Memory & Cognition
1999, 27 (6), 1008-1023

Effect of causal structure on category construction

WOO-KYOUNG AHN
Yale University, New Haven, Connecticut

In four experiments, the question of how the causal structure of features affects the creation of new categories was examined. Features of exemplars to be sorted were related in a single causal chain (causal chain), were caused by the same factor (common cause), or caused the same effect (common effect). The results showed that people are more likely to rely on common-cause or common-effect background knowledge than on causal-chain background knowledge in category construction. Such preferences suggest that the common-cause or the common-effect structures are considered more natural conceptual structures.

Studies on concept formation have mainly focused on the learning of categories that are preconstructed, either by society or by the experimental situation. A child, for example, might learn concepts of dogs and of cats when his or her mother (or father) points to them and labels them. People also learn concepts by creating categories of their own. For example, after several semesters of teaching at a college, a professor might create his or her own idiosyncratic categories of undergraduate students (e.g., those who constantly put effort into the course vs. those who work only when they like the subject matter). In creating new categories, people usually bring in their background knowledge. The professor in the above example might believe that the more effort students put into a course, the higher their grades tend to be. The present study examines how such causal background knowledge affects the construction of new categories. The following section briefly reviews previous studies on category construction.

Unidimensional Sorting Bias

The general consensus is that existing natural categories do not have defining features that are singly necessary and jointly sufficient. Rather, exemplars in the same category are generally similar to each other, and exemplars in different categories are dissimilar to each other—the so-called *family resemblance* (FR) principle (Rosch & Mervis, 1975; E. E. Smith & Medin, 1981). However, when participants are asked to create new categories from given exemplars, they show a strong bias toward sorting exemplars on the basis of a single dimension and creating categories with defining features (unidimensional, or 1-D, sorting, henceforth), rather than creating categories with FR structure (FR sorting, henceforth; Ahn, 1990b; Ahn & Medin, 1992; Imai & Garner, 1965; Medin, Wattenmaker, & Hampson, 1987; Regehr & Brooks, 1995).

For instance, participants in Medin et al. (1987) received 10 exemplars, abstractly described in Figure 1. The 10 exemplars (e.g., E1, E2, etc.) consisted of four dimensions (D1, D2, D3, and D4), each with two values (e.g., 0 and 2 for D1). For example, D1 might be a color dimension, and 0 on D1 might stand for *yellow*.[1] These exemplars were developed by first starting out with two prototypes (E1 and E6) and by generating four distortions of each prototype (E2 through E5 from E1, and E7 through E10 from E6) by crossing a value from the contrast category.

In Medin et al.'s (1987) experiments, participants were asked to create two categories with these 10 objects. According to Rosch and Mervis' (1975) FR measure, people given this set of exemplars should create categories by grouping E1 through E5 together and E6 through E10 together. However, no participant in this study created an FR structure. Rather, they selected a single most salient dimension (e.g., D1) and grouped all exemplars with the same value on the dimension in the same category (e.g., E1, E2, E3, E4, and E10 in one category and E5, E6, E7, E8, and E9 in the other category).

The robustness of the 1-D sorting bias has been demonstrated under many instructional variations and across many different types of instantiations of the stimulus. Ahn and Medin (1992) used the exemplar structure shown in Figure 2, but again failed to elicit FR sorting. L.B. Smith (1981) predicted that increasing the number of dimensions would encourage participants' to pay more attention to overall similarity, but no such evidence was found, even with eight dimensions (Ahn, 1990a; Regehr & Brooks,

This project was supported partly by National Science Foundation Grant NSF-SBR 9515085 and partly by National Institute of Mental Health Grant RO1 MH57737, awarded to the author. I thank Dorrit Billman, Martin Dennis, Mark Keane, Douglas Medin, Tom Spalding, Michael Waldmann, and two anonymous reviewers for their helpful comments on this research. I also thank Marvin Chun, Kevin Lee, Nancy Kim, and Heidi Wenk for their thorough comments on earlier drafts of this paper and Tisha Baldwin, Loranel Graham, Denise Hatton, and Yani Indrajana for collecting data for the experiments reported here. Correspondence concerning this article should be sent to W.-K. Ahn, Department of Psychology, Vanderbilt University, Wilson Hall, Nashville, TN 37240 (e-mail: woo-kyoung.ahn@vanderbilt.edu).

Copyright 1999 Psychonomic Society, Inc.

	D1	D2	D3	D4		D1	D2	D3	D4
E1	0	0	0	0	E6	2	2	2	2
E2	0	0	0	2	E7	2	2	2	0
E3	0	0	2	0	E8	2	2	0	2
E4	0	2	0	0	E9	2	0	2	2
E5	2	0	0	0	E10	0	2	2	2

Figure 1. Exemplar structure used in Medin, Wattenmaker, and Hampson (1987) and in Experiment 3.

1995). In another attempt to produce FR sorting, Regehr and Brooks used "holistic blobs," where the features were portions of the total shape rather than clearly demarcated features, such as size and pattern. Still, few participants produced FR sorting. After failing to obtain FR sorting with numerous variations on type of dimension, they finally obtained FR categories by using a "match-to-standards" procedure in which participants received two prototypes serving as standards and a stack of cards to be matched to these standards. But when the same participants were told to sort all of the cards again without using this procedure, they reverted to 1-D sorting.

Sorting in Knowledge-Rich Domains

The studies cited above used stimulus materials where participants had no knowledge about how features were related. However, real-world concepts have complex interproperty relationships (see, e.g., Carey, 1985; Keil, 1989; Murphy & Medin, 1985). For example, birds can fly because they have wings, cars can move because they have an engine and wheels, and so forth. Recently, there has been a surge of studies demonstrating the effect of background knowledge on categorization (e.g., Keil, 1989; Medin & Shoben, 1988; Murphy & Allopenna, 1994; Nakamura, 1985; Pazzani, 1991; Spalding & Murphy, 1996; Wattenmaker, 1995). However, few studies have articulated in detail the nature and role of domain theories. In general, it is suggested that theories include causality as a central and primitive (although not exclusive) component (Carey, 1985; Gelman & Kalish, 1993; Wellman, 1990). The goal of the present paper is to investigate how different types of causal structure in background knowledge affect category construction, rather than to merely demonstrate the effect of background knowledge in general.

Given any two features, there is one possible causal structure, in which one feature causes another.[2] Given three features, there can be three types of causal structure: common-cause structures (A causes B, A causes C; e.g., different symptoms present in a patient are caused by the same virus), common-effect structures (A causes B, C causes B; e.g., different personality traits of a person elicit the same response from the observer), and causal-chain structures (A causes B, which causes C; e.g., a bird has wings, which allows it to fly, which allows it to build a nest in a tree). The present study directly compares the effects of these three types of causal structure on category construction. The aim is to go beyond a mere demonstration of the background knowledge effect and to closely examine differential effects of background knowledge as a function of causal connectivity among features. If people spontaneously rely on one type of causal structure more

	D1	D2	D3	D4		D1	D2	D3	D4
E1	0	0	0	0	E6	2	2	2	2
E2	0	0	0	1	E7	2	2	2	0
E3	0	0	2	0	E8	2	2	1	2
E4	0	1	0	0	E9	2	0	2	2
E5	2	0	0	0	E10	1	2	2	2

Figure 2. Exemplar structure used in Ahn and Medin (1992) and Experiments 1 and 2.

than on others in creating new categories, such findings can also suggest what kind of causal structure in concepts people consider to be more *natural*. The following sections review previous studies in relation to these three causal structures.

Effect of common-cause and common-effect knowledge. Medin et al. (1987) examined category construction when participants had background knowledge about how surface dimensions are connected to a theme. The exemplar structure was the same as that in Figure 1. This time, the dimensions were personality traits that were instantiations of introvertness or extrovertness. For example, consider the following descriptions of five people taken from this study:

Carrie: outgoing, energetic, entertaining, bold;
Susan: outgoing, energetic, entertaining, daydreamer;
Olivia: outgoing, energetic, inhibited, bold;
Felicity: outgoing, self-conscious, entertaining, bold;
Wendy: sad, energetic, entertaining, bold.

Although there is no single feature that all five people share, the participants in Medin et al.'s experiment overcame the 1-D sorting bias and grouped these five people into one category because they tend to be *extroverted*. Since there is no defining feature at the surface level, the participants were considered to have created FR categories. One can think of extrovertness as a common cause for the surface features and consider this study as evidence showing the effect of common-cause structure on FR sorting.

