The Role of Causal Status versus Inter-Feature Links in Feature Weighting

Jessecae K. Marsh (jessecae.marsh@Yale.Edu)
Department of Psychology, Box 208205
New Haven, CT 06520 USA

Woo-kyoung Ahn (woo-kyoung.ahn@Yale.Edu)
Department of Psychology, Box 208205
New Haven, CT 06520 USA

Abstract

Studies have found that the causal status of features determines what exemplars are considered good members of a category (see Ahn & Kim, 2000). However, this causal status effect was questioned in recent studies (Rehder, 2003; Rehder & Hastie, 2001), because the preservation of causal links of a category's causal network was shown to play a significant role. We demonstrate in this study that these results are methodological artifacts arising from the use of unnatural wording of category attributes.

Introduction

Categories are believed to include not just a catalog of features, but also rich representations of the causal relations between features (e.g., Carey, 1985; Murphy & Medin, 1985). Recently, a number of proposals were made to specify the process of applying causal knowledge to categorization. In particular, two mechanisms have been proposed to describe how causal knowledge influences goodness-of-exemplar judgments (e.g., Ahn, 1998; Rehder, 2003; Sloman, Love, & Ahn, 1998). The goal of the current study is to re-assess the empirical support for each of these feature-weighting mechanisms.

One proposal states that causal knowledge indicates a feature's causal status, which, in turn, determines each individual feature's weighting. Ahn and colleagues have proposed in their causal status hypothesis that, with all else equal, features which cause other features in the same category are weighted more heavily than features that are effects of other features (e.g., Ahn, 1998; Ahn, Kim, Lassaline, & Dennis, 2000). Consider a hypothetical category with three features (X, Y, Z), which form a causal chain such that X causes Y, which causes Z (X→Y→Z). X has the highest causal status, Y has the next highest, and Z has the lowest causal status. Thus, the conceptual centrality, or importance to the concept, according to the causal status hypothesis, would be in the descending order of X, Y, and Z. For instance, upon learning that Roobans' eating of sweet fruits tends to cause Roobans to have sticky feet, which tend to allow them to climb trees, participants judged an instance missing "eating sweet fruits" to be the least likely member of Roobans, whereas an instance missing "climbing trees" to be the most likely member of the category Roobans (Ahn et al., 2000).

The second proposed mechanism states that whether the configuration of features in an exemplar is consistent with known causal relations of the category determines the exemplar's membership likelihood (e.g., Rehder & Hastie, 2001). For example, Rehder's causal model theory (2003) postulates that the likelihood that a category's causal network produces a given exemplar determines category membership for that exemplar. Exemplars that preserve a category's causal links would be viewed as better category members because they are more likely to be produced by the causal laws governing the category than exemplars that break such relationships. Using Rehder's (2003) example, an animal that does not fly and yet still builds nests in trees would be less likely to be judged to be a bird than an animal that does not fly and builds nests on the ground.

The two proposed mechanisms can make conflicting predictions for category membership. To illustrate this, let us consider again a category with X, Y, and Z, which forms a causal chain of X→Y→Z. Now consider two exemplars: one has Y and Z but not X (represented as `011`) and the other has X and Z, but not Y (`101`). The causal status hypothesis predicts that `101` is a better member than `011`, because X should be weighted more heavily than Y. However, `101` has two causal violations (X→Y and Y→Z did not occur), whereas `011` has only one causal violation (X→Y did not occur). Thus, the mechanism sensitive to preserving inter-feature causal links would predict that `011` is a better member than `101`.

Given that these two accounts for the role of causal knowledge on categorization can at times produce opposite predictions, it is important to understand which mechanism is more primary under what circumstances. Recent studies by Rehder and his colleagues (2003; Rehder & Hastie, 2001) found strong effects of inter-feature causal links in some of their experiments, to the extent that the causal status effect disappeared at times. (See the next section for details.) Such results can be taken to question the validity of the causal status hypothesis.

To the contrary, we argue that individual feature weightings would be more crucial than weighting determined by inter-feature relations. Consider a category with just two prototype features, X and Y, where X causes Y. There are four possible exemplars: `11`, `10`, `01`, and `00`. The proposals make two conflicting predictions. First, although `10` and `01` have the same number of causal violations, the causal status hypothesis predicts `10` to be a better category member than `01`. This prediction is based on essentialism: an essence is the deepest cause in a category and the absence of a cause feature (i.e., a feature that is closer to the essence in a causal network) signifies that the exemplar might have a different essence (Ahn et al., 2000). For example, an instance that flies without wings (`01`) seems to imply a causal mechanism or an essence entirely different from that of birds. Thus, exemplar `01` would be a worse bird than an instance that has wings but cannot fly (`10`). That is, not all causal violations are equal because a violation due to a missing cause suggests the presence of a different essence.

