

Fixed vs. Varying Common-Cause and Common-Effect Schemas

The results of the first experiment confirmed the prediction that categorization judgments in the presence of causal knowledge reflect the statistical structure of category members that that knowledge is expected to generate. In an additional experiment a second test of this prediction was conducted using variants of the common-cause and common-effect schemas shown in Figure 1. In this experiment, the common cause feature F_1 of the common-cause schema was described as occurring in some, but not all, of category members. As a result, this condition used what has been called a varying-cause common-cause schema (Waldmann et al. 1995), because the presence of the common-cause F_1 varies over category exemplars. In contrast, a second experiment used a fixed-cause common-cause schema in which the common cause F_1 was described as occurring in all category members. One consequence of the common cause feature being fixed rather than varying lies in the different pattern of correlations among features produced by a fixed common cause schema. Whereas, a varying common-cause schema produces correlations between the common cause and its effects and also correlations between the effects themselves (Figure 2), a fixed common-cause schema produces correlations between the common cause and its effects but no correlations between the effects. This difference in the pattern of correlations can be illustrated with the disease example presented earlier. When only a certain subpopulation has a disease (a varying-cause common-cause schema) then the disease's symptoms will be correlated. In contrast, when all members of a population have the disease (a fixed-cause common-cause schema), the symptoms are no longer correlated. This is because the presence of one symptom does not increase the probability of the presence of the disease (which is already present with probability 1). Hence the probability that other symptoms are present also does not increase.

Likewise, in the first experiment the common effect feature F_4 of the common-effect schema was described as occurring in some, but not all, of category members. In contrast, the present experiment used a fixed-cause common-effect schema in which the common effect F_4 was described as occurring in all category members. However, whereas causal-model theory predicts that the sensitivity to correlations among effect features should be absent as a consequence of describing the common cause as fixed, it predicts that a fixed-effect common-effect schema should continue to exhibit the same kind of higher-order interactions among cause features observed in the first experiment: A discounting effect such that category membership ratings should be a nonlinear function of the number of cause features present. Confirmation of these predictions will again indicate the importance of the asymmetry of causal relationships, because as in the first experiment the fixed common-cause and common-effect conditions used here are analogs of one another if one ignores the direction of the causal arrow.

To test whether category membership ratings would mirror the statistical structure of category members generated by fixed common-cause and common-effect schemas, the second experiment replicated the first except that F_1 in the common-cause condition, and F_4 in the common-effect condition, was described as occurring in 100% of all category members. Moreover, these features were made defining features by also telling participants that they appeared in no other categories, and the names of the categories were changed to further emphasize their defining nature. For example, for the Lake Victoria Shrimp category feature F_1 is “high amounts of acetylcholine” and F_4 is “high body weight.” Accordingly, in the (fixed) common-cause condition Lake Victoria Shrimp were renamed Acetylcholine Shrimp, and participants were told that all Acetylcholine Shrimp have high amounts of acetylcholine and that no other kind of shrimp does. In the (fixed) common-effect condition Lake Victoria Shrimp were renamed Heavy Shrimp, and participants were

told that all all Heavy Shrimp have an usually high body weight not seen in any other kind of shrimp. Note that the fixed common-cause condition was intended as an analog to those real-world categories in which a defining feature causes other characteristics (e.g., the DNA of a species causing the observed morphology and behaviors of a species), whereas the fixed common-effect condition was in which a defining feature has multiple causes (e.g., being “HIV positive” can be caused by sharing needles, blood transfusion, or unsafe sex).

When generating category membership ratings in the first experiment, participants were always given complete information regarding the presence or absence of each of the four binary features for the presented category exemplar. In contrast, in the real world we usually cannot directly observe defining features such as DNA or being HIV-positive. To simulate this situation, the test exemplars that participants were presented with always listed the defining feature as unknown. That is, in the common-cause condition participants were presented with the eight three-feature exemplars with F_1 missing: x000, x001, x010, x100, x011, x101, x110, and x111 (where x denotes an unknown value). To provide a comparison group for this condition, an F_1 -control condition was identical in all respects except that the three common-cause causal relationships were omitted. Likewise, the common-effect condition participants were presented with the eight three-feature exemplars with F_4 missing: 000x, 001x, 010x, 100x, 011x, 101x, 110x, and 111x. The F_4 -control condition was identical to the common-effect condition except that the three common-effect causal relationships were omitted.

Method

Materials. The categories in Appendix A were renamed as follows in the common-cause and the F_1 -control condition: Iron Sulfate Ants, Acetylcholine Shrimp, Ionized Helium Stars, Radioactive Sodium Carbonate, Butanos, and Magnetic

Computers. In those conditions the category feature F_1 was described as occurring 100% of the time in category members, and 0% of the time in members of other categories. In the common-effect condition and the F_4 -control condition the categories were named: Fast Ants, Heavy Shrimp, Planet Stars, Reactive Sodium Carbonate, Carbonos, and Bright Computers, and F_4 was described as occurring 100% of the time in category members, and never in members of other categories.