Although Medin et al.'s (1987) study is one of the pioneering demonstrations of the background knowledge effect, the exact underlying mechanism is unclear, for a number of reasons. One such reason is that this study did not use strict dimensional distinctions. For example, one dimension had *outgoing* and *sad* as its values, and another dimension had *energetic* and *self-conscious* as its values. However, *outgoing* and *self-conscious* can be two different values for a single dimension as well. As a result, it is not clear whether the background knowledge effect in Medin et al. was derived from this unclear distinction between dimensions or from an actual knowledge of common cause. A related point is that the study relied on people's existing background knowledge. Although there is merit to such demonstrations, it is not clear whether the effect was due to the presence of background knowledge per se or to the particular set of features used. In addition, if the effect was due to the background knowledge, it is not clear exactly what aspect of the background knowledge led to the effect.

Ahn (1990a, 1991) used experimentally provided background knowledge and showed that FR sorting can be obtained from the set used in Figure 1, when the features share the same underlying theme. The stimuli used in these studies can be thought of as examples of common-cause and common-effect structures. As an example of a common-effect structure, the participants heard that flowers with certain values on four dimensions were known to attract Trooder bees and that flowers with certain other values were known to attract Champin birds. As an example of a common-cause structure, the participants heard that an external virus tended to cause certain types of symptoms along four dimensions, whereas genetic components tended to cause other types of symptoms. In both cases, knowing this background information led to more FR sorting, as compared with the control group, who did not have this background knowledge. However, a direct comparison between the common-cause and the common-effect structures is not feasible from Ahn's studies, because different stimulus materials were used for the two structures.

Effect of causal-chain knowledge. Spalding and Murphy (1996) suggested another type of background knowledge that might elicit FR sorting. They pointed out that "all of the features [in Ahn (1990a, 1991)] are linked only to the theme but are not linked to each other.... This seems very unlike features of natural categories. Consider some features of birds, like 'wings' and 'flies.' It seems that there is a direct relation between these features in which the presence of 'wings' helps to support or allow 'flies"(p. 527). They claim to have demonstrated that FR sorting increased when the background knowledge revealed direct links among the features, thereby allowing features to be integrated. For example, a car's features that can be easily integrated (e.g., *made in Norway, heavily insulated, white, drives on glaciers,* and *has treads*) led to more FR sorting than those that are difficult to integrate (e.g., *green, manual transmission, radial tires, air bags,* and *vinyl seat covers*), even when the formal structure of the features was identical for the two conditions. At this point, however, it is not yet clear whether these results are solely due to the effect of a causal-chain structure. When Spalding and Murphy used the term *direct relations* among features, they may have meant multiple connections, rather than a single chain. For instance, one might have thought that, because a car is made in Norway, it is built to be drivable on glaciers and that a car can be driven on glaciers because it is heavily insulated and has treads. In this case, the structure seems to be a mixture of a common-effect and a causal-chain structure, rather than a single causal chain. Hence, it seems safe to conclude that no one has yet exclusively tested the effect of causal-chain structures on category construction.

Overview of Experiments

The purpose of the present study is to simultaneously compare three causal structures: common cause, common effect, and causal chain. In all the experiments reported here, participants received a set of 10 exemplars and were asked to create two categories of any size. Within each experiment, the exemplars to be sorted were identical across experimental conditions. Experimental conditions only differed in the instructions on how the features were causally related. Figure 3 shows schematic representations

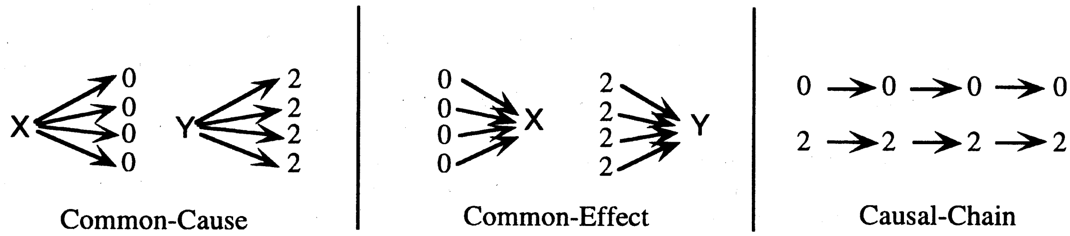

Figure 3. Schematic representations of common-cause, common-effect, and causal-chain structures for exemplar structures in Figures 1 and 2.

of the three causal structures. In the common-cause condition, the four values of one prototype (0s in Figures 1 and 2) are described as being caused by the same factor (say, X), and the four values of the other prototype (2s in Figures 1 and 2) are described as being caused by another factor (say, Y). In the common-effect condition, the four values of one prototype are described as causing the same factor (X), and the four values of the other prototype are described as causing another factor (Y). In the causal-chain condition, each set of four prototype values form a single causal chain. Finally, all the experiments employed a control condition in which no background knowledge about interproperty relations was provided.

Note that all three background knowledge conditions highlight prototype values of FR categories, by indicating how 0s (or 2s) go together. Even though the background knowledge instructions opted for the prototypes across all three conditions, it was predicted that not all background knowledge would facilitate FR sorting. More specifically, the prediction was that people would be more likely to create FR categories from the common-cause and the common-effect conditions than from the causal-chain condition.

These predictions were derived from the previous sorting studies in both knowledge-poor and knowledge-rich domains. As was noted, 1-D sorting was predominant in knowledge-poor domains without any theme provided, whereas FR sorting was significantly increased in knowledge-rich domains with a theme of common cause or common effect. This pattern of results seems to stem from people's tendency to create categories with defining features. Without background knowledge, a defining dimension would be one of the most salient surface features. With background knowledge, either a common cause or a common effect can serve as a defining dimension, while the surface features serve as diagnostic cues for the presence of common-cause or common-effect factors.[3] The preference for categories with defining features was powerfully demonstrated by Brooks and Wood (1997). In this study, even after directly experiencing members of FR categories to the degree that they could be correctly used as a basis for prediction, almost all the participants still insisted that there must have been a feature that was true of all members of the same category. This demonstration is consistent with the experience that any instructor covering the FR view runs into when students simply refuse to believe the claim that most natural categories do not have defining features (Brooks & Wood, 1997). Such conviction with regard to defining features is also consistent with psychological essentialism (Medin & Ortony, 1989), the claim that people have a strong belief that categories do have essences, even if they do not know what they are. Indeed, many researchers accounted for the learning of FR structure in terms of rule-plus-exception strategies that seem to reflect a compromise between people's belief in defining features and the lack of such defining features in natural categories (e.g., Ahn & Medin, 1992; Martin & Caramazza, 1980; Nosofsky, Palmeri, & McKinley, 1994; Ward & Scott, 1987).

If people believe that categories must have defining features and, for that reason, prefer to create categories with defining features, the common-cause and the common-effect structures are particularly compatible with FR structures, as compared with the causal-chain or no-background-knowledge situations. To begin with, common cause and common effect each seems to be a dimension that would be favored as a defining dimension, because they unify the surface dimensions, providing a coherent way of explaining all the surface features. In addition, common-cause and common-effect structures seem prevalent in natural categories. For instance, natural kinds seem to have common-cause structures where an essence (e.g., DNA structure) causes various surface features. Artifacts can be thought of as having common-effect structures in which various parts allow a common function. Once the common-cause or the common-effect is selected as a defining dimension, then for each exemplar, people would attempt to infer which common cause (or effect) is present, given the surface features (e.g., would 0002 have been caused by X or Y?). In doing so, one can examine which prototype values (0s or 2s in Figures 1 and 2) constitute the majority in the exemplar. Note that this strategy is essentially the FR principle.