The second conflicting prediction concerns `00`. Rehder (2003) suggests that the fact that no causal relation was violated in `00` serves as positive support for the exemplar being a category member, increasing its membership likelihood. However, a Chihuahua, which does not have wings and does not fly (`00`), is a very poor "bird" although it preserves the expected causal relation of the category (i.e., not having wings causes a Chihuahua not to be able to fly.) If link preservation should figure more heavily than actual feature presence, then a Chihuahua should be judged as a more likely bird than a penguin, an exemplar that only has one category feature and thereby breaks the causal relationship instantiated in the category. In some sense, treating `00` as a good member of a category appears to be psychologically implausible, as is well illustrated in the famous raven paradox: seeing green grass does not boost our belief that all ravens are black (Hempel, 1945).

For these reasons, we argue that sensitivity to inter-feature causal links is not likely to play the predominant role in determining category membership. In the next section, we discuss why previous results showing the effect of inter-feature causal links appear to be due to the artificiality of stimulus materials used in those experiments. We will first describe Rehder's (2003) experiments in detail, which serve as the basic paradigm in the current study.

Causal Chain Structure

In Rehder (2003), participants first learned a category (see Figure 1, under "Ambiguous version" for the sample description for Kehoe ants). Participants in the causal chain condition also learned inter-feature causal relationships (F1→F2→F3→F4). After sufficiently studying the category, participants judged the membership likelihood of 16 possible exemplars that can be formed from 4 binary dimensions (e.g., `1111`, `1000`; see under "Item" in Table 1). Across three experiments, only partial support for the causal status effect was found, but the effect of inter-feature correlations was found to be consistently significant. For instance, `0000` had no causal violations and was judged to be a better category member than `0001`, `0010`, `0100`, and `1000`, each of which had at least one causal violation. Such results are at odds with the causal status effect (e.g., `1000` possesses a causally central feature and should therefore be rated higher than `0000`.)

We argue that these results were obtained because of the repetitive use of the wording "normal" for non-prototype values of features (see `0000` in Figure 1, "Ambiguous version"). There are two reasons why this wording might have inflated sensitivity to inter-feature relations.

First, the wording of "normal" could easily be misunderstood as speaking to a high prevalence of the feature within the category. Therefore, `0000` may have been rated as a better category member not because it preserves the links of the category's causal network, but because it is in all ways a normal category member. Indeed, close examination of Rehder's experiments indicates that the more ambiguous the meaning of the "normal" value becomes (e.g., when feature base-rates were not specified such that "normal" can be thought of as a more characteristic value), the weaker were the causal status effects (Rehder, 2003, Experiment 2).

Second, by repetitively using the same wording "normal" for `0` values for all four attributes, the inter-feature correlations (or the lack thereof) became artificially salient to participants. Consider the following 3 ways -- shown in 3 columns -- of representing a set of isomorphic exemplars.

0000	0000	2569
0010	0070	2579
0100	0800	2869
1011	3074	3574

Conventionally, `0`'s and `1`'s are used across all dimensions as shown in the first column so that readers can easily notice the correlated structure. In reality, however, a feature represented by `0` in one dimension (e.g., `0` could equal small for the dimension of size) is different from a feature represented by `0` in a different dimension (e.g., `0` could equal red for the dimension of color). The format used in the third column reflects these variations by using different numbers to represent four binary dimensions (e.g., `2` and `3` correspond to `0` and `1` for the first dimension, `5` and `8` to `0` and `1` for the second dimension). When such non-aligning feature values are used, the correlated structure is much more difficult to notice. Indeed, the sensitivity to feature correlations in concept learning is notoriously difficult to obtain (e.g., Malt & Smith, 1984; Murphy & Wisniewski, 1989; Ahn, Marsh, Luhmann, & Lee, 2002). Rehder's stimuli (2003) correspond to the second column by using the wording "normal" for all the non-prototype values (i.e., `0`'s). It is easy to see that the violation of the correlated pattern is more visible, compared to the third column. That is, the sensitivity to inter-feature causal links could have been unnaturally inflated in Rehder (2003; also in Rehder & Hastie, 2001) due to the unusual wording.

