 1987).

Finally, it is illuminating to compare the relative importance of coherence and the causal status effect. Not only did the coherence effect never account for less than 90% of the variance in Experiments 1-6, it has dominated participants' categorization judgments in every 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). And, 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 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 especially puzzling given that such a "strong case" was present in their own data. Instead, these analyses indicate that by far the most important factor that categorizers consider when using causal laws to generate categorization judgments is whether an object displays a configuration of features that make sense in light of those laws. Compared to coherence, changes to the importance of individual features runs a distant second.

Summary

This article has presented five main findings. First a causal status effect is promoted by probabilistic causal links. Second, it is also promoted by an essentialized category, albeit only for probabilistic causal links. Third, the causal status effect is mediated by features' subjective category validity. Fourth, coherence effects were highly robust, appearing in every condition in which they were assessed. Finally, coherence effects never accounted for less than 90% of the variance induced by causal knowledge. The causal status effect, in comparison, never accounted for more than 10% of the variance. All of these findings are consistent with a generative model of classification and inconsistent with a dependency model.

References

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

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

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

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

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

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: 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 Experiemental Psychology: General, 131*, 451-476.

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

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.

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.

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

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

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

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

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

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

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. (2007). *Classifying with essentialized categories*. Submitted for publication.

Rehder, B. & Kim, S. (2008). The role of coherence in causal-based categorization. Paper submitted to the 30th Annual Conference of the Cognitive Science Society.

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., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence.

Cognitive Science, *22*, 189-228.

Tversky, A., & Gati, I. (1978). Studies in similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 79-98). Hillsdale, NJ: Erlbaum.

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.

Appendix

Empirical Results from Experiments 4-6

The average classification ratings for each of the 8 test items in each condition of Experiments 4-6 are presented in Tables A1, A2, and A3, respectively.

Table A1
Classification ratings from Experiment 4. Standard errors are shown in parentheses.

Test Item	Chain-Random	Control-Random	Chain-Canonical	Control-Canonical
000	51.0 (6.1)	29.3 (7.9)	59.4 (6.5)	25.1 (4.9)
001	21.7 (3.4)	42.1 (6.9)	19.7 (2.8)	32.1 (3.8)
010	21.2 (3.4)	39.2 (5.6)	15.2 (2.3)	31.7 (3.8)
100	26.7 (3.8)	44.9 (5.6)	20.2 (2.8)	32.7 (3.7)
011	37.4 (5.1)	68.5 (3.6)	27.5 (4.1)	62.2 (4.2)
101	29.8 (4.6)	65.6 (5.5)	21.9 (3.8)	63.6 (3.5)
110	30.1 (4.3)	63.9 (5.4)	29.6 (4.3)	63.8 (3.4)
111	93.3 (1.9)	94.6 (1.8)	94.9 (1.8)	84.9 (4.1)

Table A2
Classification ratings from Experiment 5. Standard errors are shown in parentheses.

Test Item	Chain-Substitutive	Chain-Additive	Control- Substitutive	Control-Additive
000	56.5 (6.9)	52.4 (5.2)	32.2 (6.7)	39.2 (6.1)
001	19.4 (2.7)	33.1 (3.4)	43.2 (5.9)	51.8 (4.7)
010	15.4 (3.0)	31.8 (3.6)	42.4 (5.2)	51.1 (4.6)
100	20.2 (4.4)	37.3 (4.1)	43.5 (5.8)	52.6 (4.7)
011	29.9 (5.2)	47.2 (4.5)	71.3 (2.7)	72.9 (2.8)
101	28.5 (4.8)	45.4 (4.3)	69.2 (2.9)	73.9 (3.0)
110	28.6 (5.0)	49.2 (4.6)	67.5 (2.7)	72.4 (3.3)
111	95.4 (1.2)	93.0 (1.7)	92.5 (2.7)	92.5 (3.6)

Table A3
Classification ratings from Experiment 6. Standard errors are shown in parentheses.

Test Item	Chain-Mechanism	Chain-No-Mechanism	Control
000	60.2 (6.3)	54.9 (6.7)	21.7 (4.4)
001	17.8 (2.6)	27.5 (3.5)	33.8 (2.3)
010	16.7 (2.7)	25.9 (3.4)	34.8 (2.4)
100	18.4 (2.7)	29.9 (3.2)	33.0 (2.3)
011	30.3 (4.3)	38.9 (4.2)	64.3 (2.5)
101	29.0 (4.3)	40.2 (4.2)	62.4 (2.8)
110	30.4 (4.7)	44.2 (4.4)	63.0 (2.6)
111	95.3 (1.7)	95.8 (1.2)	91.8 (2.2)

Author Note

Bob Rehder and ShinWoo Kim, Department of Psychology, New York University.

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

Footnotes

---

[i] 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 & Wellman, 1991; Gelman, 2003; Johnson & Solomon, 1997). And, an essential property not just one that just happens to be present in all category members (and absent in all nonmembers), it is one that is present in all category members that *could* exist. But while 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 causal status effect.

[ii] Causal strengths were set to 80% rather than 75% as in Experiment 1's Chain-75 condition so as to not be inconsistent with characteristic features occurring in "most" category members. If E causes X with probability 80%, then X should be present in 80% of category members. It then follows that Y should be present in 64% of category members and Z should be present in 51%. A causal strength of 75% would have resulted in Z appearing in fewer than 50% of category members.

[iii] Whereas the difference between the missing-Z and missing-X items was 43.1 points in Ahn et al. for Romanian Rogos, that difference was -4.6 in Experiments 1's Chain-75 condition and 11.2 in Experiments 2's Essentialized-Chain-80 condition.

Table 1
Features and causal relationship for Myastars, an artificial category.

	Features
X	Very hot
Y	High density
Z	Large number of planets
	Causal Relationships
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.

Table 2
Classification ratings from Experiment 1. Standard errors are shown in parentheses.

Test Item	Chain-100	Chain-75	Control
000	47.4 (6.3)	52.4 (6.2)	19.7 (4.6)
001	13.4 (2.3)	24.5 (2.8)	40.6 (3.1)
010	14.1 (2.2)	25.2 (2.8)	39.0 (3.6)
100	13.9 (2.4)	32.6 (3.2)	40.2 (3.4)
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)
111	94.2 (2.1)	93.8 (1.8)	93.3 (2.2)

Table 3
Classification ratings from Experiment 2. Standard errors are shown in parentheses.

Test Item	Essentialized-Chain-80	Unconnected-Chain-80	Control
000	31.3 (5.1)	49.7 (5.4)	23.1 (3.7)
001	29.2 (2.8)	23.2 (2.4)	38.8 (2.,7)
010	24.2 (2.9)	20.3 (2.5)	38.0 (3.3)
100	32.9 (2.0)	25.1 (2.9)	39.2 (3.1)
011	46.6 (3.4)	30.0 (3.8)	67.2 (2.1)
101	46.7 (3.5)	28.3 (3.6)	66.5 (1.9)
110	61.0 (3.0)	34.4 (4.1)	63.2 (2.3)
111	92.9 (1.5)	94.0 (1.6)	92.7 (2.1)

Table 4

Classification ratings from Experiment 3. Standard errors are shown in parentheses.