In contrast, the causal-chain structure poses more difficulty in terms of FR sorting. Unlike common-cause and common-effect structures, causal-chain structures do not seem to be immediately salient structures in real-

Nopamine	N
Deroline	D
Quatasine	Q
Zovabine	Z

Nopamine	M
Deroline	L
Quatasine	S
Zovabine	B

Nopamine	N
Deroline	D
Quatasine	Q
Zovabine	V

Nopamine	M
Deroline	L
Quatasine	S
Zovabine	Z

Nopamine	N
Deroline	D
Quatasine	S
Zovabine	Z

Nopamine	M
Deroline	L
Quatasine	T
Zovabine	B

Nopamine	N
Deroline	R
Quatasine	Q
Zovabine	Z

Nopamine	M
Deroline	D
Quatasine	S
Zovabine	B

Nopamine	M
Deroline	D
Quatasine	Q
Zovabine	Z

Nopamine	P
Deroline	L
Quatasine	S
Zovabine	B

Figure 4. Exemplars used in Experiment 1.

world objects. Creating FR categories means being able to tolerate nonprototype features, such as 1 in 0100. Given the background knowledge of $0 \rightarrow 0 \rightarrow 0 \rightarrow 0$, for instance, the presence of 1 in D2 leaves two causal relations unexplained: Why did not 0 in D1 lead to 0 in D2, and how did 0 in D3 occur without 0 in D2? In a causal chain with a single path, it is difficult to find alternative explanations for the presence of nonprototype features.[4] In the case of common-cause and common-effect structures, however, the nonprototype features are more acceptable, because the surface features are independently connected to the common factor and the presence of a nonprototype feature can be compensated for by the presence of other prototype features.

Four experiments test these predictions. Experiments 1, 2, and 3 compare these three background knowledge conditions and the control condition by using both artificial stimuli (Experiment 1) and more natural stimuli (Experiments 2 and 3). Experiment 4 provides a stronger test for the effect of the causal-chain structure.

EXPERIMENT 1

In Experiment 1, the stimulus materials were made to be as artificial as possible, so that no participant would have a priori knowledge about the interproperty relations. The exemplar structure was the set in Figure 2, which failed to produce FR sorting without background knowledge (Ahn & Medin, 1992). It was predicted that the amount of FR sorting would be larger in the common-effect and the common-cause conditions than in the causal-chain and the no-background-knowledge conditions.

Method

Participants. The participants were 134 undergraduate students at the University of Louisville, participating in this experiment in partial fulfillment of introduction to psychology course requirements. All the participants were randomly assigned to one of the experimental conditions.

Materials and Design. The stimulus materials described rocks with varying mineral compositions. Each rock varied on the dimensions of the fictitious minerals nopamine, deroline, quatasine, and zovabine. Instead of using levels such as high, medium, and low, the values of each dimension were made nominal by using letters of the alphabet, in order to avoid any possible confound owing to the dimensional extremity. For example, nopamine varied by having levels N, P, or M. The abstract exemplar structure is provided in Figure 2. Figure 4 shows the ten exemplars used in Experiment 1.

The common-cause condition pertained to whether the rock was formed through the forces of an earthquake or a volcano. The participants in this condition were told, "Geologists have found that the mineral composition of rocks is formed by a volcano or an

Table 1
Results of Experiment 1: Amount of Family Resemblance (FR), Unidimensional (1-D), and Other Types of Sorting From Each Condition

Type of Background Knowledge	Type of Sorting					
	FR		1-D		Other	
	N	%	*N*	%	*N*	%
Common cause	23	74.2	7 [5, 1, 0, 0]	22.6	1	3.2
Common effect	14	48.3	10 [9, 0, 1, 0]	34.5	5	17.2
Causal chain usually	8	24.2	15 [11, 1, 2, 1]	45.5	10	30.3
Causal chain else	5	21.7	8 [5, 1, 1, 1]	34.8	10	43.5
Control	3	15.0	14 [9, 2, 2, 1]	70.0	3	15.0

Note—Numbers in brackets indicate the number of participants who selected D1, D2, D3, or D4, respectively, as a basis of their 1-D sorting.

earthquake. As you can see in the table under 'Volcano,' the rocks formed by a volcano tend to have levels N, D, Q, and Z. When rocks were formed by an earthquake, as shown under 'Earthquake' in the table, the levels tend to be M, L, S, and B." Then the participants received a table summarizing this information. In order to make it clear that the above background knowledge does not necessarily imply a certain correlational structure, the participants received the following instructions:

> Geologists, however, do not know whether the mineral composition, as described above, was determined all at once. All we know is that each level of mineral is caused by a volcano or an earthquake as described above. We do not yet know whether a volcano or an earthquake determined the levels of all 4 types of minerals at the same time. For example, if rock has N, D, Q, and Z, then it is possible that a series of 4 volcanoes could have each led to the formation of each type of minerals in the rocks. It is also possible that one volcano determined all four levels in the rock all at once. However, we do not yet know whether this is the case at this stage of scientific knowledge.

The common-effect condition explained that minerals in the rocks have healing power for indigestion or headaches. Specifically, the participants were told: "A group of physicians have found that the mineral composition of rocks has some healing power. . . . As you can see in the table under 'Indigestion,' possessing rocks with levels N, D, Q, and Z tends to help digestion. As shown in the table under 'Headache,' possessing rocks with levels M, L, S, and, B tends to cure headaches." Then the participants received a table summarizing this information. As in the common-cause condition, the participants in the common-effect condition were explicitly told that there is no known information about the correlation or competition among the features, as follows:

> All we know is that each level of mineral tends to cure indigestion or headaches as described above. For example, it is possible that N, D, Q, and Z work together to cure indigestion. It is also possible that N alone cures indigestion no matter what other levels are. However, we do not yet know whether this is the case at this stage of scientific knowledge.

There were two causal-chain conditions, depending on the specific instructions they received. In both causal-chain conditions, they were told that the 0 value in D1 in Figure 2 causes the 0 value in D2, which causes the 0 value in D3, which causes the 0 value in D4 (i.e., 0→0→0→0). They received similar causal instructions for value 2 (i.e., 2→2→2→2). Given these causal instructions, it is quite possible that the participants might assume that the causal relationship is deterministic and/or the specified causes and effects are the only possible causes and effects for given features. This interpretation would certainly prevent the participants from creating FR structure, by not allowing, for example, 0→0→1→0 as a member of a category where 0 values serves as characteristic values. In order to provide the optimal situation for creating FR structure, which is a stronger test of the prediction, two types of instructions were further added. In the causal-chain-usually condition, the participants were told that one value *usually* causes the other value, so that 0→0→1→0, for example, does not directly conflict with the provided background knowledge. As an example, they were told that the HIV virus *usually* leads to development of AIDS but it isn't *always* followed by AIDS. In the causal-chain-else condition, the participants were told, after the causal-chain instructions, that a feature can be caused by *something else*. For instance, they were told that "we know that rain causes the grass to get wet, but it is also possible that something else, such as a sprinkler may cause the grass to get wet. Similarly, we know that wet grass causes the grass to look shiny. At the same time, wet grass causes something else, such as your shoes to get wet."[5]

The order in which each dimension is presented within each exemplar was held constant across all five conditions. The specific order chosen was the one that was most compatible with the causal-chain structure (i.e., D1, D2, D3, and D4). This order would allow a conservative test of the present hypothesis, because it gives as much advantage as possible to the causal-chain condition.

Procedure. The participants received a set of 10 exemplars, each of which was written on a separate index card. They were told to lay out the cards in front of them and to read the cards very carefully while noting the four dimensions in each card. Then, participants in each condition received the appropriate background knowledge instructions. Following these instructions, they were asked to create two categories. They were told that there could be any number of exemplars in each group as long as there were two groups. In all conditions, to emphasize that they did not have to follow the classification suggested in the given background knowledge, the instructions indicated that there would be many ways to classify the cards and that there was no one correct answer. Furthermore, they were told that they did not have to have a specific reason for their categorization. After the sorting task was completed, the participants recorded their own responses by writing down the card numbers on the back of the cards under each category.

Results and Discussion

The data were classified as either FR sorting (creating two categories corresponding to the two columns in Figure 2), 1-D sorting (creating two categories on the basis of a single dimension in such a way that one category has all four exemplars with the same value of the criteria dimension and the other has the rest), or other (sorting that is neither FR nor 1-D). Table 1 summarizes the results from all five conditions.

Analysis of the common-cause and common-effect conditions. FR sorting was created much more frequently from the common-cause (74.2%) and common-effect (48.3%) conditions than from the control (15.0%) condition (Fisher's exact tests, $p < .05$ for both comparisons). These results indicate that when people learn how features cause or are caused by another dimension, they use this unifying dimension more than any of the surface dimensions, resulting in the creation of FR structures. Without this background knowledge, the participants failed to notice the FR structure of exemplars.

The common-cause condition produced more FR sorting than did the common-effect condition [$\chi^2(1, N =$

60) = 4.26, $p < .05$]. It is not clear why this difference was obtained. One possible explanation is the difference in the plausibility of the two sets of background knowledge instructions. Hence, in an attempt to equate plausibility, Experiments 2 and 3 use different stimulus materials and measure the plausibility of the background information.

