A Piecemeal Processing Strategy Model for Causal-Based Categorization

Guillermo Puebla (guillermo.puebla@ed.ac.uk)
Department of Psychology, School of PPLS, The University of Edinburgh
Edinburgh EH8 9JZ, United Kingdom

Sergio E. Chaigneau (sergio.chaigneau@uai.cl)
School of Psychology, Universidad Adolfo Ibáñez
Av. Presidente Errázuriz 3328, Las Condes, Santiago, Chile

Abstract

Over the last 20 years, causal-model theory has produced much knowledge about causal-based categorization. However, persistent violations to the normative causal-model theory are prevalent. In particular, violations to the Markov condition have been repeatedly found. These violations have received different explanations. Here, we develop a model that starts from generally accepted cognitive phenomena (e.g., processing limitations, the relevance of inference in cognitive processing) and assumes that people are not fully causal nor fully associative when performing causal-based categorization, offering a new explanation for Markov violations.

Keywords: causal-based categorization; causal-model theory; causal inference; Markov condition

Introduction

Causal-model theory has taught us much about causal cognition (see Rehder, 2017; Sloman & Lagnado, 2015). However, problematic violations of the theory's predictions persist. The most important violation is that of the Markov condition. This condition states that when the state of a variable's immediate causes is known, then that variable is rendered conditionally independent of all its non-descendants (Pearl, 2000). Illustrative examples of violations are found in Rehder and Burnett (2005), and in Puebla and Chaigneau (2014). There, participants needing to infer the state of an unknown variable used information about other properties, even if those properties were conditionally independent from the unknown variable.

Some authors explain these violations by arguing that they are only apparent, because subjects do not necessarily use the same causal model specified by experimenters (Park & Sloman, 2013). Other authors argue that people may resort to associationist thinking and interpret directed causal links as bidirectional associations (Rehder, 2014). In contrast, here we hypothesize that violations occur because people may combine a partial understanding of causality with underlying similarity-based processing. In what follows, we present a process model of causal categorization, use it to make specific predictions about properties' conceptual weights, and fit it to empirical data.

Informed probabilities influence property-weights

Prior research consistently reports that when people need to infer a central property for categorization (i.e., because they lack information about the state of that property), properties that are causally related to the absent property increase their relevance for categorization proportionately to their informativeness about the missing property (Chaigneau, Barsalou, & Sloman, 2004; Puebla & Chaigneau, 2014; Rehder & Kim, 2009). In the current work we extend these findings to conditions in which the central property's state is explicitly known.

We hypothesize that even if a central property's state is made explicit, there may still remain some uncertainty regarding the property's true state. Thus, other causally related properties may acquire their weight depending on their contribution to decreasing that uncertainty. This idea is discussed in Rehder and Burnett (2005), and preliminary evidence for it can be found in Chaigneau et al. (2004, Exp. 7). In causal-model research, information about a property's inferential contribution is generally provided in the form of probabilities of effects given causes (i.e., p(effect|cause)). Consequently, we assume that when cues indicate that a given property is central, other causally linked conceptual properties acquire their weight as a function of how informative they are of the central property. In particular, in our Exp. 1 we used a causal chain model (A→B→C; with an additional D property which was not causally linked to other variables), and told participants that property C was central (i.e., it gave the category its name), with the expectation that its directly linked properties (i.e., B) would acquire their weight proportional to their p(effect|cause), and that its indirectly linked properties (i.e., A) would be weighted proportional to their probabilistic contribution to the central property's direct causes (i.e., B). Note here that we are assuming that people are intransitive when using causal models to categorize (Johnson & Ahn, 2015).

Making the last property in a causal chain the central property is representative of many categories that are defined by their functions. For artifacts (Carrara & Mingardo, 2013; Chaigneau et al., 2004) and for functionally conceptualized natural kinds (e.g., Barsalou, Sloman, & Chaigneau, 2005; Lombrozo & Rehder, 2012), the goals that they achieve in their normal settings are central for their classification (e.g., an artificial heart is believed to belong to the heart category depending on it being able to pump blood to a greater extent than on it using any particular physical mechanism to achieve that goal).

Cognitive limitations

In typical causal classification experiments participants need to integrate several pieces of information, e.g., information about the direct causal links and their associated probabilities (p(effect|cause)), the indirect causal links (two-way relations that are mediated by other properties), and also the particulars of the materials provided. Researchers generally assume that people are able to integrate all this information. In fact, as discussed in (Rehder, 2003a), the causal-model theory assumes that people classify entities as category members to the extent that the entity's distribution of properties would be expected from the category's ideal causal model.