Common Cause and Common Effect

Rehder and Hastie (2001) taught participants common-cause and common effect structures (see Table 1 for illustration) rather than a causal chain structure. They found that in the common effect condition F4, which was described as an effect of F1, F2, and F3, was weighted much more heavily than its causes, arguing that these results countered the causal status hypothesis. The feature attributes used in Rehder and Hastie (2001) were identical to Rehder (2003), and therefore, suffer from the previously discussed wording

Ambiguous Version	Unambiguous Version
Initial Description (F1) Some Kehoe Ants have blood that is very high in iron sulfate. Others have blood that has normal levels of iron sulfate. (F2) Some Kehoe Ants have an immune system that is hyperactive. Others have a normal immune system. (F3) Some Kehoe Ants have blood that is very thick. Others have blood of normal thickness. (F4) Kehoe Ants build their nests by secreting a sticky fluid that then hardens. Some Kehoe Ants are able to build their nests quickly. Others build their nests at a normal rate.	**Initial Description** (F1) Kehoe Ants tend to have blood that is very high in iron sulfate. (F2) The immune system of Kehoe Ants tends to be hyperactive. (F3) The blood of Kehoe Ants tends to be very thick. (F4) Kehoe Ants build their nests by secreting a sticky fluid that then hardens, and they tend to build nests quickly.
0000 Blood that has normal levels of iron sulfate Normal immune system Blood of normal thickness Build nests at a normal rate	**0000** Blood very low in iron sulfate Underactive immune system Thin blood Build nests very slowly

Figure 1. Sample stimuli used in the current experiment. The Ambiguous version was used in Rehder (2003).

problems. For instance, the inter-feature correlations would have been unnecessarily salient by the use of "normal" values. However, there is an additional complication due to the uniqueness of the common effect structure. As acknowledged in Rehder and Hastie and demonstrated in Ahn and Kim (2000), a common effect structure (assuming a reasonably strong causal relationship) strongly implies that the base-rate of the common effect (F4) be higher than the base-rates of its causes (F1, F2, and F3): F4 is more likely because it can be produced in a multitude of ways. Even though the base-rates were initially specified to the participants to be 75% in these experiments, the base-rate implied by the given causal structure, coupled with the repetitive presentation of the confusing wording of "normal" values in the test exemplars, might have encouraged participants to overlook the provided base-rates. Instead, they might have utilized the base-rates strongly implied by the common effect structure.

Experiment

To summarize, Rehder and his colleagues presented a number of findings that apparently demonstrated sensitivity to inter-feature causal links and insensitivity to the causal status of individual features. We have argued so far that these results were obtained due to the use of the wording "normal" to represent non-prototype features. To demonstrate that the previous findings were artifacts of unnatural wording in the stimuli, the current experiment utilized two stimuli versions: an Ambiguous version, in which the absence of a feature was described as being "normal" as in Rehder (2003), and an Unambiguous version, in which the absence of a feature was described as being the opposite state of the prototype feature. Figure 1 shows sample stimuli. Causal chain, common cause, and common effect structures were examined.

Materials and Design

Six categories and their causal structures taken from Rehder (2003; Rehder and Hastie, 2001) were used. In the cover story of each category, prototype features of the category (F1, F2, F3, and F4) were described. In the Ambiguous version of the materials the feature descriptions were the same as the ones used in Rehder's studies in that each feature was described as taking two possible forms: a "normal" occurring form and a variant of that form. In the Unambiguous version only the prototype value was given. As in Rehder's Experiment 2 (2003), which produced the strongest results against the causal status hypothesis, information about the base-rates of the characteristic features was not provided to the participants in either version of the stimuli.

Each cover story also described the causal relationship between the features of the category. The causal relationships could take one of three structures: F1 caused F2, F3, and F4 (common cause condition); F1, F2, and F3 each caused F4 (common effect condition); or F1 caused F2, which caused F3, which caused F4 (causal chain condition).

Sixteen test exemplars (all possible combinations of the four binary attributes; see Table 1) were created for each category. In the Ambiguous version, the absence of a feature was described as the instance possessing a normal amount of the given feature. In the Unambiguous version, the absence of a feature was described as the opposite state for the appropriate feature. (See 0000 of each version in Figure 1 for examples.) Note that using the opposite values in the Unambiguous version can result in stronger causal violations. For instance, upon learning that blood high in iron sulfate causes a hyperactive immune system, an exemplar that contains blood with <u>normal</u> levels of iron sulfate and a hyperactive immune system would be less of a violation than an exemplar that contains blood <u>low</u> in iron sulfate and a hyperactive immune system. Thus, if people value causal violations in membership judgments, the Unambiguous condition is more likely to show this effect.