Analysis of the causal-chain condition. One of the highlights of the present study is in the claim that not all background knowledge is equal. One can argue that the effect of common-cause and common-effect background knowledge reported earlier is not surprising, because the common-cause and common-effect background knowledge listed the prototype values of FR categories and, therefore, basically gave away information about which categories to create. However, the causal-chain background knowledge in Experiment 1 also refers to prototype features by accentuating how 0s and 2s go together. Therefore, the causal-chain background knowledge also gives away prototype information of FR categories and, hence, which categories to create. However, unlike the common-cause or the common-effect situations, where one feature can serve as a basis for a defining dimension unifying all the surface features, the causal chains, which lacked such a dimension, did not increase FR sorting. Chi-square tests for independence indicated that FR sorting from both types of causal-chain conditions (21.7% and 24.2%) was significantly less than FR sorting from the common-cause and common-effect conditions ($ps < .01$). Furthermore, Fisher's exact tests indicate no significant difference between the two causal-chain conditions and the control condition ($p > .10$).

To summarize, when compared with the control condition, the common-cause and common-effect conditions increased the amount of FR sorting. In addition, the causal-chain condition did not increase FR sorting, clearly demonstrating that not all background knowledge is equal.

Analysis of 1-D sorting. As is shown by the numbers in brackets in Table 1, under 1-D sorting, D1 was most likely to be selected to be the basis of 1-D sorting across all five conditions. Given that there is no reason to expect that a fictitious dimension *nopamine* is more important than dimension *deroline* and others, these results are most likely to be due to the order in which the dimensions were presented to the participants. Since the order was held constant across all conditions, the above results on FR sorting cannot be attributed to any possible interaction effect with the order of dimensions.

EXPERIMENT 2

The purpose of Experiment 2 was to generalize the results to more meaningful features. In Experiment 1, the stimulus materials were purposely developed to be as artificial as possible, in order to prevent any confounds of prior background knowledge. Experiment 2 uses more meaningful dimensions, while still using exemplars in which any existing categories or feature correlation would be unlikely. As in Experiment 1, three causal structures were used along with the control condition. Across the four conditions, identical exemplars were used. In addition, the plausibility of each set of background knowledge instructions was measured, in order to ensure that any differential effects of background knowledge were not due to differences in plausibility.

Method

Participants. Thirty-four undergraduate students at Yale University participated in this experiment in partial fulfillment of introduction to psychology course requirements.

Materials. The abstract structure of the exemplars was the same as that in Experiment 1 (see Figure 2). Four sets of stimulus materials and their stimulus dimensions were used. Three of them are shown in the Appendix, and one is shown below. For instance, in the *building* stimulus, there were four dimensions with three values each, as are shown in the parentheses: floor type (straw-mat, concrete, wood), wall (nonreinforced, reinforced, half-reinforced), number of windows (10, 6, 3), and door size (large, medium, small). Within each dimension, the first value corresponds to 0 in Figure 2, the second value corresponds to 1, and the third value corresponds to 2. Figure 5 shows the 10 exemplars that the participants received for the *house* stimulus as instantiations of the set in Figure 2.

In the common-cause condition, the participants received background knowledge on how each set of prototype values could be caused by the same factor (e.g., private purpose or public purpose) as follows:

> If a house is designed for a private purpose, the builders use straw-mat for soft flooring.
>
> If a house is designed for a private purpose, the builders use nonreinforced walls so the owner has the flexibility to remodel.
>
> If a house is designed for a private purpose, the builders install only a few windows to ensure privacy.
>
> If a house is designed for a private purpose, the builders use small doors because not many people go in and out of the house.
>
> If a house is designed for a public purpose, the builders use concrete for durable flooring.
>
> If a house is designed for a public purpose, the builders use reinforced walls to maximize the weight capacity of the house.
>
> If a house is designed for a public purpose, the builders install many windows to create a sense of openness.
>
> If a house is designed for a public purpose, the builders use large doors to account for the heavy flow of people in and out of the house.

In the common-effect condition, the participants received background knowledge, explained how each set of prototype values could cause the same factors (private use or public use) as follows:

> Using softer straw-mat for floor materials of a house allows the house to be used for a private purpose.
>
> Using flexible, nonreinforced walls allows the house to be used for a private purpose with many options for remodeling.
>
> Having a small number of windows allows the house to be private.
>
> Using a small door allows the house to be private.
>
> Using durable concrete floors in a house allows the house to be used for a public purpose.
>
> Using reinforced walls allows the house to withstand heavy usage and hence to be used for a public purpose.
>
> Having many windows causes openness and hence allows the house to be used for a public purpose.
>
> Having large doors allows the house to be used for a public purpose and to accommodate heavy traffic.

floor type	concrete
wall	reinforced
# of windows	10
door size	large

floor type	straw-mat
wall	non-reinforced
# of windows	3
door size	small

floor type	concrete
wall	reinforced
# of windows	10
door size	small

floor type	straw-mat
wall	non-reinforced
# of windows	3
door size	Medium

floor type	concrete
wall	reinforced
# of windows	6
door size	large

floor type	straw-mat
wall	non-reinforced
# of windows	10
door size	small

floor type	concrete
wall	non-reinforced
# of windows	10
door size	large

floor type	straw-mat
wall	half-reinforced
# of windows	3
door size	small

floor type	wood
wall	reinforced
# of windows	10
door size	large

floor type	concrete
wall	non-reinforced
# of windows	3
door size	small

Figure 5. Exemplars used in Experiment 2.

Finally, in the causal-chain condition, the participants were told that the prototype values in each set form a single causal chain as follows:

> Using straw-matted floors causes the builders to use nonreinforced walls because straw-matted floors cannot support heavy walls.
>
> Using nonreinforced walls forces builders to add only a small number of windows because nonreinforced walls cannot support many windows.
>
> Having a small number of windows forces builders to use small doors to fit with the overall proportion.
>
> Using concrete floors allows the builders to use reinforced walls because concrete floors can support reinforced walls.
>
> Using reinforced walls allows the builders to add many windows because reinforced walls can support many windows.
>
> Having many windows forces builders to use large doors to fit with the overall proportion.

In the control condition, no background information was provided.

Then, for each participant, a booklet containing four problems was prepared. A Latin-square design was used to select one of each set of content material and one of each background knowledge condition. For instance, one participant received the house materials in the control condition, the reptile disease materials in the common-cause condition, the plant materials in the common-effect condition, and the tribe materials in the causal-chain condition. The order of the four tasks within each booklet was randomized. As in Experiment 1, the order in which each dimension was presented within each exemplar was held constant across all four conditions (i.e., D1, D2, D3, and D4), so that it would be most compatible with the causal-chain structure, providing a conservative test of the present hypothesis.

Design and Procedure. The procedure was identical to that in Experiment 1, except that the participants performed four sorting tasks, and after creating the categories, they rated, on a 7-point scale (1 for *very implausible* and 7 for *very plausible*), the plausibility of the background knowledge they had received (i.e., the causal relations among the features). The plausibility ratings did not vary across the common-cause ($M = 5.24$, $SD = 1.23$), the common-effect ($M = 5.06$, $SD = 1.66$), and the causal-chain ($M = 5.25$, $SD = 1.30$) conditions ($p > .30$), and they were all significantly higher than the midpoint of the scale (i.e., 4; $p < .001$), indicating that all three types of background knowledge were reasonably plausible.[6] The mean plausibility ratings and the standard deviation for each content material are provided in the Appendix.

Results and Discussion

The scoring criterion was the same as that in Experiment 1. The percentages of sorting type for each condition are summarized in Table 2.

As in Experiment 1, the common-cause (55.9%) and the common-effect (64.7%) conditions encouraged FR sorting, as compared with the control (35.3%) and the causal-chain (14.7%) conditions. Because Experiment 2 was a within-subjects design, McNemar's test (McNemar,

Table 2
Results of Experiment 2: Amount of Family Resemblance (FR), Unidimensional (1-D), and Other Types of Sorting From Each Background Knowledge Condition

Background Knowledge	Type of Sorting: FR *N*	FR %	1-D *N*	1-D %	Other *N*	Other %
Common cause	19	55.9	9 [2, 3, 3, 1]	26.5	6	17.7
Common effect	22	64.7	8 [7, 0, 1, 0]	23.5	4	11.8
Causal chain	5	14.7	16 [6, 7, 0, 3]	47.1	13	38.2
Control	12	35.3	18 [7, 5, 3, 3]	52.9	4	11.8

Note—Numbers in brackets indicate the number of participants who selected D1, D2, D3, or D4 as a basis of their 1-D sorting, respectively.