Procedure. The training procedure was identical to that in first experiment except for the new information about F_1 (or F_4). After category learning, participants were presented with test exemplars where the presence or absence of the defining feature was unknown. That is, in the common-cause and F_1 -control conditions, participants produced category membership ratings for the eight three-feature exemplars x000, x001, x010, x100, x011, x101, x110, and x111. In the common-effect and F_4 -control conditions participants produced ratings for the eight three-feature exemplars 000x, 001x, 010x, 100x, 011x, 101x, 110x, and 111x.

Participants. One hundred and eight New York University undergraduates received course credit or \$10 for their participation, with 36 participants each assigned to the common-cause and common-effect conditions, and 18 each to the F_1 -control and F_4 -control conditions.

Results

Category membership ratings for the eight test exemplars averaged over participants in the common-cause, common-effect, and two control conditions are presented in Table A1, and in Figure A1 for selected exemplars. As in the first experiment, the presence of inter-feature causal relationships in the common-cause and common-effect conditions affected the category membership ratings given to a number of the test exemplars relative to their respective control conditions. For instance, in the common-cause condition (Figure A1a), exemplar x000 was given a

significantly lower rating (29.1) and exemplar x111 was given a significantly higher rating (92.6) than in the F_1 -control condition (52.8 and 76.1, respectively).

Presumably, these differences in ratings obtained because exemplar x000 broke expected pairwise correlations between the effect features and the unobserved common-cause and exemplar x111 preserved those correlations. Similarly, in the common-effect condition (Figure A1b), exemplar 000x was given a significantly lower rating (30.8) and exemplar 111x was given a significantly higher rating (91.8) than in the F_4 -control condition (59.2 and 74.6, respectively). These differences in ratings likely obtained because exemplar 000x broke expected pairwise correlations between the cause features and the unobserved common-effect and exemplar 111x preserved those correlations.

Because the common-cause and F_1 -control conditions tested different exemplars than the common-effect and F_4 -control conditions, these two sets of conditions were analyzed separately. A two-way ANOVA of the common-cause and F_1 -control conditions confirmed a significant interaction between condition and exemplar ($F(7, 364) = 12.74$, $MSE = 109.5$, $p < .0001$) reflecting the fact that the pattern of ratings given to exemplars differed in the two conditions (Figure A1a). Likewise, in a two-way ANOVA of the common-effect and F_4 -control conditions there was a significant interaction between condition and exemplar ($F(7, 364) = 13.87$, $MSE = 173.3$, $p < .0001$) (Figure A1b).

As in the first experiment multiple regressions were performed on each participant's categorization ratings. Because in the common cause and F_1 -control conditions exemplars never displayed a value for the first feature, the eight predictors involving f_1 and its interactions used in the first experiment (i.e., f_1 , f_{12} , f_{13} , f_{14} , f_{123} , f_{124} , f_{134} , f_{1234}) were omitted as predictors in those conditions. Likewise, because in the common effect and F_4 -control conditions exemplars never displayed a value for the fourth feature, the eight predictors involving f_4 and its interactions (i.e.,

$f_4, f_{14}, f_{24}, f_{34}, f_{124}, f_{134}, f_{234}, f_{1234}$) were omitted in those conditions. The regression weights averaged over participants in each conditions are presented in Figure A2.

Figure A2 indicates that the presence of causal knowledge influenced the importance of individual features and interactions among features relative to the control conditions. For example, in both common-cause and common-effect conditions the three observable features were weighed much more heavily than in the respective control conditions. In the common-cause condition the greater weights for the effect features $F_2, F_3,$ and F_4 (Figure A2a) arose because their presence constitutes evidence for the presence of the cause $F_1,$ and F_1 was described as defining of category membership. Said differently, when a feature is described as being caused by another feature, categorizer's expect that feature to be more prevalent among category members. The feature constitutes stronger evidence in favor of category membership as a result.

Note that the greater feature weights for features $F_2, F_3,$ and F_4 correspond to the two-way interaction terms between features directly-connected by causal relationship in the first experiment's common-cause condition (i.e., $f_{12}, f_{13},$ and f_{14}). That is, the presence of an effect feature confirms expected correlation between it and its common cause. In the present experiment this effect manifests itself as weights on the individual effect features rather than interactions between those features and the common cause because in the current experiment the common cause was not directly observable.

Analogous to the feature weights in the common-cause condition, weights for features $F_1, F_2,$ and F_3 in the common-effect condition (Figure A2b) were greater than in the corresponding F_4 -control condition. These greater weights arose because the presence of these cause features constitutes evidence for the presence of the common effect $F_4,$ and F_4 was described as defining to category membership. These greater feature weights correspond to the two-way interaction terms between features

directly-connected by causal relationship in the first experiment's common-effect condition (i.e., f_{14} , f_{24} , and f_{34}). That is, the presence of a cause confirms the causal relationship between it and its common effect even though the common effect is not directly observable.