Test Item	Essentialized-Chain-100	Unconnected-Chain-100
000	36.2 (6.1)	57.3 (6.3)
001	17.2 (2.9)	19.0 (3.0)
010	17.2 (2.6)	12.5 (2.1)
100	16.9 (2.6)	13.5 (2.6)
011	38.7 (5.1)	26.3 (4.6)
101	33.6 (4.6)	22.1 (4.5)
110	35.8 (4.9)	23.3 (4.3)
111	95.3 (2.0)	93.1 (1.9)

Table 5

Classification ratings from two previous studies and Experiments 1-7. Results indicating a significant causal status effect are in bold. Significant coherence effects are shown in italics.

			Feature Regression Weights				Feature Interaction Weights			Feature Frequency Ratings				Test Item Ratings			
														Missing X	Missing Y	Missing Z	
Study	Condition	N	X	Y	Z	X – Z	Direct	Indirect	Diff.	X	Y	Z	X – Z	(011)	(101)	(110)	Z – X
Ahn et al. (2000, Expt. 1)	Causal	15												27.2	40.5	62.0	**34.8***
	Control	15												43.2	47.3	47.8	4.6
Rehder & Kim (2006)	1-1-1	192	7.6	6.4	6.2	**1.4***	*4.4*	*2.9*	1.4					48.5	48.7	52.5	**3.9***
Experiment 1	Chain-100	36	6.2	7.7	6.6	0.4	*13.6*	*12.5*	1.1	77.8	77.3	77.0	0.7	24.4	18.7	22.2	-2.1
	Chain-75	36	8.6	7.3	5.1	**3.5***	*10.3*	*8.2*	2.1*	76.7	73.7	70.2	**6.5***	40.6	38.7	46.7	6.1
	Control	36	12.4	12.2	13.0	0.5	0.9	0.7	0.2	77.2	74.7	76.7	0.4	65.1	66.0	66.1	1.0
Experiment 2	Essentialized-Chain-80	36	12.8	10.6	8.3	**4.5***	*6.7*	*3.2*	3.5*	79.0	70.5	65.3	**13.7***	46.6	46.7	61.0	**14.4***
	Unconnected-Chain-80	36	7.3	6.6	5.8	1.6	*11.9*	*10.0*	1.9†	77.3	76.8	75.5	1.8	30	28.3	34.4	4.4
	Control	36	11.8	11.7	12.7	-0.9	1.4	1.5	–0.1	76.7	77.4	78.9	-2.1†	67.2	66.5	63.2	-4.0
Experiment 3	Essentialized-Chain-100	36	9.0	10.4	9.9	-0.8	*10.1*	*9.2*	0.9	88.5	86.4	87.5	1.0	38.6	33.6	35.8	-0.8
	Unconnected-Chain-100	36	4.6	5.4	6.7	-2.1	*14.4*	*12.8*	1.6*	80.0	79.4	79.1	0.9	26.3	22.1	23.3	-3.0
Experiment 4	Chain-Random	36	6.1	6.6	6.6	-0.5	*11.7*	*9.9*	1.8	78.3	78.6	79.3	-1.0	37.4	29.8	30.1	-7.3
	Chain-Canonical	36	5.4	5.6	5.7	-0.3	*14.7*	*11.8*	2.8*	79.2	76.7	77.2	1.9	28.2	20.5	27.2	-1.0
	Control-Random	18	11.2	10.5	11.7	-0.4	2.4	1.2	1.2	79.7	78.5	80.1	-0.4	68.5	65.6	63.9	-4.6
	Control-Canonical	18	11.7	11.1	11.2	0.5	1.8	1.8	0	80.9	78.7	80.1	0.9	62.2	63.6	63.8	1.6
Experiment 5	Chain-Substitutive	36	6.5	5.6	6.6	-0.1	*13.5*	*12.2*	1.3	77.9	73.9	76.8	1.1	29.9	28.5	28.6	-1.3
	Chain-Additive	36	7.6	6.6	6.0	1.6†	*8.5*	*7.0*	1.5†	74.2	74.8	73.7	0.5	47.2	45.4	49.2	2.0
	Control-Substitutive	18	10.5	10.7	11.3	-0.9	1.7	1.4	0.3	77.2	79.8	79.3	-2.1	71.3	69.2	67.5	-3.8
	Control-Additive	18	9.6	8.9	9.5	0.1	0.8	0.9	-0.1	73.7	76.9	73.1	0.6	72.9	73.9	72.4	0.5
Experiment 6	Chain-Mechanism	36	6.0	5.9	5.8	0.2	*13.7*	*13.0*	0.7	80.3	79.9	79.7	0.5	30.3	29.0	30.4	-0.1
	Chain-No-Mechanism	36	7.9	6.5	5.8	2.0	*10.6*	*9.5*	1.0	77.4	73.8	73.4	3.9†	38.9	40.2	44.2	5.3
	Control	36	12.0	12.9	12.5	-0.5	2.0	2.1	0	77.4	78.9	78.4	-1.1	64.3	62.4	63.0	-1.3
Experiment 7	Chain-No-Mechanism-3Q	36								78.7	74.2	71.2	7.5*	38.8	44.0	63.8	**25.0***
	Control	36								77.6	78.3	76.1	1.6	69.2	68.6	69.2	0

Note. Difference scores are tested against 0. † $p < .10$. * $p < .05$.

Figure 1
Three-element causal chain.

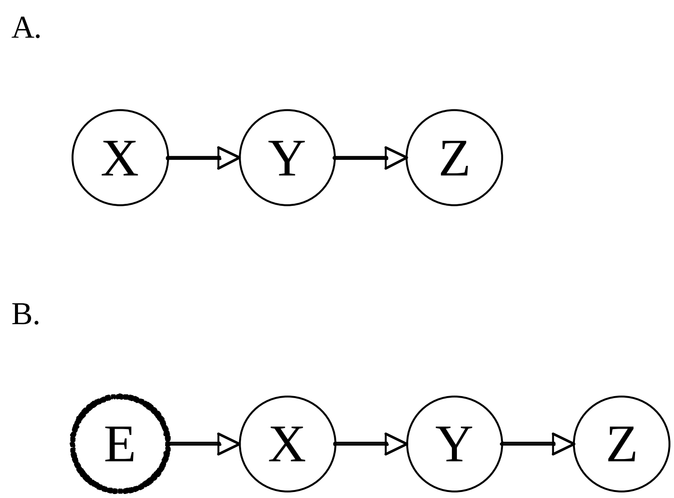

Figure 2
Previous results from (A) Ahn et al. (2000) and (B) Rehder and Kim (2006).