1947) was used to examine FR sorting. The amount of FR sorting was reliably greater in the common-cause and common-effect conditions than in the causal-chain condition [$\chi^2(1, N=34) = 12.3, p < .001$, and $\chi^2(1, N=34) = 13.8, p < .001$, respectively], and also reliably greater than in the control condition [$\chi^2(1, N = 34) = 5.44, p < .05$, and $\chi^2(1, N=34) = 8.33, p < .01$, respectively]. There was no reliable difference between the common-cause and the common-effect conditions with respect to FR sorting ($p > .10$). Finally, there was no reliable difference between the causal chain condition and the control condition ($p > .10$).

In order to ensure that the effect was not due to particular content materials, the amount of FR sorting was further broken down by content material and is reported in Table 3. The common-cause condition led to more FR sorting than did the causal-chain and control conditions in three out of four content materials. (When the common-cause condition was worse, the difference was less than 10%.) The common-effect condition led to more FR sorting than did the causal-chain and control conditions in all four content materials.

Because Experiment 2 used a within-subjects design, order effects might be an issue of concern. However, when only the data from the first trial were examined, the main pattern of results was still obtained, with the common-cause and common-effect conditions producing more FR sorting than did the causal-chain and control conditions. In addition, Experiment 3 used a between-subjects design to further address this concern.

Another set of analyses was conducted to examine differences in 1-D sorting. According to the McNemar's test, both the common-cause and the common-effect conditions led to less 1-D sorting than did the causal-chain condition [$\chi^2(1, N=34) = 3.77, p = .06$, and $\chi^2(1, N = 34) = 4.00, p < .05$, respectively] and also less than did the control condition [$\chi^2(1, N = 34) \doteq 6.23, p < .01$, and $\chi^2(1, N= 34) = 7.14, p < .01$, respectively]. There was no difference between the common-cause and the common-effect conditions [$\chi^2(1, N = 34) = 0.11, p > .10$].

To summarize, Experiment 2 replicated Experiment 1, using more meaningful features. FR sorting was more prevalent in the common-cause and common-effect conditions than in the causal-chain and control conditions.

EXPERIMENT 3

So far, Experiments 1 and 2 failed to provide evidence that people utilize causal-chain background knowledge. Experiment 3 tests whether people rely on causal-chain background knowledge under different circumstances. It could be that people are willing to utilize the causal-chain background knowledge, but instead of creating FR categories, they might prefer to create two categories on the basis of the most fundamental cause in the causal chain. However, this strategy might have been difficult to apply, because the exemplar structures used in Experiments 1 and 2 (Figure 2) contained three values in each dimension, including the dimension serving as the most fundamental cause, and the task was to create only two categories. Hence, it is important to examine the causal-chain condition when there are binary values on each dimension and to test whether this would encourage the use of causal-chain knowledge. That is, although causal-chain background knowledge might not encourage FR sorting, it might have another effect on category construction—namely, encouraging 1-D sorting on the basis of the most fundamental cause.

Method

As in Experiment 2, there were four conditions varying in background knowledge. All four sets of content materials in Experiment 2 were used, except that the exemplar structure was the set in Figure 1. Each participant carried out only one sorting task per background knowledge condition. Thirty-nine undergraduate students at Yale University participated in this experiment, in partial fulfillment of introduction to psychology course requirements. There were 7 participants in the common-cause condition, 8 in the common-effect condition, 12 in the causal-chain condition, and 12 in the control condition. There were more participants in the causal-chain and control conditions because they were the focus of Experiment 3.

Results and Discussion

The scoring criteria were the same as those in Experiment 1. The percentages of sorting type for each condition

Table 3
Amount of Family Resemblance (FR) Sorting in Experiment 2 for Each Content Material

Background Knowledge	Content Materials: Reptile Disease *N*	Reptile Disease %	Tribe *N*	Tribe %	Plant *N*	Plant %	Building *N*	Building %
Common cause	7/9	77.8	3/8	37.5	6/10	60.0	3/7	42.9
Common effect	6/7	85.7	5/9	55.6	3/8	37.5	8/10	80.0
Causal chain	0/10	0	3/7	42.9	1/9	11.1	1/8	12.5
Control	0/8	0	5/10	50.0	2/7	28.6	5/9	55.6

Note—The numbers indicate the number of participants who created FR categories out of the total number of participants who received the corresponding content material for a given condition.

Table 4
Results of Experiment 3: Amount of Family Resemblance (FR), Unidimensional (1-D), and Other Types of Sorting From Each Background Knowledge Condition

Type of Background Knowledge	Type of Sorting					
	FR		1-D		Other	
	N	%	*N*	%	*N*	%
Common cause	4	57.1	3	42.9	0	0.0
Common effect	3	37.5	3	37.5	2	25.0
Causal chain	0	0.0	8	47.1	4	38.2
Control	0	0.0	11	52.9	1	11.8

are summarized in Table 4. Even with different exemplar structures, FR sorting was created more frequently from the common-cause (57.1%) and common-effect (37.5%) conditions than from the control and causal-chain conditions (0% for both conditions; Fisher's exact tests, $ps < .05$). The difference between the common-cause and the common-effect conditions in creating FR categories was not statistically significant [$\chi^2(1, N = 15) = 0.58, p = .45$].

Among those who created 1-D categories, the basis of their sorting was examined. Five out of 12 participants in the causal-chain condition selected the first dimension (i.e., the most fundamental cause in the causal-chain condition) as a basis for 1-D sorting. Likewise, 7 out of 12 participants in the control condition selected the same dimension as a basis for 1-D sorting. Hence, this dimension was chosen as a 1-D sorting basis equally often, regardless of whether the background knowledge indicated it to be the most fundamental cause in a chain. Again, as was explained in Experiment 1, the reason D1 was selected most frequently was presumably because it was the dimension presented first in each exemplar. Present results show that this preference for the first dimension mentioned did not increase any further, given the causal-chain background knowledge.

In addition, the overall FR sorting from the common-cause and common-effect conditions is somewhat less than that in Experiment 2, which used identical content materials and instructions. This result seems to be due to the differences in exemplar structure. As is shown in Figure 1, all the exemplars used in Experiment 3, except for the prototypes, contained a value in the contrasting category, whereas fewer than half of the exemplars used in Experiment 2 (Figure 2) contained a value in the contrasting category (i.e., E3, E5, E7, and E9). It has been shown that the overall amount of FR sorting is a function of the amount of this kind of crossed-over features (Ahn & Medin, 1992).

In sum, Experiment 3 replicated the advantage of the common-cause and common-effect background knowledge over the causal-chain background knowledge and also demonstrated the lack of FR sorting, given causal-chain background knowledge with exemplars that have binary values.

EXPERIMENT 4

Across three experiments, the causal-chain condition produced sorting no different from that for the control condition (no background knowledge). Experiment 4 provides a final test for the causal-chain condition with a different exemplar structure. Previous exemplars in Experiments 1, 2, and 3 have so-called *crossed-over* features, in that features that are characteristic of one category are crossed over to the opposite category. For instance, in Figure 2, the 0 value is characteristic of a category on the left column, but it also occurs in the contrasting category. Spalding and Murphy (1996) suggested that crossed-over features could discourage people from applying causal-chain background knowledge, because they cause contradictions. Referring to the sets in Figures 1 and 2, they stated that such a design with crossed-over features (or the so-called characteristic design) "leads to a real problem when the features are related by prior knowledge. The characteristic design would lead to examples like the following: A bird that flies, builds nests in trees, perches on power lines, and does not have wings" (p. 533). In Experiment 4, it is examined whether the causal-chain condition still fails to lead to FR sorting even when crossed-over features are not used and there cannot be

	D1	D2	D3	D4		D1	D2	D3	D4
E1	0	0	0	0	E6	2	2	2	2
E2	0	0	1	1	E7	2	2	1	1
E3	0	1	1	0	E8	2	1	1	2
E4	1	1	0	0	E9	1	1	2	2
E5	1	0	0	1	E10	1	2	2	1

Figure 6. Exemplar structure used in Experiment 4.