Statistical analysis of the regression weights shown in Figure A2 confirms the effect on feature weights. A 2 (condition) by 3 (feature: 2, 3, or 4) ANOVA of the feature weights of the common-cause and F_1 -control conditions, and a 2 (condition) by 3 (feature: 1, 2, or 3) ANOVA of the common-effect and F_4 -control conditions each revealed a main effect of condition, $F(1, 52) = 26.12$, $MSE = 45.5$, $p < .0001$, and $F(1, 52) = 39.81$, $MSE = 53.7$, $p < .0001$, respectively. There was no significant effect of feature or interaction between feature and condition in either of these analyses.

Figure A2 also presents the regression weights on interactions between features. Earlier it was shown how a fixed-cause common-cause schema, unlike a varying-cause common-cause schema, predicts no correlations between effect features. Consistent with this prediction, Figure A2a indicates that interactions between effects features in the common-cause condition (i.e., f_{23} , f_{24} , f_{34} , and f_{234}) did not differ from those in the F_1 -control condition ($F < 1$). In contrast, Figure A2b reveals the presence of interactions among cause features (i.e., f_{12} , f_{13} , f_{23} , and f_{123}) in the common-effect condition. In 2 (common-effect vs. F_4 -control) by 4 (interaction term: f_{12} , f_{13} , f_{23} , f_{123}) ANOVA of the regression weights there was a significant interaction indicating that the pattern of interaction weights differed between the two conditions, $F(3, 156) = 3.41$, $MSE = 15.1$, $p < .05$. As in the first experiment, to gain understanding of the significance of these interactions the logarithm of category membership ratings were plotted as a function of the number of effect features present in the common-cause condition, and as a function of the number of cause features present in the common-effect condition (Figure A3). As in the first experiment, whereas the common-cause condition shows a linear increase in (the

logarithm of) categorization ratings as a function of the number of effect features (Figure A3a), in the common-effect condition categorization ratings exhibited a discounting effect such that ratings increased more with the introduction of the first cause feature as compared to subsequent cause features (Figure A3b). Taken together, the disanalogies between the common-cause and common-effect results shown in Figures 9 and 10 constitute evidence for participants perceiving causal relationships as asymmetrical.

Theoretical Modeling: Causal-Model Theory

To fit causal-model theory to the current results, the likelihood equations presented in Table 4 were adapted to account for the presence of dimensions with unknown feature values. In general, likelihood equations for exemplars with unknown dimensions can be derived by summing over the likelihoods associated with each possible feature value. For example, the likelihood equations for the eight test exemplars with an unknown dimension in the common-cause condition can be derived from the common-cause likelihood equations in Table 4 by summing across the two possible values on the first dimension. For instance, $\underline{L}_{CC}(x000) = \underline{L}_{CC}(0000) + \underline{L}_{CC}(1000)$. However, because in this condition feature F_1 was described as occurring 100% of the time $\underline{L}_{CC}(0000) = 0$ and hence $\underline{L}_{CC}(x000) = \underline{L}_{CC}(1000)$. Note that because the likelihood of any exemplar with F_1 missing is defined to be 0, the likelihood equations for the eight remaining category exemplars with nonzero likelihoods need to be renormalized so they sum to 1. In the common-cause condition the result of this renormalization is for parameter c to be eliminated, a result expected given that a fixed common-cause schema entails $\underline{c} = 1$.

Similarly, the likelihood equations for the eight test exemplars with an unknown dimension in the common-effect condition can be derived from the common-effect likelihood equations in Table 4 by summing across the two possible

values on the fourth dimension, for example $\underline{L}_{CE}(000x) = \underline{L}_{CE}(0000) + \underline{L}_{CE}(0001)$. But, because feature F_4 was described as occurring 100% of the time, $\underline{L}_{CE}(0000) = 0$ and hence $\underline{L}_{CE}(x000) = \underline{L}_{CE}(1000)$. Again, because the likelihood of all exemplars with F_4 missing is defined to be 0, the likelihood equations for the eight possible category exemplars are renormalized so as to sum to 1.

However, fitting the causal models directly to the current data is problematic because the omission of F_1 and F_4 means that the models are underconstrained by the data. For example, failure to observe F_1 in the common cause condition means that the increased weight associated with the effect features can be accounted for either by increasing the efficacy of the causal mechanism linking them to the common cause (by adjusting parameter \underline{m}) or by the efficacy of their background causes (by adjusting parameter \underline{b}). Similarly, failure to observe F_4 in the common effect condition means that the increased weight associated with the cause features can be accounted for by either by increasing the efficacy of the causal mechanism linking them to the common effect (by adjusting parameters \underline{m} or \underline{b}) or by their base rates (by adjusting parameter \underline{c}).