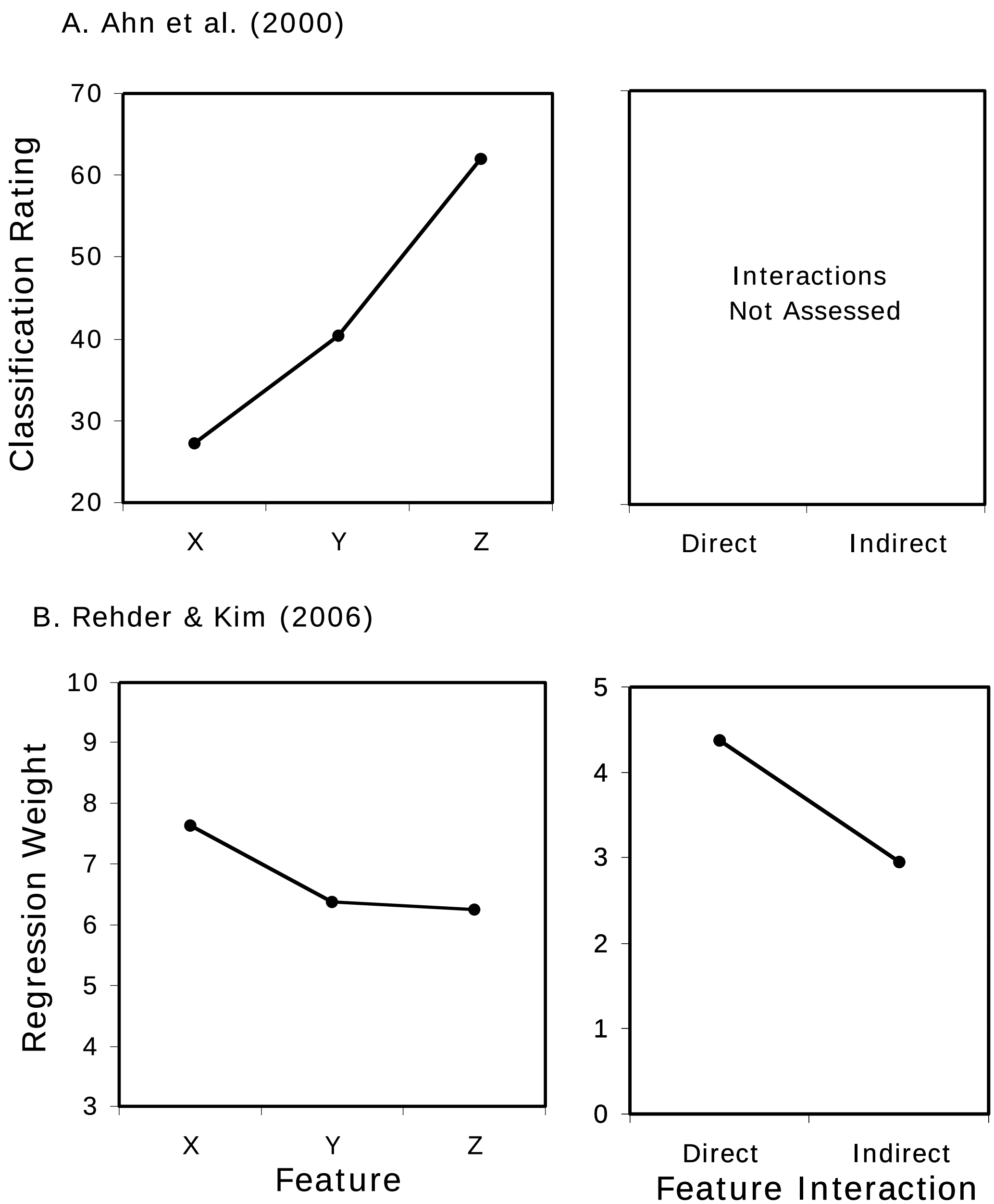

Figure 3

Predictions from two theoretical models. (A) The dependency model. (B) The generative model.