Procedure

At the beginning of a block participants were given a short set of instructions that included the name of the category for which they were about to make judgments. Participants then proceeded to make judgments for the sixteen exemplars corresponding to that category. The top half of each exemplar screen presented the cover story for the category, the description of the four category features, and the outline of the causal relationship between the features. In the bottom half of the screen the information pertinent to the exemplar for that trial was presented. First, the question "Is this a [X]?" with the X replaced by the appropriate category name was listed, followed by the four features found in that trial's exemplar. Participants answered the question by entering on a keyboard a number on a scale from 0 (Definitely not an X) to 20 (Definitely an X). If a participant neglected to provide a rating on an exemplar, the missing data was replaced in the statistical analyses with the mean value for that exemplar in that causal structure and version[1]. All 16 test items pertaining to the same category (i.e., same causal structure) were presented in the same block in a randomized order. Upon completing one block, participants proceeded to two more blocks, each corresponding to a different category with a different causal structure.

Compared to Rehder (2003), one discernible procedural difference was that causal background information was available to participants during judgments, saving the need for extensive training of category information. This allowed participants to make judgments about three different categories in three different causal structures without the fear of confusing what features belonged to a category. As shown below, the results from the Ambiguous version replicated Rehder (2003) and Rehder and Hastie (2001), ensuring that this procedural difference was not critical.

Counterbalancing the different orders of causal structure and material type resulted in six experimental sequences. Participants completed three blocks worded either in the Ambiguous version (N=36) or the Unambiguous version (N=36). Participants proceeded at their own pace. All experimental procedures were conducted on an eMac using the RSVP experimentation package. Yale undergraduates completed the experiment either for pay or partial fulfillment of a course.

Results

We will first describe the results from the causal chain structure, followed by the common cause and common effect structures. The mean ratings for each exemplar in each causal structure are presented in Table 1, separated by version. To determine the impact of each individual feature and inter-feature link on membership judgments, regression analyses were completed, following Rehder and Hastie (2001)[2]. Regression weights were calculated for each individual participant within each causal structure type, and the average weights are shown in Table 2.

Two general patterns emerged across the three causal structures. First, greater weight was placed on individual feature presence than link preservation for the Unambiguous condition as compared to the Ambiguous condition, whereas the three relevant links for a given causal structure were weighted more heavily in the Ambiguous version than in the Unambiguous version for all causal structures. Second, the causal status hypothesis was supported in the Unambiguous condition but not in the Ambiguous condition.

Table 1. Average ratings across the three causal structures. Significant differences between the Unambiguous (Un) and Ambiguous (Amb) versions are shaded in gray.

	Causal Chain		Common Cause		Common Effect	
	F1→F2→F3→F4		↗ F2 F1 → F3 ↘ F4		F1 ↘ F2 → F4 F3 ↗	
Item	Un	Amb	Un	Amb	Un	Amb
0000	7.1	71.0	8.1	71.8	3.1	69.6
0001	10.6	42.2	16.3	52.1	14.6	25.1
0010	14.6	37.9	17.8	49.8	13.1	47.5
0100	15.1	36.6	17.9	54.1	16.8	51.0
1000	23.1	41.5	21.1	29.3	17.1	49.3
0101	24.3	31.8	30.0	43.1	29.3	52.6
0110	27.0	40.3	28.1	47.5	32.7	40.1
0011	27.2	48.2	29.7	43.7	28.1	56.8
1010	35.4	35.7	35	41.4	32.9	40.4
1001	35.6	31.8	33.9	40.6	36.0	54.2
1100	39.7	46.4	39.9	48.8	42.7	40.0
0111	50.8	45.8	45.9	40.1	62.4	73.1
1011	54.6	43.6	59.3	59.0	59.6	69.7
1101	60.4	47.9	64.6	61.9	67.9	69.9
1110	68.9	51.7	64.0	61.5	60.3	34.0
1111	91.3	85.3	87.6	89.6	91.1	88.3

Causal Chain

As shown in Table 1, the two versions led to quite different patterns of results in the Causal Chain condition. For instance, the mean ratings for 0000 were 71.0 in the Ambiguous version, but only 7.1 in the Unambiguous version. This difference is most likely due to the fact that in the Ambiguous version, 0 values were described as "normal," making participants think that these feature values had high base-rates in the category. Note, however, that this inflated rating on 0000 in the Ambiguous condition might have made it look as if participants were highly sensitive to causal violations. For instance, in this condition, 0000 was rated to be much higher than 0001, 0010, 0100, and 1000, each of which has at least one causal violation. However, this was not the case in the Unambiguous condition: 0000, although preserving all known causal relations, was rated lower than any of the other fifteen exemplars in the category.