Nevertheless, to demonstrate that the common-cause and common-effect causal models predict the basic pattern of empirical results, I investigated whether these results would be generated with the parameter values that were estimated in the first experiment. Because the same procedure and materials were used in the two experiments, the average parameter values obtained in the first experiment's common-cause and common-effect conditions should approximate performance of those same groups in the second experiment. To this end, the current common-cause and common-effect group-level categorization ratings were predicted with values $\underline{m} = .214$, and $\underline{b} = .437$ in the common-cause condition (\underline{c} is absent because of renormalization), and $\underline{c} = .522$, $\underline{m} = .325$, and $\underline{b} = .280$ in the common-effect condition (see Table 5). Because of the possibility of differences in scale usage between the two

experiments, a new scaling constant K was estimated in each condition. The resulting estimate of K in the two conditions are presented in Table A2, along with measures of the degree of fit of the models. For the model fits presented in Table A2 the number of data points (N) was 8. For causal-model theory, the number of parameters (P) was 1 and for the prototype model the number of parameters was 2.

It was shown earlier in Figure A1a that the common-cause group gave a lower category membership rating to exemplar x000 and a higher rating to exemplar x111. Figure A1a presents the fit of the common-cause causal model to these exemplars. According to causal-model theory, exemplar x000 is a poor category member because it is unlikely to be generated by a fixed common-cause category (because all the effects are missing), and x111 is a good category member because it is likely to be generated by a fixed common-cause category (because all the effects are present). Likewise, Figure A1b indicates that common-effect model predicts a low categorization rating to exemplar 000x because that pattern of features is unlikely to hold for a fixed common-effect category (because all the causes are missing), whereas 111x is a good category member because all the causes are present.

Following the first experiment, to analyze the predictions of causal-model theory's fits in terms of feature weights and interactions between features, multiple regressions were performed on the predicted ratings, and are presented in Figure A2 superimposed on the observed data. Figure A2 confirms that causal-model theory reproduces participants' increased weight on the effect features in the common-cause condition, and on the cause features in the common-effect condition. Moreover, Figure A2 indicates that causal-model theory is also able to account for the interactions among features that distinguish the common-effect condition from the common-cause condition. Figure A3 presents the predicted category membership ratings in the common cause condition (Figure A3a) and the common effect condition (Figure A3b) as a function of the number of effect and cause features, respectively. As

the figure indicates, causal-model theory reproduces the observed linear increase in ratings (in log coordinates) as one adds effect features to a fixed common-cause category, and the observed nonlinear increase in ratings as one adds cause features to a fixed common-effect category. This difference in the pattern of predictions in the two conditions indicates that causal-model theory, like the undergraduate participants, is sensitive to the asymmetry inherent in causal relationships.

Statistical test of the quality of causal-model theory's fits revealed that predicted ratings did not differ from observed ratings in the common-cause condition, $\chi^2(8) = 10.4$. Although the difference between predicted and observed ratings just reached significance in the common-effect condition, $\chi^2(8) = 15.8$, $p < .05$, note that fits in both conditions were obtained with parameter estimates derived from a different group of subjects, namely those in the first experiment.

Theoretical Modeling: Prototype Model

As in the first experiment, an important question is whether the classification performance observed in the second experiment could be explained instead by a similarity-based model such as a prototype model that includes extra dimensions that encode whether expected correlations are preserved or broken. On the one hand, the Extended-Features Prototype Model presented earlier can be in principle applied to the data from the second experiment despite the fact that exemplars had dimensions with missing values, because there are well-known techniques for computing similarity between exemplars with missing dimensions. As described in detail in Appendix C, I assume that similarity is computed according to an intersection rule (Estes, 1994) in which a dimension has no influence on similarity when a value on that dimension is unknown in one of the exemplars being compared. However, the unknown value on the common-cause dimension in the common-cause condition, and on the common-effect dimension in the common-effect

condition, means that the values of the higher-order dimensions representing whether expected correlations are preserved or broken are also unknown. For example, in the common-cause condition exemplar x101 would be encoded as x101xxx, because the presence or absence of the common cause would be unknown, as would whether the three expected pairwise correlation are preserved or broken. As described in Appendix C, the result is that the Extended-Features Prototype Model reduces to a simple prototype model when applied to the categorization data of the second experiment.

Use of the intersection rule also means it is straightforward to apply the Exemplar-Fragments Model. However, the unknown values on the common-cause and common-effect dimensions means that the exemplar fragments assumed to be stored in memory by the Exemplar-Fragments Model can no longer account for differences in category membership ratings that arises as a consequence of the presence or absence of causal knowledge. For example, as described in Appendix D, in the common cause condition exemplar x101 is as similar to the set of stored common-cause exemplar fragments (i.e., 11xx, 1x1x, 1xx1, 00xx, 0x0x, 0xx0) as is exemplar x010. As a consequence, the Exemplar-Fragments Model also reverts to a simple prototype model.