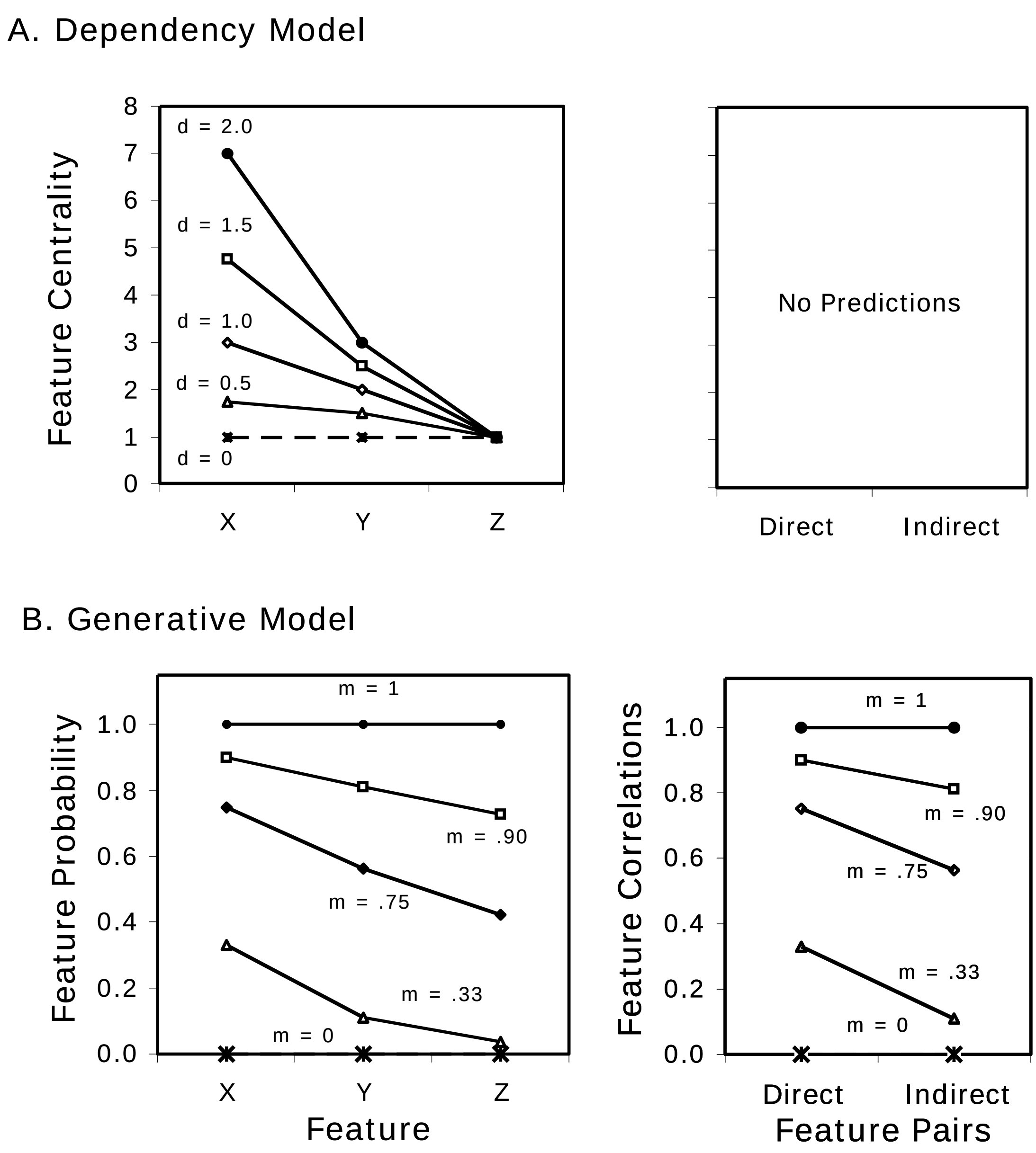

Figure 4

Causal structures tested Experiment 1. (A) Chain-100 condition. (B) Chain-75 condition. (C) Control condition.

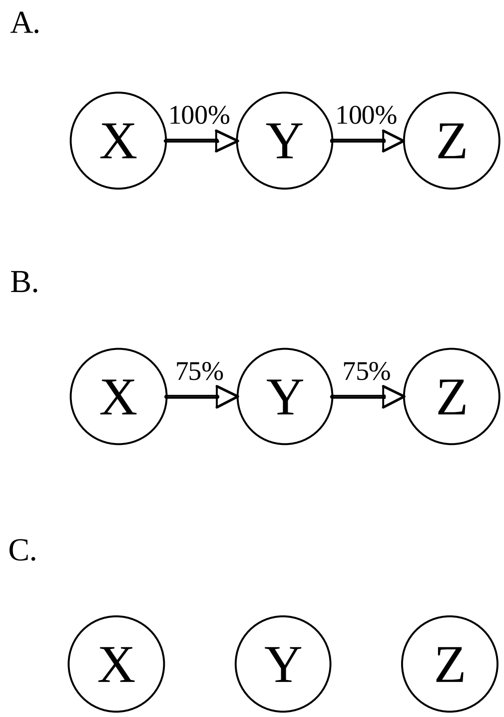

Figure 5

Classification results from Experiment 1. (A) Chain-100 condition versus Control condition. (B) Chain-75 condition versus Control condition. p_{lin} is the significance of the linear trend in each condition.