Table 5
Results of Experiment 4 for the Causal-Chain and Control Conditions Broken Down for the Absent and Present Features

Type of Background Knowledge	FR *N*	FR %	1-D *N*	1-D %	Other *N*	Other %
Causal-chain						
Absent	20	62.5	1	3.1	11	34.4
Present	12	37.5	4	12.5	16	50.0
Control						
Absent	20	62.5	2	6.3	10	31.3
Present	7	21.9	7	21.9	18	56.3

Note—FR, family resemblance; 1-D, unidimensional.

contradictions in applying the causal-chain background knowledge. In this way, Experiment 4 provides a stronger test for the causal-chain condition.

Method

Participants. The participants were 32 undergraduate students at the University of Louisville. The participants received partial credit toward a requirement in their introductory psychology class or were monetarily reimbursed for their time. They were randomly assigned to each condition of the experiment.

Materials. The exemplar structure used in Experiment 4 did not contain crossed-over features (see Figure 6). Previous studies (Ahn & Medin, 1992) have shown that, without background knowledge, many participants produce FR sorting if there are no crossed-over features (see the discussion of Experiment 4 for further explanation). This particular set of exemplars was selected on the basis of previous results so that there would still be room for FR sorting to increase when background knowledge was added.

Four sets of objects were created following this structure. For each object, the values that were not characteristic of either category (i.e., 1s in Figure 6) were either some other value (*present*, henceforth) or missing values (*absent*, henceforth). When the values were missing or absent, a question mark was presented in its place, and participants were told that no information was available about the dimension. For example, for reptile materials, the participants saw the following for the exemplar 2112 in Figure 6:

(1) temperature : high
(2) blood pressure : ?
(3) insulin level : ?
(4) pupil dilation : large.

Using missing values for nonprototype values allowed us to minimize the possibility of contradictions among features when causal-chain knowledge was applied. That is, if the causal-chain background knowledge still does not increase the amount of FR sorting under this situation, it is safe to conclude that the lack of a causal-chain knowledge effect is not due to contradictions among features.

For each object, two sets of instructions were developed, one for no background knowledge and the other for causal-chain knowledge. As before, the causal-chain background knowledge instruction indicated how the prototype values of each category form a single causal chain. The rest of the instructions for both conditions was the same as those in Experiment 1.

Design and Procedure. Each participant received four problems. Two problems were from the causal-chain condition, and two were from the control condition. Within each condition, one problem had missing values for the 1s in Figure 6, and the other problem had neutral values for the 1s. For the same participant, each problem had a different content. These constraints ensured that each participant would never see the identical content more than once while they sorted both absent and present materials with and without the causal-chain background information. A Latin-squares design was used to determine which participants would receive which four versions within each exemplar structure set. The procedure was identical to that in Experiment 2.

Scoring criterion. The participants' responses were classified as FR, 1-D, or other type of sorting. The two FR categories for each set correspond to the two major columns in Figure 6. Sorting was considered 1-D if one of the two categories created by each participant had all instances with the same value on a single dimension but none of the instances with the different value on the same dimension. For instance, for the neutral values, grouping 0000, 0011, and 0110 in one category and the rest in the other category would be 1-D along the first dimension. This criterion offers evidence for selective attention to a single dimension, because 1100 and 1001 would be placed in the alternative category, despite the fact that they share many values with the members in the first category along the noncriterial dimension. Finally, if a sorting was neither FR nor 1-D, it was classified as *other*.

Results and Discussion

Effect of causal-chain background knowledge. Given no background knowledge, 43.3% of the responses were FR sorting. As before, the number did not significantly increase with the causal-chain background knowledge (51.5%). Because each participant sorted two sets of exemplars for each type of background knowledge, *t* tests were carried out for analyses by coding each FR sorting as 1 and each non-FR sorting as 0. This way, a participant's score for each condition ranged between 0 and 2. The difference between the causal-chain and the control conditions was not statistically significant [$t(1) = 1.15, p = .26$]. Even when the nonprototype values were missing, there was no additional effect of causal-chain background knowledge in producing FR sorting (62.5% for the control condition and 62.5% for the causal-chain condition). As is shown in Table 5, there seems to be a slight increase in FR sorting from the causal-chain condition when the nonprototype values were present (37.5%), as compared with the control condition (21.9%). This difference was not statistically significant according to McNemar's (1947) test ($\chi^2 = 0.67$), and furthermore, the direction was opposite to the alternative hypothesis of Experiment 4. That is, if contradictions among features were responsible for the lack of a causal-chain knowledge effect, FR sorting should have decreased when the nonprototype values were explicitly present.

Compared with Experiments 1, 2, and 3, there was more FR sorting in Experiment 4 even from the control condition. This was because of the particular structure of the exemplars used in Experiment 4. Experiments 1, 2, and 3 used exemplar structures that had crossed-over features (i.e., prototype features of one category occurred in the contrasting category). A detailed explanation of the effect of crossed-over features on 1-D sorting is beyond the scope of this paper (see Ahn & Medin's, 1992, two-stage model). To briefly describe their account, in the first stage, 1-D sorting is carried out on the most salient dimension (e.g., large vs. small objects). In the second stage, exemplars that were not classified in the first stage (e.g., medium objects) were sorted on the basis of their overall similarity to categories created in the first stage. According to the model, FR sorting is unlikely to occur with crossed-over features, because it destroys FR structure even at the first stage. For instance, with the structure in Figure 2, if D2 is selected as a basis of 1-D sorting in the first stage, E9 would be classified with E1, E2, E3, and E5, violating the FR principle. But without crossed-over features, as in Figure 6, no FR structure would be broken in the first stage. This is a well-established phenomenon in that, when no crossed-over features exist, FR sorting was obtained from 20%–55% of the participants, even without any background knowledge.

At the same time, this high percentage of FR sorting from the control condition in Experiment 4 does not seem to have contributed to the lack of difference between the causal-chain condition and the control condition. Ahn (1990a) showed that when common-cause or common-effect background knowledge was provided for exemplars without crossed-over features, FR sorting could be obtained up to 98%. Given that the proportions of FR sorting in the control conditions of Experiment 4 were 62.5% in the absent condition and 21.9% in the present condition, clearly there was room for further increase in the amount of FR sorting, should there have been the effect of causal-chain background knowledge .

Collapsing over Experiments 1–4, the null effect of causal-chain background knowledge, as compared with the control condition, is striking. Out of 166 sorting responses in the causal-chain condition, 30.1% were FR, and out of 128 in the control condition, 32.8% were FR. Furthermore, Experiment 4 provided one of the most optimal conditions for the causal-chain condition by eliminating contradictions among features. However, no increase in FR sorting was observed for causal-chain knowledge over FR sorting in the control condition.

Absent versus present conditions. Although not the main purpose of the present paper, there were more FR sortings from the absent condition (62.5%, collapsed across the control and the causal-chain condition) than from the present condition (29.5%; $p < .001$). These results are reasonable in that, if noncharacteristic values (i.e., 1s in Figure 6) are absent, rather than taking specific values, the dissimilarity among exemplars would be reduced.

GENERAL DISCUSSION

Across three different exemplar structures and three different sets of content materials, the present experiments demonstrated that when common-cause and common-effect background knowledge specified the relations among prototype values, the creation of FR categories was facilitated. Causal-chain background knowledge, although explicitly accentuating prototype values, did not lead to FR sorting. This finding was obtained when each dimension had two values (Experiments 1 and 2) as well as three values (Experiment 3). Even when no contradiction could occur because missing values were used (Experiment 4), responses made in the causal-chain condition did not differ from those made in the control condition.

In the introduction, predictions were made by referring to previous sorting studies demonstrating the 1-D sorting bias. This bias was interpreted not as a strategy participants would adopt as a quick and dirty way of creating categories, but rather as a preference for creating categories with defining features. When surface features are connected to a common cause or a common effect, sorting on the basis of these common factors provides a means to create categories with defining features, and, at the same time, the method encompasses all of the dimensions present. In this way, the present account provides a coherent explanation for sorting in both knowledge-poor and knowledge-rich domains on the basis of a single mechanism.