As discussed earlier, an important characteristic of prototype models is that they are unable to account for correlations among features in judgments of category membership. Thus, on the one hand a simple prototype model should be able to fit the data from the current common-cause condition by adjusting per-dimension saliency parameters. On the other hand, a prototype model is in principle unable to fit the data from the current common-effect condition in which interactions among features were observed (Figures A1 and A2). To demonstrate this fact, a simple prototype model was fit to the common-cause and common-effect conditions. To make these fits comparable to those of causal-model theory, the fits were carried out

on the aggregated group data. Note that because in both the common-cause and common-effect condition the saliency of the three effect and cause features, respectively, did not differ, a single saliency parameter s was used in each model fit (see Appendix C). The prototype model fits are presented in Table A2.

As expected, the prototype model was able to account for the increased sensitivity to the presence or absence of features. For example, in the common-cause condition the fits to exemplars $x000$ and $x111$ were 40.5 and 96.5, reflecting the generally lower rating given to $x000$ and the higher rating given to $x111$ by common-cause participants relative to the F_1 -control group. Likewise, in the common-effect condition the fits to exemplars $000x$ and $111x$ were 39.0 and 99.8, reflecting the generally lower rating given to $000x$ and the higher rating given to $111x$ by common-effect participants relative to the F_4 -control condition.

However, because the Prototype Model is unable to account for interactions among features, it is in principle unable to account for the results from the common-effect condition. Figure A4 present the fits of a prototype model to data presented earlier in Figure A3. As the figure indicates, whereas the Prototype Model does a fair job of accounting for the common-cause categorization ratings as a function of the number of effect features (Figure A4a), it is unable to account for the common-effect categorization ratings as a function of the number of cause features (Figure A4b), because it is unable to account for the nonlinear increase in categorization ratings as the number of cause features increases.

The failure of the Prototype Model to fit the common-effect data is reflected in measures of the degree of fit. Table A2 indicates that the Prototype Model produced worse fit according to RMSE (6.4) as compared to causal-model theory (4.7), and chi-square test revealed a significant difference between observed and predicted values in that condition ($\chi^2(8) = 19.9$, $p < .05$). Note that the better fits of causal-model theory were obtained even though the parameter estimates it used came from a different

group of subjects. As expected, the predictions of the Prototype Model to the common-cause ratings did not differ significantly from the observed ratings.

Discussion

One purpose of this additional experiment was to simulate the causal structure of many real-world categories in which a defining feature presumed to be present in all category members is thought to be causally responsible for observable features. A causal structure based on a defining feature a fixed common-cause schema, and, unlike the varying-cause common-cause schema used in the first experiment, a fixed common-cause schema implies no correlations among effect features. As predicted, in the (fixed) common-cause condition participants treated effect features as if they were independent, that is, their category membership ratings exhibited no sensitivity to interactions among the effects.

Another common characteristic of defining features is that they are not directly observable, and the current experiment's common cause condition also tested the effects on categorization performance when the defining common cause feature was unobservable. The results suggested that participants engaged in a form of causal reasoning while categorizing. That is, the fact that the presence of an effect feature was taken as strong evidence in favor of category membership suggests that categorizers reasoned from the presence of the effect to the presence of the unobserved common cause. The result was that as evidence in favor of the common cause increased (in the form of greater number of effect features), so too did categorization ratings. In contrast, in the F_1 -control condition those same features had less influence on category membership ratings, a result that obtained presumably because the absence of causal relationships meant that participants had no basis for reasoning from those features to the defining feature F_1 .

Similarly, the presence of a cause feature in the (fixed) common-effect

condition was taken as stronger evidence in favor of category membership than in the F_4 -control condition, because common effect participants reasoned from the presence of a cause feature to the presence of the defining common-effect feature. Unlike a fixed common-cause schema however, a fixed common-effect schema also implies higher-order correlations among the cause features. As predicted, in the common-effect condition participants exhibited a discounting effect such that the presence of the first cause feature was taken as stronger evidence in favor of the common-effect (and hence category membership) than additional causes.

Once again, this sensitivity to interactions among cause features in the common-effect condition but not effect features in the common-cause condition indicates a disanalogy between the common-cause and common-effect conditions, reflecting again the importance of treating causal relationships as an asymmetrical relationship. This disanalogy can also be understood in terms of the different patterns of causal reasoning that participants engage in to infer a hidden common cause versus a hidden common effect. For example, one can confidently infer the presence of a hidden common effect on the basis of a single cause even when other causes are known to be absent, because the presence of the common effect does not imply the presence of any of the other causes. In contrast, one should be less confident inferring the presence of a hidden common cause on the basis of only a single effect when other effects are known to be absent, because if the common cause is indeed present it implies the presence of those other effects. The likelihood equations derived for the fixed common-cause and common-effect embody these different patterns of causal reasoning for the common-cause and common-effect networks.

Quantitative model fitting confirmed that each of the empirical effects observed in the current experiment are predicted by fixed common-cause and common-effect causal models. In addition, it was shown that the extensions to

similarity-based prototype and exemplar models proposed following the first experiment reverted to simple prototype models. As such, these models provide no principled account of why observable features should be weighed more heavily in the common-cause and common-effect conditions than in their corresponding control conditions. Moreover, a simple prototype model is in principle unable to account for interactions among observable features, such as those found in the common-effect condition.