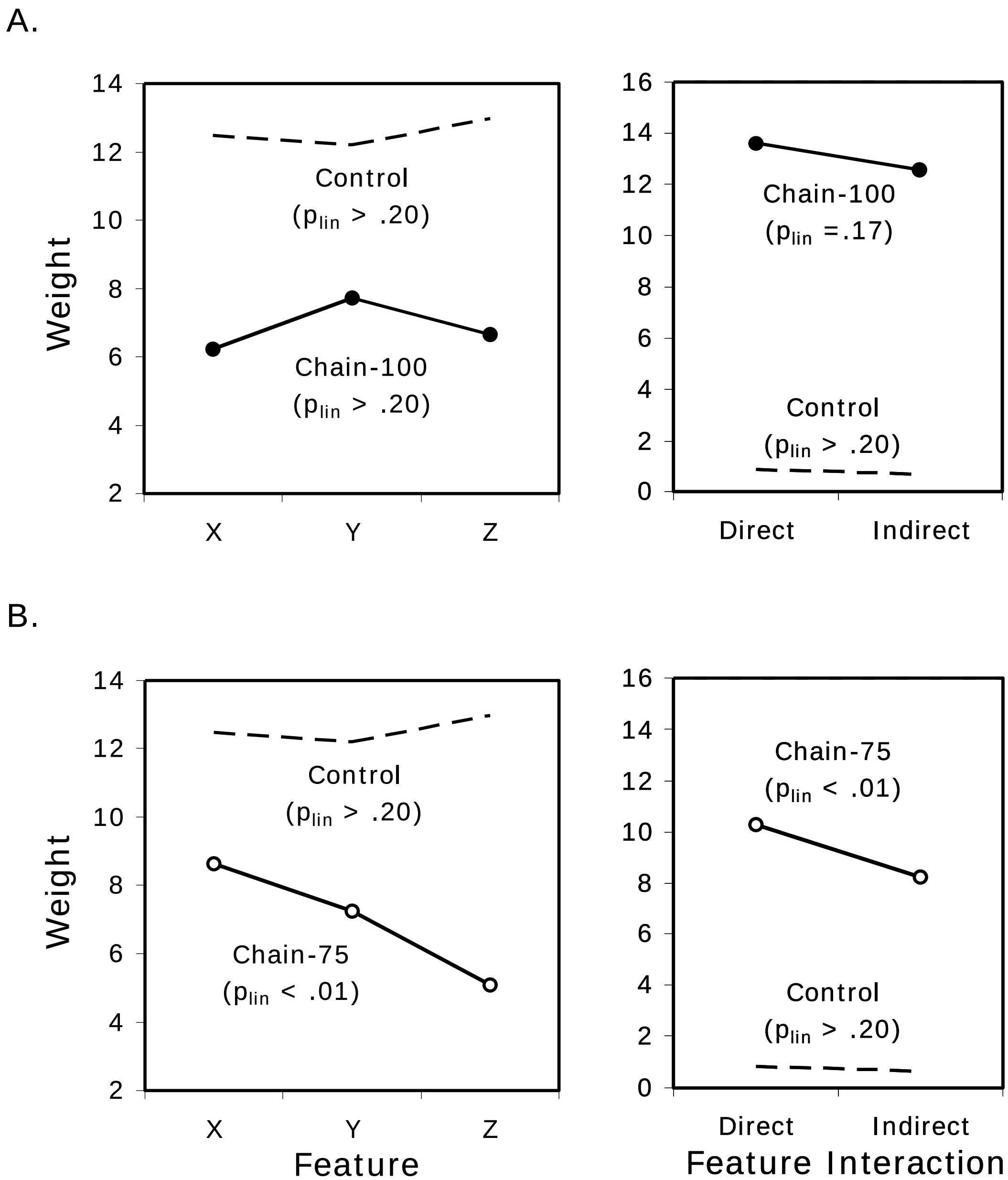

Figure 6

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

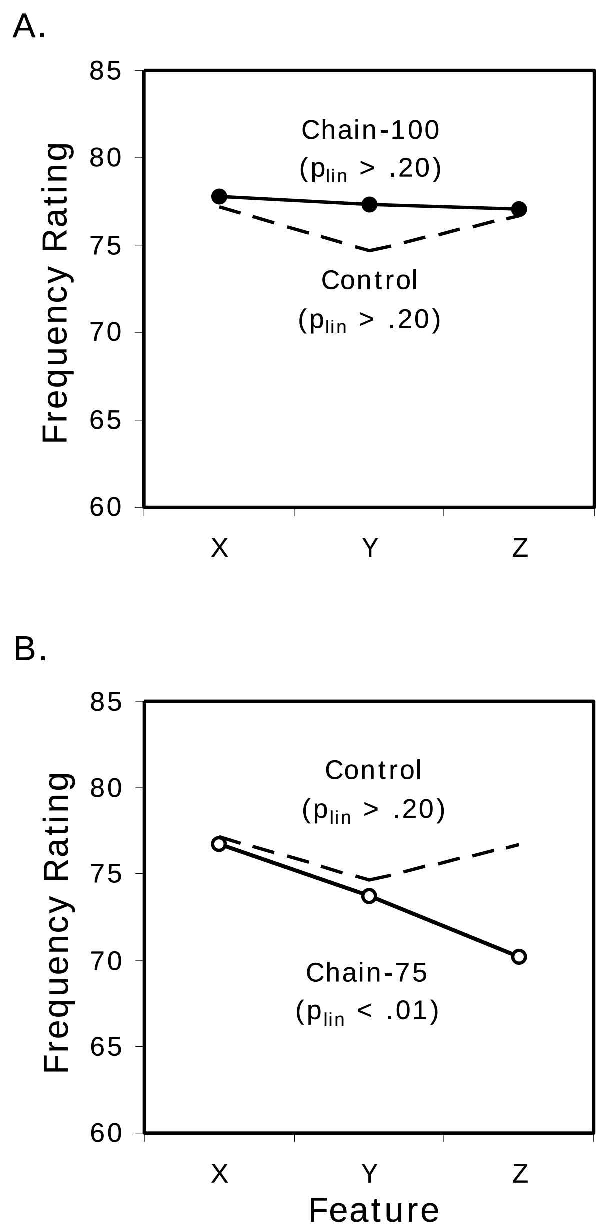

Figure 7

Classification ratings from Experiment 1. (A) For test items in the Chain-75 condition missing only feature X, only feature Y, and only feature Z. (B) For all test items as a function of their number of characteristic features.

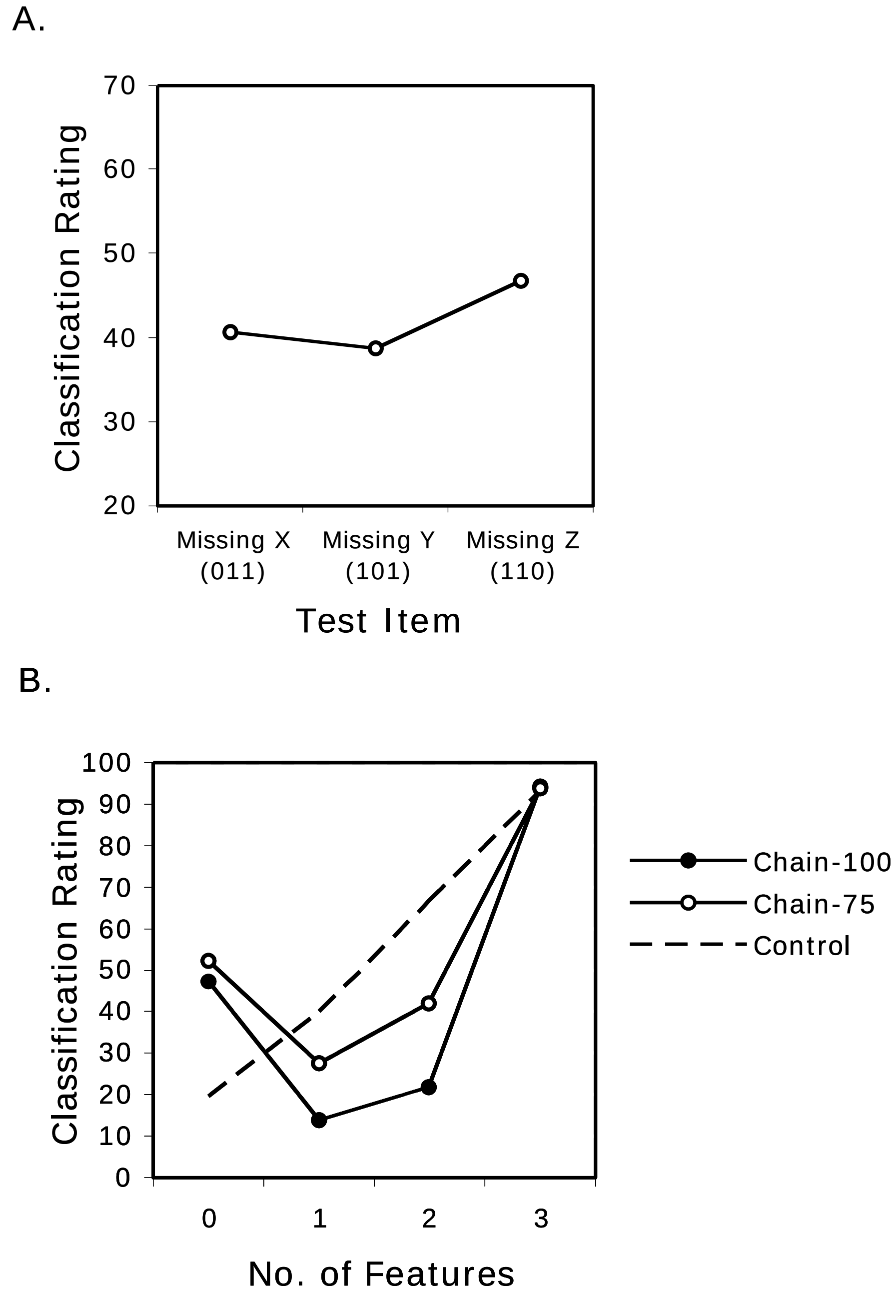

Figure 8

Causal structures tested Experiment 2. (A) Essentialized-Chain-80 condition. (B) Unconnected-Chain-80 condition. (C) Control condition.