Is 1-D sorting on the basis of a surface feature essentially the same as sorting on the basis of a deeper dimension, such as a common cause or a common effect? On the one hand, the basic mechanism might be quite similar, in that they both assume defining features. On the other hand, 1-D sorting on the basis of a surface feature is much more rigid, because if there is a mismatch along the criterial dimension, that exemplar is weeded out. In addition, 1-D sorting on the basis of a surface dimension is more rigid, in that selective attention might be drawn only to that dimension and overall similarity between exemplars ignored. In contrast, sorting on the basis of a deeper dimension is more flexible, in that, as long as there is enough surface evidence indicating the presence of a deeper dimension, some nonprototype features can be tolerated. For that reason, FR sorting was predicted from the common-cause and common-effect conditions. Similarly, Markman (1989) suggested that FR structures are superficial manifestations of a deeper explanatory principle. Using Medin et al.'s (1987) example, she explained that the properties of the category *bird*, such as feathers, wings, and building nests in trees, can be interpreted as adaptations to flying. Members of the bird category can manifest this underlying principle in different manners to different degrees, resulting in FR structures at the surface level.

Implications for Structures of Natural Categories

The present results clearly show that people *spontaneously* applied the common-cause and common-effect structures to produce FR sorting, even when they could have used, instead, an easier strategy of sorting on the basis of a single surface feature. As has been noted by Lassaline and Murphy (1996), FR sorting is definitely more difficult than 1-D sorting, because one has to look over multiple dimensions simultaneously. Despite this added difficulty and the specific instructions to the participants that they could create any categories, FR sorting was increased when the common-cause and common-effect instructions were provided. The persistent lack of a causal-chain background knowledge effect also suggests that the results from the common-cause and common-effect conditions are not due to some kind of demand characteristic imposed by the background knowledge instructions. Taken together, these results provide indirect evidence that people perceive the common-cause and common-effect structures as more natural category structures than the causal-chain structures.

Indeed, the common-cause and common-effect structures seem to conform to the causal structures of natural categories. Natural kinds, such as animals and plants, are believed to have essences that are responsible for their surface features (Medin & Ortony, 1989; see Markman, 1989, for similar accounts), reflecting a common-cause structure. In addition, one can argue that artifacts have common-cause structures in which surface physical features of the same object are caused by a designer's intention (Bloom, 1996). One might alternatively think of artifacts as having common-effect structures, in that the surface features of an artifact object all allow performance of a certain function. For instance, all the parts of a car are present for the common effect of transportation. Because of this similarity to the structure of natural categories, people might have spontaneously relied on such causal background knowledge, creating categories corresponding to these structures.

In contrast, the causal-chain structure does not seem to be representative of the structure of natural categories. Causal chains do exist in real-life categories, as in the bird example in which having wings causes or allows flying, which in turn causes or allows building nests in trees. Still, causal-chain structures seem, rather, to be part of a more fundamental structure, such as a common-cause or a common-effect structure. That is, when we think of a bird concept as a whole, the most prominent structure is how the underlying essence would lead to surface features, rather than how the surface features are related to each other into a single causal chain. A more direct examination of causal structures in natural categories awaits future research.

Effect of Causal-Chain Structures

Clearly, null results from causal-chain structures in the present study do not necessarily mean that there is no effect of causal-chain structures on category construction. For instance, when a causal chain structure is embedded in a common-cause or a common-effect structure, there might be an additional effect of causal-chain structure as an organizing force. Another potentially important factor is the domain. The present experiments examined only object categories (e.g., buildings, rocks, plants, tribes, etc.) and failed to find the effect of causal-chain structures. However, event categories (e.g., traffic accidents, ceremonies) tend to be formed in a sequence of causal chains, and hence, people might spontaneously rely on causal-chain background knowledge for organizing such events.

Other Possible Roles of Background Knowledge in Categorization

The present study examined only one specific aspect of the background knowledge effect. That is, it presented one way of classifying different kinds of background knowledge—namely, causal connectivity. This limited focus should be interpreted as a research strategy, rather than as a claim that it is the only role of background knowledge in categorization.

One possible knowledge effect that has not been studied in this paper is discussed in Wattenmaker (1995). In that study, more FR sorting occurred with social categories (i.e., "the categorization of people based on traits or behavioral characteristics," p. 277) than with object categories (i.e., "the categorization of concrete entities in the environment such as animals, plants, and human artifacts," pp. 277-278). According to Wattenmaker, one of the reasons for this domain difference might be that social features are more flexible than object features. For instance, a behavior statement, "he placed himself in front of a truck at the gates of the nuclear power plant," can be interpreted as *principled*, *criminal*, *conscientious*, *eccentric*, *courageous*, and so forth. According to Wattenmaker, this flexibility of social features encourages FR sorting, because it is easier to tolerate any inconsistency caused by noncharacteristic features and it allows us to easily identify a common theme in social features.

Conclusion

The present results suggest that people do not seem sensitive to correlational structures within categories when they are not meaningful with respect to background knowledge. As discussed by Murphy and Medin (1985) and Keil (1981), the amount of correlation one can notice within an object is computationally intractable. If a correlation exists because of a common cause or a common effect, people seem to act *as if* they are utilizing correlation in constructing categories. Indeed, it might be quite rational to expect FR construction to occur only when background knowledge permits. The present study demonstrated that the causal-chain structure does not provide a meaningful basis for noticing an FR structure. Future studies should investigate the rational basis for such differential effects of causal background knowledge.

REFERENCES

AHN, W. (1990a). Effects of background knowledge on family resemblance sorting. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society* (pp. 149-156). Hillsdale, NJ: Erlbaum.

AHN, W. (1990b). *A two-stage model of category construction*. Unpublished doctoral dissertation, University of Illinois, Urbana.

AHN, W. (1991). Effects of background knowledge on family resemblance sorting and missing features. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 203-208). Hillsdale, NJ: Erlbaum.

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

BLOOM, P. (1996). Intention, history, and artifact concepts. *Cognition*, **60**, 1-29.

BROOKS, L. R., & WOOD, T. (1997, November). *Identification in service* *of use: Characteristic of everyday concept learning*. Paper presented at the 38th annual meeting of the Psychonomic Society, Philadelphia.

Byrne, R. M. J. (1997). Cognitive processes in counterfactual thinking about what might have been. In D. Medin (Ed.), *The psychology of learning and motivation* (Vol. 37, pp. 105-154). San Diego: Academic Press.

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

Gelman, S. A., & Kalish, C. W. (1993). Categories and causality. In R. Pasnak & M. L. Howe (Eds.), *Emerging themes in cognitive development* (Vol. 2). New York: Springer-Verlag.

Imai, S., & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, **69**, 596-608.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. *Psychological Review*, **88**, 197-227.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic Bulletin & Review*, **3**, 95-99.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

Martin, R. C., & Caramazza, A. (1980). Classification in well-defined and ill-defined categories: Evidence for common processing strategies. *Journal of Experimental Psychology: General*, **109**, 320-353.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**, 153-157.

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

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, **20**, 158-190.

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

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

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **11**, 299-339.

Nakamura, G. V. (1985). Knowledge-based classification of ill-defined categories. *Memory & Cognition*, **13**, 377-384.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 416-432.

Regehr, G., & Brooks, L. R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 347-363.

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

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smith, L. B. (1981). Importance of the overall similarity of objects for adults' and children's classifications. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 811-824.

Spalding, T. L., & Murphy, G. L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 525- 538.

Ward, T. B., & Scott, J. (1987). Analytic and holistic modes of learning family-resemblance concepts. *Memory & Cognition*, **15**, 42-54.

Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. *Cognitive Psychology*, **28**, 274-328.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

NOTES

1. It should be noted that the actual value of 0 on D1 is different from the actual values of 0 on other dimensions (e.g., D2). This is a notation commonly used for ease of understanding prototype values in FR structure. That is, in these exemplars, 0s are prototype values for one FR category and 2s are prototype values for the other FR category.

2. The present study does not consider cyclic relations where A causes B which causes A in turn.

3. It is acknowledged that sorting on the basis of a deeper dimension (e.g., a common cause) might not be exactly the same as 1-D sorting on the basis of a surface dimension. See the General Discussion section for more discussion on this issue.

4. Although some studies show that people are very good at fitting features into alternative explanations of a causal sequence (e.g., Byrne, 1997), the point here is that, as compared with common-cause and common-effect structures, causal-chain structures require more effort to come up with alternative explanations.

5. Since this experiment is the first-known direct test of causal-chain structures on sorting, the effect of these specific instructions (i.e., the use of *usually* and *something else*) is yet unclear. Later experiments reported in the paper use different causal-chain instructions (i.e., no use of *usually* and *something else*) for generality. Likewise, one might question the differences in plausibility of all of the background knowledge instructions. Later experiments use different materials and also empirically equate plausibility.