Appendix C

Fitting the Extended-Features Prototype Model to the Second Experiment

In contrast to the first experiment in which all rated exemplars possessed complete information about the presence or absence of features on all four dimensions, exemplars in the second experiment were missing information on the common cause dimension in the common-cause condition, and on the common effect dimension in the common-effect condition. As a result of this missing feature information, the values of the higher-order features on the fifth, sixth, and seventh dimensions are also unknown. On the one hand, accommodating missing features is straightforward in similarity-based models, and I follow Estes (1994) in assume similarity is computed by an “intersection rule” in which dimensions with missing features are simply omitted from the similarity computation. However, the result is that there are no estimates for s_1 , s_5 , s_6 , and s_7 in the common-cause condition, or for s_4 , s_5 , s_6 , and s_7 in the common-effect condition. As a consequence, the Extended-Prototype Model reverts to a simple prototype model with three dimensions when applied to the results of the second experiment.

Following the fits of causal-model theory, this simple prototype model was fit to the second experiment’s group-level data rather than to individual subject data. Because results from the regression analyses (Figure A2) suggested no difference in the saliency of individual features in both the common-cause and common-effect conditions, a single saliency parameter s replaced parameters s_2 , s_3 , and s_4 in the common-cause condition, and s_1 , s_2 , and s_3 in the common-effect condition. As a result, model fitting in both conditions of the second experiment involved two parameters: \underline{K} and s .

Appendix D

Fitting the Exemplar-Fragments Model to the Second Experiment

Use of the intersection rule in computing similarity means that the missing feature information in second experiment (F_1 in the common-cause condition, and about F_4 in the common-effect condition) means that dimension 1 has no influence on similarity computations in the common-cause condition and dimension 2 has no influence on similarity in the common-effect condition. However, the result is that the exemplar fragments representing causal knowledge are no longer able to account for differences in exemplars' category membership ratings as a function of whether those exemplars preserve or break expected correlations. For example, in the common-cause condition the similarity of exemplar x101 to the six exemplar fragments is,

$$\begin{aligned} \text{Sim}_{\text{Exemplars}}(\text{x101}) &= \text{Sim}(\text{x101}, 11\text{xx}) + \text{Sim}(\text{x101}, 1\text{x1x}) + \text{Sim}(\text{x101}, 1\text{xx1}) + \\ &\quad \text{Sim}(\text{x101}, 00\text{xx}) + \text{Sim}(\text{x101}, 0\text{x0x}) + \text{Sim}(\text{x101}, 0\text{xx0}) \\ &= 1 + s_3 + 1 + s_2 + 1 + s_4 \\ &= 3 + s_2 + s_3 + s_4 \end{aligned}$$

For an exemplar with exactly the opposite set of features, x010, the similarity to the six exemplar fragments is,

$$\begin{aligned} \text{Sim}_{\text{Exemplars}}(\text{x010}) &= \text{Sim}(\text{x010}, 11\text{xx}) + \text{Sim}(\text{x010}, 1\text{x1x}) + \text{Sim}(\text{x010}, 1\text{xx1}) + \\ &\quad \text{Sim}(\text{x010}, 00\text{xx}) + \text{Sim}(\text{x010}, 0\text{x0x}) + \text{Sim}(\text{x010}, 0\text{xx0}) \\ &= s_2 + 1 + s_4 + 1 + s_3 + 1 \\ &= 3 + s_2 + s_3 + s_4 \end{aligned}$$

That is, the effect of the presence of the exemplar fragments is to add a constant to each exemplar's category membership rating. In other words, like the Extended-Prototype Model, the Exemplar-Fragments Model converts to a version of a simple prototype model when the common cause and common-effect features are unobservable.

Table A1

Categorization ratings observed the second experiment, and the ratings predicted by causal-model theory in the common-cause and common-effect conditions. Standard errors for observed ratings are shown in parentheses.

Exemplar	Common Cause		F ₁ Control	Exemplar	Common Effect		F ₄ Control
	Observed	Predicted	Observed		Observed	Predicted	Observed
x000	37.6 (3.0)	42.2	52.8 (3.5)	000x	29.1 (4.0)	26.7	59.2 (4.0)
x001	56.9 (2.5)	55.1	61.5 (3.8)	001x	58.9 (4.0)	53.5	63.9 (3.2)
x010	55.8 (2.2)	55.1	63.6 (3.6)	010x	57.9 (3.5)	53.5	59.6 (3.2)
x100	57.0 (1.8)	55.1	66.3 (3.1)	100x	61.1 (3.6)	53.5	64.7 (3.5)
x011	71.8 (2.2)	72.1	66.9 (3.3)	011x	72.4 (2.9)	76.4	66.3 (3.1)
x101	73.7 (2.0)	72.1	67.1 (2.7)	101x	74.8 (2.7)	76.4	68.1 (3.5)
x110	72.2 (2.0)	72.1	69.7 (3.1)	110x	74.9 (2.4)	76.4	66.1 (3.1)
x111	92.6 (1.7)	94.2	76.1 (4.5)	111x	91.8 (2.0)	96.6	74.6 (3.8)

Table A2

Model fitting results from the second experiment. RMSE = Root mean square error.