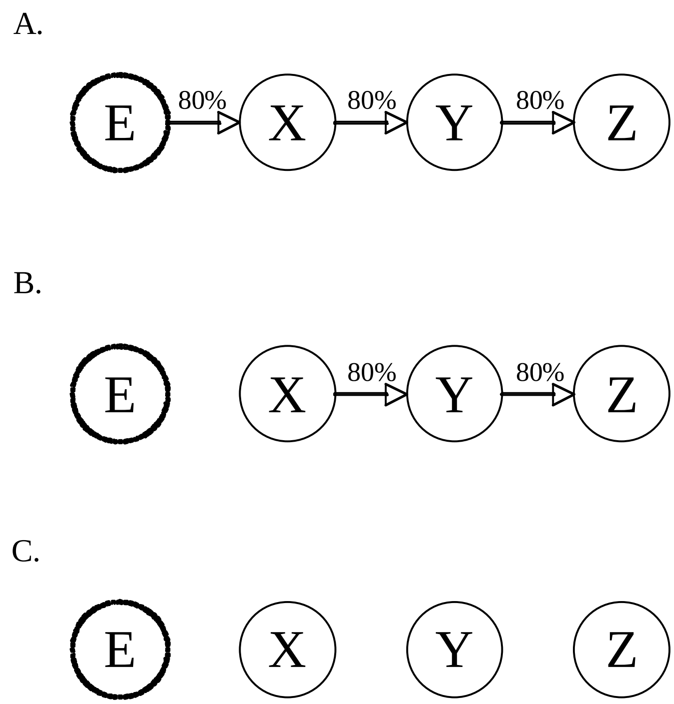

Figure 9

Classification results from Experiment 2. (A) Essentialized-Chain-80 condition versus Control condition. (B) Unconnected-Chain-80 condition versus Control condition. p_{lin} is the significance of the linear trend in each condition.

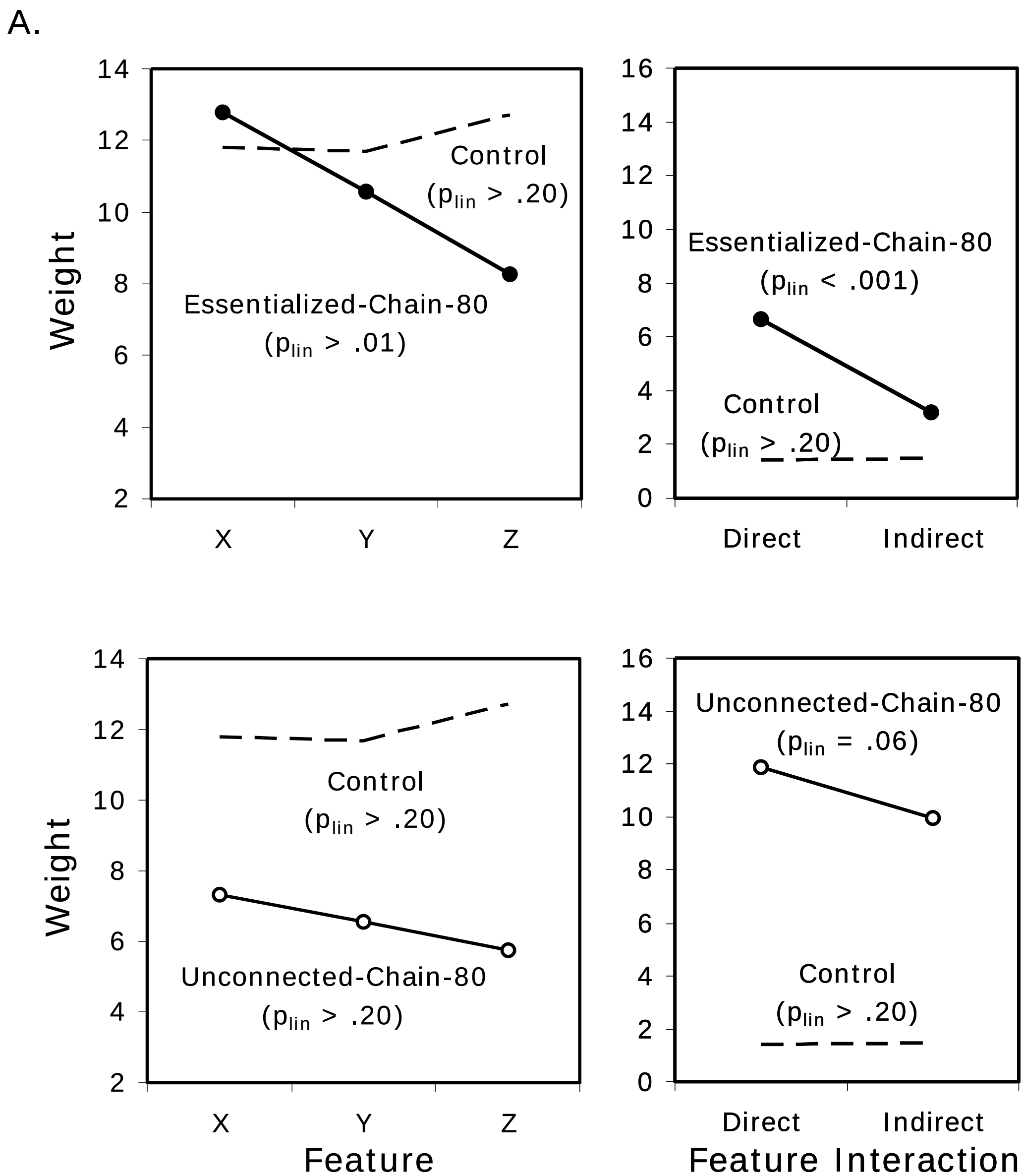

Figure 10

Feature frequency ratings from Experiment 2. (A) Essentialized-Chain-80 condition versus Control condition. (B) Unconnected-Chain-80 condition versus Control condition. p_{lin} is the significance of the linear trend in each condition.

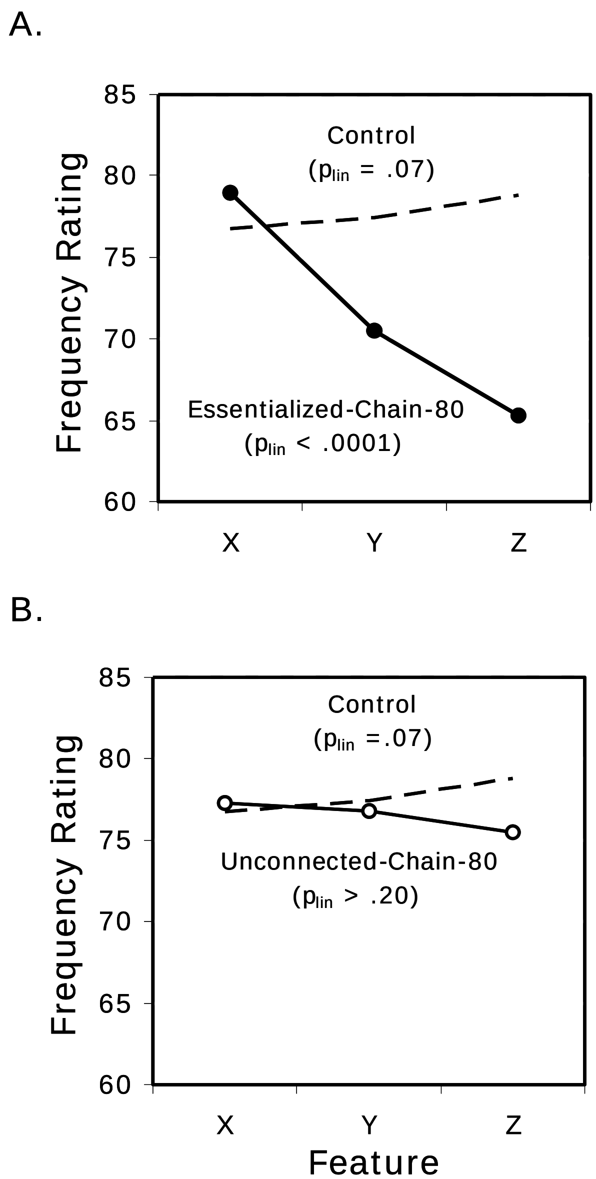

Figure 11

Classification ratings from Experiment 2. (A) For test items in the Essentialized-Chain-80 condition missing only feature X, only feature Y, and only feature Z. (B) For all test items as a function of their number of characteristic features.