In contrast, in our model we hypothesize that people simplify their task by analyzing information in a piecemeal fashion (thus, we call it the Piecemeal Strategy Model or PSM). In particular, we assume that they only evaluate pairs of directly connected properties, and that unconnected properties are considered in isolation (e.g., in the causal chain model A→B→C, with D as an isolated property, subjects would separately evaluate A→B, B→C and D). Regarding the type of computation subjects perform, we assume that they consider each directly connected pair (and each isolated property) in the ideal model presented to them, as a separate prototype with which to compare the particular instances they need to judge. To implement these ideas, we used Nosofsky (1992) Multiplicative Prototype model (MPM). This implies computing a distance, as given by,

$$\delta_{XY} = \sum_{i=X}^{Y} \left(\frac{p_i}{p_X + p_Y} \right) |x_i - M_i| \quad (1)$$

where X and Y are two directly connected properties in the causal model, p_i is the inferential contribution of a property, x_i corresponds to the state of the ith property in the currently considered instance, and M_i corresponds to the ideal state of the i$^{\text{th}}$ property in the causal model (i.e., the prototype). Note that the denominator inside the parenthesis allows Eq. 1 to comply with the MPM requirement that the weights (p_i) in the distance computation all add to one. For isolated properties (D in our scenarios), the corresponding distance is defined to be,

$$\delta_D = p_D |x_D - M_D| \quad (2)$$

where p_D is a free parameter estimated from the data, reflecting the inferential weight of the isolated property ($0 \leq p_D \leq 1$), x_D is the state of the D property in the currently considered instance, and M_D corresponds to the ideal state of the D property in the causal model (i.e., the prototype).

Distances cannot be considered by themselves, because they are linear. Similarity, in general, behaves like a generalization gradient (Shepard, 1987). For this reason, distances in Eqs. 1 and 2 need to be transformed into similarities by,

$$S_{XY} = e^{-b(\delta_{XY})} \quad (3)$$

where s_{XY} is similarity, and b is a sensibility parameter that determines the rate at which similarity falls with distance. In our model fitting, we fixed b = 1 (i.e., b was not estimated from the data). To compute the similarity s_D for the isolated D property, δ_{XY} is substituted by δ_D in Eq. 3.

Finally, we assume that the similarities from all the partial models under consideration are averaged to obtain an estimate of the overall similarity of the instance being judged relative to the prototype (i.e., the received causal model) by,

$$S_o = \frac{1}{n} \sum_{i=1}^{n} s_i \quad (4)$$

where so is the overall similarity of the instance being judged, n is the total number of separate pieces of information being considered (A→B, B→C, D), and s_i is the similarity according to Eq. 3. Because the PSM implies considering some properties twice (property B in the causal chain model), for modeling purposes we introduced an adjustment to pi simply by dividing it by 2 to reflect that those properties were being taken into account twice.

In summary, we propose that pairs of features that are causally related (and any features that are causally unrelated) are treated as features in separate multiplicative similarity prototype models, with classification ratings being a function of the averaged similarity of those feature pairs to their corresponding prototypes. A closely related model was proposed and tested by (Rehder, 2003a), but he concluded that the model failed to account for the data. In the Discussion section we will consider possible explanations for why our results suggest a different conclusion.

Experiments

Participants were trained on a causal model representing a given category, until they were able to answer correctly a set of 9 conditional and counterfactual questions. They then received the set of all possible combinations of present and absent properties involved in the causal model and were asked to rate how representative each combination was of the trained category.

Ratings were analyzed using the regression method (Rehder & Hastie, 2001). In this method, participants provide category membership ratings for all possible combinations of m properties in two possible states (present or absent), producing a total of 2^m combinations. For each combination, subjects provided a categorization rating on a 1 to 7 scale. When present and absent properties are coded respectively as 1 and -1 (i.e., effect coding), these values can be entered into individualized regression equations to predict a participant's categorization ratings. Furthermore, 2-way and higher-order interaction terms can be computed by entering the product of the corresponding property coded values as predictors into the equations. The corresponding regression coefficients can then be used as individual data points reflecting, across participants, the contribution of each predictor variable to the ratings.