6. In addition, an independent group of 13 participants was asked to rate only on plausibility, without having to sort exemplars beforehand. These participants received three critical causal background scenarios and one filler item inserted to allow for the same Latin-square design used in Experiment 4 (i.e., one from each background knowledge condition instantiated in four different content materials). Again, the plausibility ratings did not vary across the common-cause (4.85), the common-effect (4.92), and the causal-chain (5.46) conditions ($p > .40$), and they were all significantly higher than the midpoint of the scale ($p < .05$).

APPENDIX
Stimulus Materials Used in Experiments 2 and 3

(Note: Means and standard deviations reported in parentheses are the results from the plausibility ratings.)

1. Reptile Disease

Stimulus Dimensions

Body Temperature (high, normal, low)
Blood Pressure (high, normal, low)
Insulin Level (low, moderate, high)
Pupil Dilation Size (small, medium, large)

Common Cause ($M = 4.89$, $SD = 1.05$)

In reptiles, enlarged livers cause high body temperature. In reptiles, enlarged livers cause high blood pressure. In reptiles, enlarged livers cause a low insulin level. In reptiles, enlarged livers cause small pupil dilation.

APPENDIX (Continued)

In reptiles, atrophied livers cause low body temperature. In reptiles, atrophied livers cause low blood pressure. In reptiles, atrophied livers cause a high insulin level. In reptiles, atrophied livers cause large pupil dilation.

Common Effect ($M = 4.17$, $SD = 2.14$)

In reptiles, high body temperature causes enlarged livers. In reptiles, high blood pressure causes enlarged livers. In reptiles, a low insulin level causes enlarged livers. In reptiles, small pupil dilation causes enlarged livers.

In reptiles, low body temperature causes atrophied livers. In reptiles, low blood pressure causes atrophied livers. In reptiles, a high insulin level causes atrophied livers.

In reptiles, large pupil dilation causes atrophied livers.

Causal Chain ($M = 5.1$, $SD = 1.60$)

In reptiles, high body temperature causes high blood pressure. In reptiles, high blood pressure causes a low insulin level. In reptiles, a low insulin level causes small pupil dilation.

In reptiles, low body temperature causes low blood pressure. In reptiles, low blood pressure causes a high insulin level. In reptiles, a high insulin level causes large pupil dilation.

2. Tribe

Stimulus Dimensions

Food attainment (farming, gathering, hunting)

Leadership (hierarchically organized leaders, two leaders, single leader)

Religion (one god of Phaus, 2 gods of Venom, 12 gods of Zenux)

Methods of disposing the dead (burial, throwing the corpse into the river, cremation)

Common Cause ($M = 4.38$, $SD = 1.51$)

In Central Africa, settling down in one place causes tribes to obtain food through farming, because the natural resources are soon depleted. In Central Africa, settling down in one place causes tribes to have hierarchically organized leaders so that they can deal with the complex social structures that result from the settlement. In Central Africa, settling down in one place causes tribes to believe in the one god of Phaus, due to the lack of contact with other religions. In Central Africa, settling down in one place causes tribes to bury the dead so that they can visit the tomb.

In Central Africa, wandering around causes tribes to obtain food through hunting, because they do not stay in one place long enough to obtain food through farming. In Central Africa, wandering around causes tribes to have a single leader, so decision making is fast and efficient. In Central Africa, wandering around causes tribes to believe in the 12 gods of Zenux, because they are influenced by many different cultures as they travel. In Central Africa, wandering around causes tribes to cremate the dead, because they cannot visit the tomb anyway.

Common Effect ($M = 5.89$, $SD = 0.93$)

In Central Africa, obtaining food through farming causes tribes to settle in one place. In Central Africa, having hierarchically organized leaders causes tribes to settle in one place so that they can deal with the complex social structure. In Central Africa, believing in the one god Phaus causes tribes to settle in one place, because they believe the place is blessed by Phaus. In Central Africa, burying the dead causes tribes to settle in one place, because they would like to visit the tomb.

In Central Africa, hunting causes tribes to wander around and follow the paths of the animals. In Central Africa, having a single leader allows tribes to wander around, because fast and efficient decisions can be made. In Central Africa, believing in the 12 gods of Zenux causes tribes to wander around, because they believe these 12 gods exist in different parts of the continent. In Central Africa, cremating the dead causes tribes to wander around, because they do not want to be haunted by the cremated dead.

Causal Chain ($M = 5.5$, $SD = 0.84$)

In Central Africa, farming causes tribes to have hierarchically organized leaders, because farming requires specialized decisions along with the coordination of these specialized decisions. In Central Africa, having hierarchically organized leaders causes tribes to believe in the god Phaus so that their complex social structures can be unified under one god. In Central Africa, believing in the god Phaus causes tribes to bury the dead, because Phaus is supposed to resurrect the dead at the end of the world.

In Central Africa, hunting causes tribes to have a single leader, because quick and efficient decisions need to be made during hunting. In Central Africa, having a single leader causes tribes to believe in the 12 gods of Zenux, so that they can allow different value systems among the members, even though they are under a single leader. In Central Africa, believing in the 12 gods of Zenux causes tribes to cremate the dead, because they believe the 12 gods fight for the dead body, and cremation would allow the ashes to be divided up among the 12 gods.

3. Plant

Stimulus Dimensions

Amount of nectar in mg produced in 24 hours (more than 2 mg; 1.5 mg; less than 1 mg)

Type of pollination ("cross-pollinated" meaning pollination occurs across plants; "cross- and self-pollinated" meaning both methods are possible; "self-pollinated" meaning pollination occurs within the same plant)

Petal shape (disk shape meaning flat, disk-like shape; lip shape meaning two petals attached like lips; tubular shape meaning trumpet-like shape)

APPENDIX (Continued)

Pollen grains (sticky, coarse, smooth)

Common Cause (*M* = 6.0, *SD* = 0.94)

Flowers that are pollinated by insects have evolved to produce more than 2 mg of nectar a day, in order to attract as many insects as possible. Flowers that are pollinated by insects have evolved to be cross-pollinated plants, because insects tend to hover across different plants. Flowers that are pollinated by insects have evolved to have flat disk petals in order to make landing easier for insects. Flowers that are pollinated by insects have evolved to have sticky pollen grains that stick to insects.

Flowers that are pollinated by wind have evolved to produce less than 1 mg of nectar a day, because there is no need to produce nectar. Flowers that are pollinated by wind have evolved to be self-pollinated plants, because even mild wind can make pollination possible within the same plant. Flowers that are pollinated by wind have evolved to have tubular petals, because the tubular shape prevents pollens from being lost by a strong wind. Flowers that are pollinated by wind have evolved to have smooth pollen grains, so the pollen grains can float easily.

Common Effect (*M* = 4.88, *SD* = 1.96)

Producing more than 2 mg of nectar a day allows flowers to be pollinated by insects who look for nectar. The need for cross-pollination causes flowers to be pollinated by insects, because insects tend to hover across plants. Disk petals allow flowers to be pollinated by insects, because disk petals are easier for insects to land on. Sticky pollen grains allow flowers to be pollinated by insects, because pollen grains can easily stick to insects.

Producing less than 1 mg of nectar a day causes flowers to be pollinated by wind, because they cannot attract insects. The need for self-pollination causes flowers to be pollinated by wind, because even mild wind can make pollination possible within the same plant. Tubular petals allow flowers to be pollinated by wind, because the tubular shape prevents pollens from being lost by strong wind. Smooth pollen grains allow flowers to be pollinated by wind, because smooth pollen grains can float easily.

Causal Chain (*M* = 5.22, *SD* = 1.64)

Producing more than 2 mg of nectar a day allows plants to be cross-pollinated through insects. The need for cross-pollination caused the flowers to have flat disk petals, because flat disks optimize dissemination of pollens to other plants. Disk-shape petals caused flowers to have sticky pollen grains, so that pollen grains can be contained in the flowers without falling to the ground.

Producing less than 1 mg of nectar a day causes flowers to be self-pollinated, because few insects would visit these flowers. The need for self-pollination caused flowers to have tubular petals so that the chance of pollination within the same plant can be increased. Tubular petals allow flowers to have smooth pollen grains, because the smooth pollen grains can be easily contained in the tube.

4. Buildings

(See the text for the stimulus dimensions and scenarios.)

Common Cause (*M* = 5.67, *SD* = 0.51)

Common Effect (*M* = 4.80, *SD* = 1.48)

Causal Chain (*M* = 5.63, *SD* = 0.74)

(Manuscript received June 2, 1998;
revision accepted for publication January 3, 1999.)