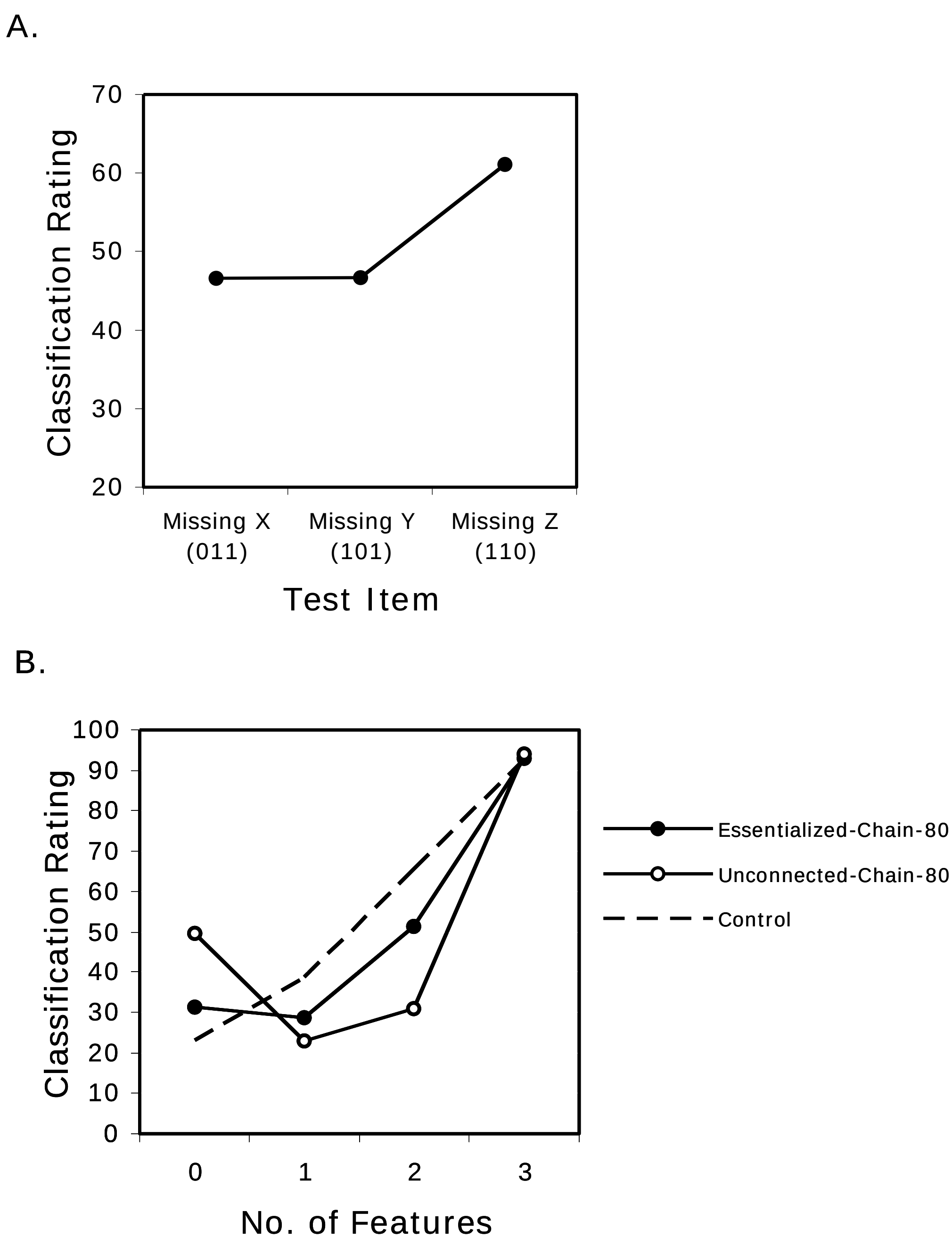

Figure 12

Classification results from Experiment 3. (A) Essentialized-Chain-100 condition versus Control condition. (B) Unconnected-Chain-100 condition versus Control condition. p_{lin} is the significance of the linear trend in each condition.