[1] For the 3456 judgments collected across all participants, 22 were replaced using this method.

[2] For each exemplar, four dummy variables representing the four features were coded as +1 if the prototype value was present and -1 if absent. Six dummy variables representing the possible links between pairs of features were coded as +1 if the two features involved in the link were jointly present/absent and coded as -1 if only one of the features was present. Table 2 shows regression weights on the relevant causal links only.

In addition, causal status effects are present in the Unambiguous condition, but not in the Ambiguous condition. For the causal chain model, the causal status hypothesis predicts that the most causally central feature (i.e., F1) should be most heavily weighted, with a decrease in weighting as centrality decreases. In the Unambiguous condition, the mean ratings of `0001`, `0010`, `0100`, and `1000`, as well as the mean ratings of `0111`, `1011`, `1101`, and `1110` are in the ascending order as predicted by the causal status hypothesis. Indeed, regression weights for individual features from the Unambiguous version were consistent with this prediction (Table 2). Individual planned t-tests were carried out comparing the weight given to the individual features of the causal chain. As predicted, the regression weight for F1 was higher than any of the other features (all t's > 2.4; all p's < .02). F2 was not weighted more heavily than F3 (p = .70), but was weighted more heavily than F4 (t(35) = 2.52, p < .02). Likewise F3 was weighted more heavily than F4 (t(35) = 2.41, p < .03).

No such pattern was found in the Ambiguous version. None of the individual feature weights were significantly different from each other (all p's > .6). Instead, more weight was allotted to the preservation of inter-feature links than to feature presence. The regression weights for each of the features (F1 through F4) for the Unambiguous version were significantly higher than those for the Ambiguous version (all p's < .0001). However, the weights for link preservation tend to be higher in the Ambiguous version than in the Unambiguous version: weights for the link between F1 and F2, as well as the link of F3 and F4 were significantly higher in the Ambiguous version (t's > 3.64, p's < .001) the link between F2 and F3 was not significantly different between the two conditions (p = .46).

Common Cause and Common Effect Structures

Rehder and Hastie (2001), after examining categorization based on the common cause and common effect structures, concluded that their results did not support the causal status hypothesis. They stated, "In this study, however, the fact that the common cause was a cause did not confer it any additional importance above and beyond its centrality *[defined as participation in many causal relationships]*, as indicated by the fact that across Experiments 1-3 the average transfer categorization regression weight given to the common cause (11.7) was slightly less than the average weight given to the common effect (12.2)" (p. 349-350; phrase in brackets added). Likewise, in the Ambiguous condition of the current study, the average regression weight given to the common cause (F1: 1.85) was lower than the weight given to the common effect (F4: 6.56) (t(35) = 2.49, p < .02), replicating their results. More specifically, `1110` in the common effect structure lacks the effect feature, but in the Ambiguous version was given a mean rating that was much lower (34.0) than other exemplars missing only one cause feature (`0111`, `1011`, `1101`, M=73.1, 69.7, 69.9, respectively; all t's > 6.1, all p's < .0001). Furthermore, `1110` in the Ambiguous version of the common effect structure received somewhat lower ratings compared to `0111` in the Ambiguous version of the common cause structure (M=40.1), even when they both involve breaking three links with the only difference being the causal status of the missing feature.

In the Unambiguous version, however, the pattern changed. The regression weight for F4 in the common effect structure was somewhat lower (11.2) than F1 in the common cause structure (13.2), reversing the direction of regression weights reported in Rehder and Hastie (2001). As discussed earlier, the common effect structure works against the causal status effect because three independent causes are capable of producing the same effect, and therefore, the terminal effect of this network is more likely to occur than any of its three causes. Despite this disadvantage, when the Unambiguous version was used, `0111` in the common cause structure (i.e., missing the common-cause feature) received much lower ratings (M=45.9) than `1110` in the common effect structure (i.e., missing the common effect feature, M=60.3), t(35) = 3.08, p < .005.