Parameters	Common-Cause Model	Common-Effect Model
<u>Causal Model Theory</u>		
<u>K</u>	521	513
RMSE	3.3	4.7
Chi Square	10.4	15.8
<u>Prototype Model</u>		
<u>s</u>	.765	.759
<u>K</u>	94.2	96.0
RMSE	2.3	6.4
Chi Square	5.5	19.9

Figure A1

Categorization ratings from the second experiment for selected exemplars. (a) In the common-cause and F_1 -control conditions. (b) In the common-effect and F_4 -control conditions. Ratings predicted by causal-model theory are shown in each panel.

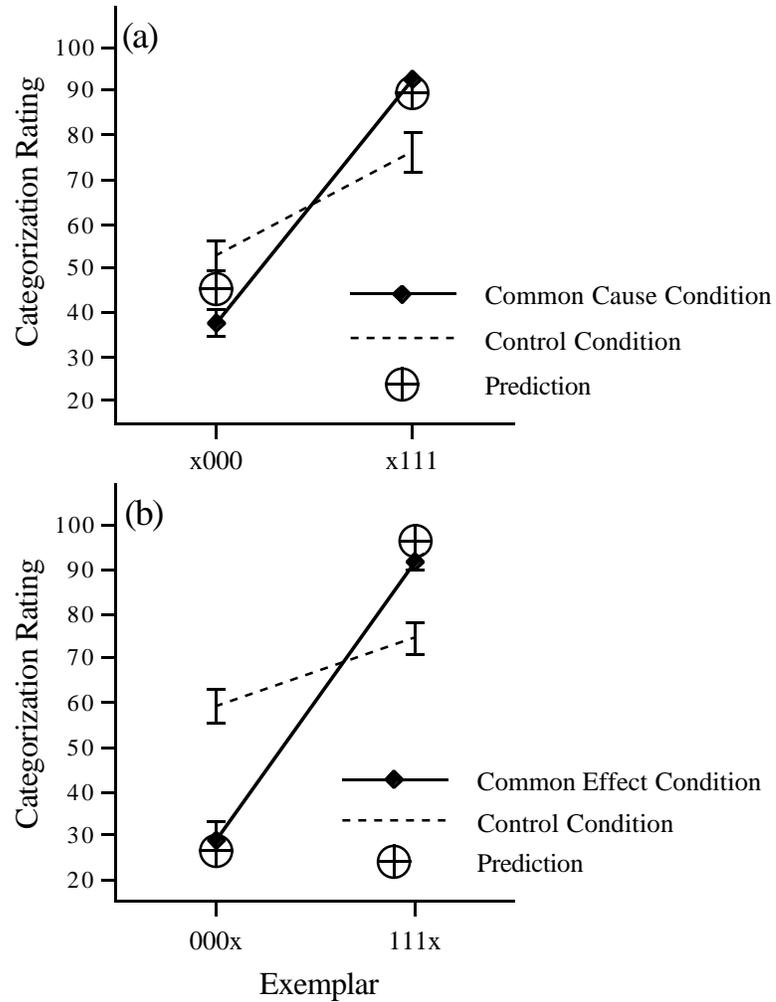


Figure A2

Regression weights for the second experiment. (a) For the common-cause and F_1 -control conditions. (b) For the common-effect and F_4 -control conditions. Weights predicted by causal-model theory are shown in each panel.