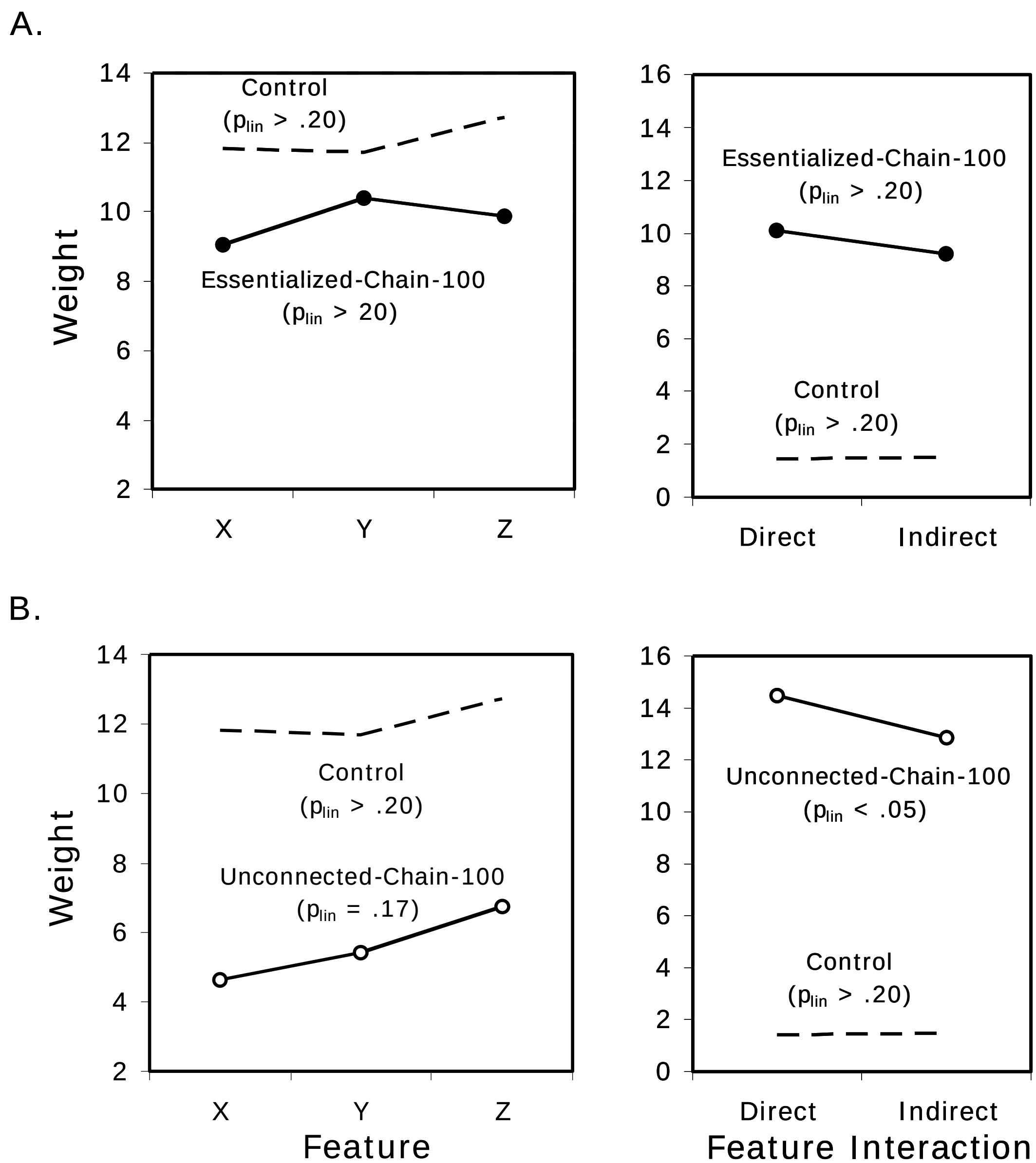

Figure 13

Feature frequency ratings from Experiment 3. (A) Essentialized-Chain-100 condition versus Control condition. (B) Unconnected-Chain-100 condition versus Control condition. p_{lin} is the significance of the linear trend in each condition.

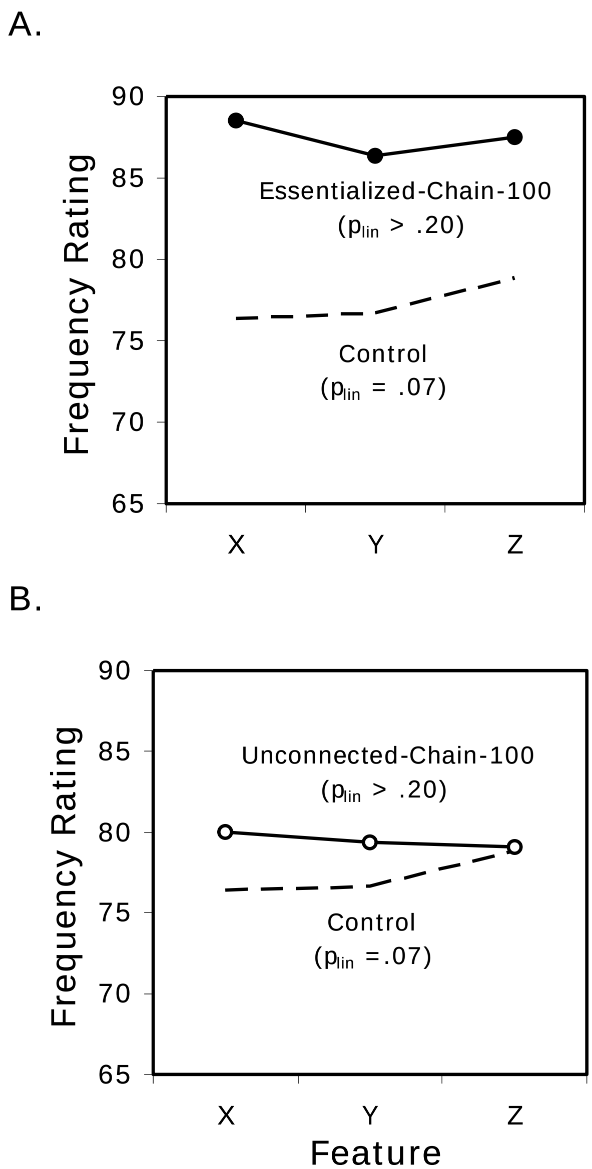

Figure 14
Classification ratings from Experiment 4-6 for all test items as a function of their number of characteristic features.

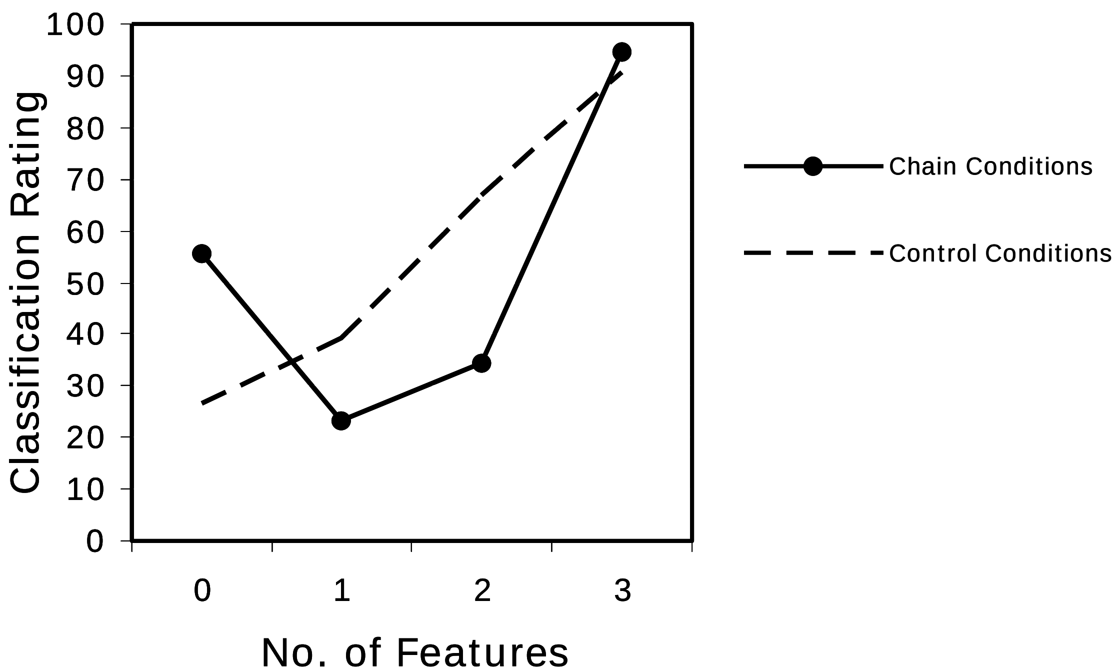

Figure 15

Experiment 7's classification ratings of test items missing only feature X, only feature Y, and only feature Z.