In both common cause and common effect structures, as in the causal chain structure, the regression weights on inter-feature links were much higher in the Ambiguous condition than the Unambiguous condition (all t's > 3.29, all p's < .005), whereas the weights on individual features were much higher in the Unambiguous condition than in the Ambiguous condition (all t's > 3.23, all p's < .005).

Table 2. Average regression weights for the three causal structures. (Un=Unambiguous; Amb=Ambiguous; L=Link, e.g., L12=link between F1 and F2; - denotes irrelevant links

	Causal Chain		**Common Cause**		**Common Effect**	
	F1→F2→F3→F4		↗ F2 F1 → F3 ↘ F4		F1↘ F2→F4 F3 ↗	
	Un	Amb	Un	Amb	Un	Amb
F1	14.5	1.88	13.2	1.85	13.0	1.88
F2	10.6	2.12	9.80	3.68	12.4	2.27
F3	10.2	1.65	8.59	1.46	8.95	3.19
F4	7.12	1.78	8.36	2.10	11.2	6.56
L12	3.37	7.72	3.55	7.75	--	--
L23	3.12	4.00	--	--	--	--
L34	2.01	6.19	--	--	2.13	8.37
L13	--	--	2.03	7.64	--	--
L14	--	--	2.52	6.43	1.05	8.15
L24	--	--	--	--	2.63	6.76

Discussion

In the current study, we tested an Unambiguous version where prototype features were indeed described to be typical of the categories and non-prototype features did not share the same wording across features, thereby avoiding unnaturally emphasizing correlations among features. Using these more realistic stimuli, we found that participants' membership likelihood judgments were based more on individual features' weighting as determined by causal status than the preservation of inter-property links. The support for the causal status effect was found even in the common effect structure where the causal structure strongly implied base-rates of features that could work in opposition to the causal status effect.

Rehder (2003) claimed that the lack of clear causal status effects in his experiments can be attributed to "participants' not linking the root of the causal chain to the category's underlying causal mechanisms" and that explicit linking to an essentialist cause was necessary to producing the effect (p. 1155). Our results show that with more natural stimuli as in the Unambiguous version, additional reinforcement in essentialist beliefs was not necessary to obtain the causal status effect.

As mentioned earlier, we believe that the Unambiguous version was successful in producing causal status effects for two different reasons. First, the Unambiguous version indicated that the prototype features were indeed typical of the categories (i.e., possessed higher base-rates.) Thus, 0000 was judged to be the worst member even though no causal violation occurred. Second, by not using the same wording "normal" for non-prototype features, we did not inflate the salience of the correlated structures to the participants. We believe both aspects of this manipulation are critical for the following reasons.

If base-rate alone drove the observed differences between the two versions, then the results of the current experiment's Unambiguous version would be predicted to replicate the high base-rate manipulation used by Rehder in several versions of his experiments (e.g., 2003, Experiment 1). In these high base-rate conditions participants were told that the causally active form of the feature was present in 75% of category members. Unlike the current Unambiguous condition, the results from this base-rate specification found only partial support for the causal status hypothesis. We argue that these results occurred because even with the high base-rate information the non-prototype values were still described as "normal" in the test exemplars, highlighting the violation of the correlated category structure to the participants.

Rehder and Hastie (2001) argued, "our more fundamental criticism of proposals such as the causal-status hypothesis and feature mutability is that they only consider the effect causal knowledge has on the importance of individual attributes. In contrast, a central contribution of the current article is to demonstrate that causal knowledge affects which combinations of attributes become acceptable to category membership" (p. 350). While we are not entirely denying the role of preserving inter-feature links on feature centrality, the current study demonstrates that a strong case for the role of inter-feature links has yet to be made using stimuli that reflect the way in which natural categories are structured. When the correlated structure of categories was not artificially emphasized, and features used in causal links were described as prototype features of a category, it was found that individual features' causal status, rather than feature combinations, was the predominant determinant of feature weighting.

Acknowledgments

The current research was supported by a National Institute of Mental Health Grant (2 R01 MH57737-05A1) given to the second author. We would like to thank Bob Rehder for helpful comments on an earlier draft.

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., & Kim, N. S. (2000). The causal status effect in categorization: An overview. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation, Vol. 40*. San Diego, CA: Academic Press.

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

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

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Plenum.

Hempel, C. G. (1945). Studies in the logic of confirmation. *Mind, 54*, 1-26, 97-121.

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

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications*. Chichester, U.K.: Ellis Horwood.

Rehder, B. (2003). 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). Casual knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130*, 323-360.

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