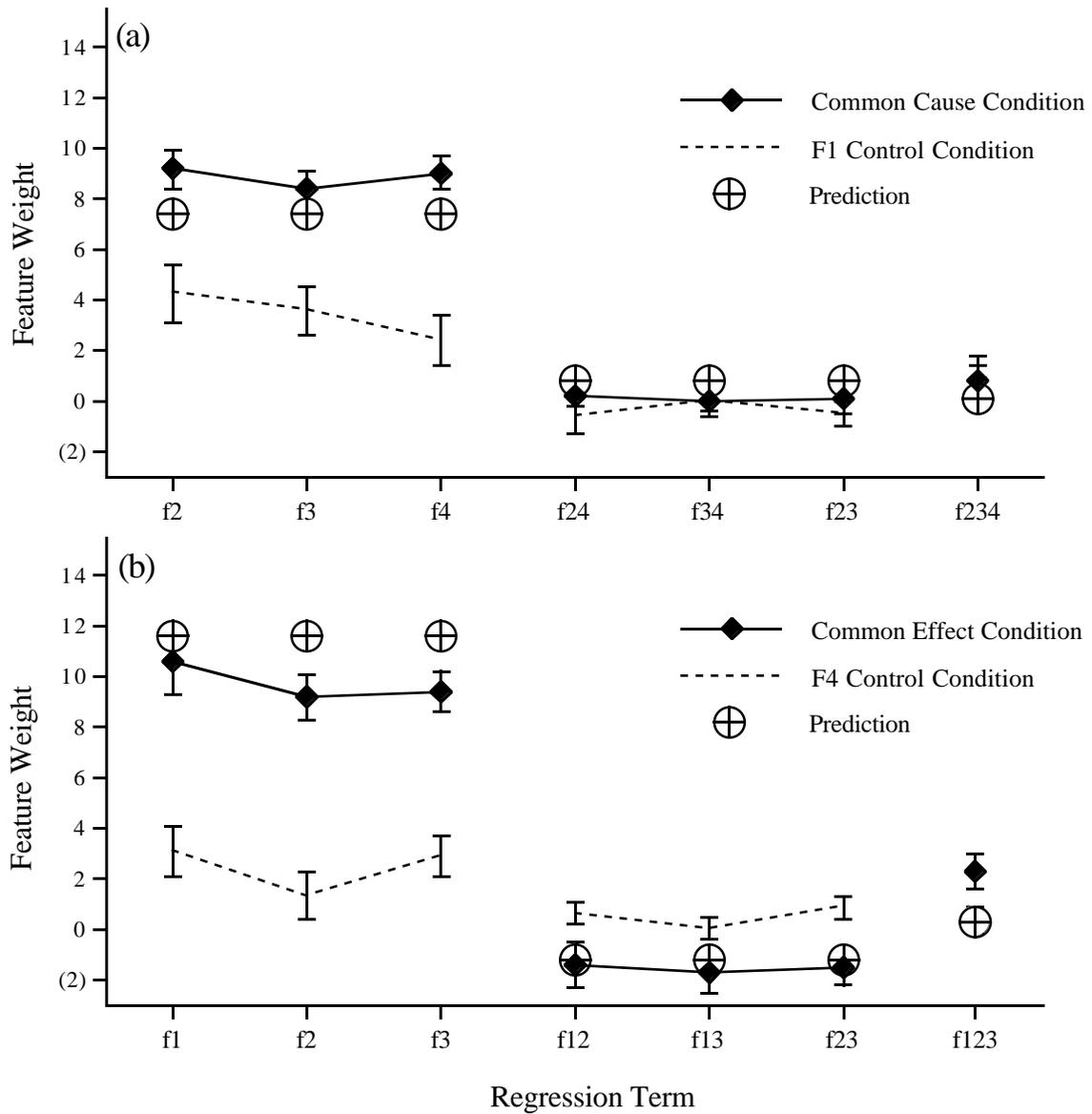


Figure A3

The logarithm of observed categorization ratings from the second experiment. (a) In the common-cause condition as a function of the number of effect features. (b) In the common-effect condition as a function of the number of causes. Ratings predicted by causal-model theory are shown in each panel.

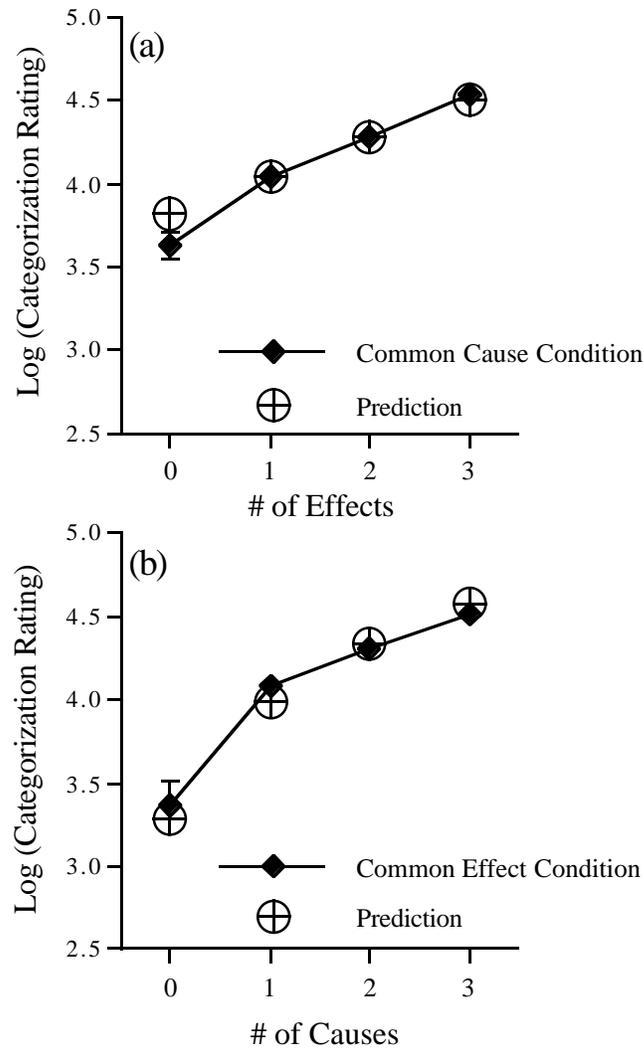


Figure A4

Fits of a simple prototype model to the data presented in Figure A3.