Subjects were randomly assigned to one of two between-subjects conditions (domain: living things, artifacts) and provided data for two within-subjects conditions (information:

complete, incomplete). In the complete information condition, subjects received descriptions containing information about all properties (A, B, C, and D). In the incomplete information condition, subjects received descriptions lacking information about property C. Because prior research suggests that, in the context of causal classification the incomplete information condition promotes using other properties to infer the state of the unknown property (e.g., Puebla & Chaigneau, 2014), this design allowed us to compare conceptual properties' regression weights across the within-subjects condition. An increase in regression weights in the incomplete information relative to the complete information condition, would show that participants used a given property to infer the state of the unknown property C.

Predictions

The PSM makes the following predictions. Due to the piecemeal strategy, we predicted higher regression coefficients for directly connected properties interaction terms than for not directly connected properties. Furthermore, Eq. 3 predicts the type of interaction that we will find. People will prefer instances where properties X and Y are both in the same state as in the received model (e.g., X = 1, Y = 1), and any deviation (e.g., X = 1, Y = -1) will produce a large decrease in similarity (due to the *b* parameter). Note that all this means small interaction coefficients (i.e., smaller than main effect coefficients). This contrasts with predictions from the causal-model theory, where people are predicted to produce large interaction terms that are as large as main effects.

The PSM predicts that independent properties in our causal models will not interact. This is the same result that the causal-model theory would lead us to expect (i.e., properties A and C in the causal chain are independent conditional on the state of B). However, the PSM predicts this pattern of interactions, not because people conform to conditional independence principles, but because of the piecemeal simplification strategy. Thus, we expect our data from the complete information condition to only mimic adherence to the causal Markov condition. This should become evident in participants' performance in the incomplete information condition. When comparing regression weights across the information factor, the lack of information about the central C property should produce an increase in the regression weights of the independent properties (A and D in the causal chain) due to those properties being associated to the unknown property C. This is a violation of the Markov condition because only direct causes are normatively relevant to predict the state of the unknown property C. Thus, we predict an apparent adherence to Markov in the complete information condition, and a failure to adhere in the incomplete information condition.

Regarding the main effects, the PSM predicts that properties' conceptual weights will follow their inferentially derived weights (p_i). For the chain model in Exp. 1, we predict that regression coefficients for C will be greater than the average of A and B; D will be smaller than the average of A, B and C; and A will not be different from B.

Experiment 1

Design and Participants Exp. 1 followed a mixed factorial 2 (domain: living things, artifacts) x 2 (information: complete, incomplete) design, with the last being a within-subjects factor. Property D served as an inbuilt control condition for each subject and provided a baseline regression coefficient to which properties in the causal model could be compared. Also, D's interaction with other properties (AD, BD and CD) also provided a baseline for interaction terms' regression coefficients. Subjects (N = 66) were Adolfo Ibáñez University undergraduates (N = 41, males = 16) who participated for course credit, and undergraduate volunteers from other local universities (N = 25, males = 7).

Materials and Procedures The materials were verbal and graphical descriptions of two categories characterized by a chain causal structure. In the living things condition, materials described the structure of a fictional biological cell. In the artifacts condition, materials described the structure of a fictional particle accelerator. Stimuli were presented on screen by means of a locally programed software.

In the learning phase, participants were trained in the causal chain graph. Subjects learned that causes produced their effects with a 0.75 probability. Regarding property D, participants were informed that it occurred in category members with a probability of 0.75. Thus, property D was predictive of the category, but not causally related to the other properties. By keeping property D's probability equal to the conditional probabilities for the other properties, we kept everything other than belonging or not to the causal model constant for property D as compared to properties A, B and C. Importantly, subjects learned that property C gave the category its name (i.e., C was the central property).

In the classification phase, subjects had the causal graph in full view. In the complete information condition, participants received descriptions containing information about all properties either present or absent (16 combinations). In the incomplete information condition, participants received descriptions which lacked information about the state of the central property C (8 combinations). In total, participants classified 24 descriptions, presented in random order. For each description, subjects had to respond whether it was or not a member of the focal category using a 6-point rating scale.

Results Effect coding variables representing 10 variables per subject (4 main factors and 6 interactions, see Fig. 1) were entered as predictors in individualized regression equations with rating as dependent variable. The resulting individualized regression coefficients were submitted to a mixed 2 (domain: living things, artifacts) $\times$ 10 (coefficients) mixed ANOVA. The mixed ANOVA showed there was no effect of the domain factor ($F(1, 64)=1.37$, $MS_e=.04$, $p=.25$, $\eta p^2=.02$, power=.21) and it did not interact with the coefficients factor ($F(9, 576)=1.02$, $MS_e=.16$, $p=.39$, $\eta p^2=.02$, power=.30). Consequently, we collapsed this factor.

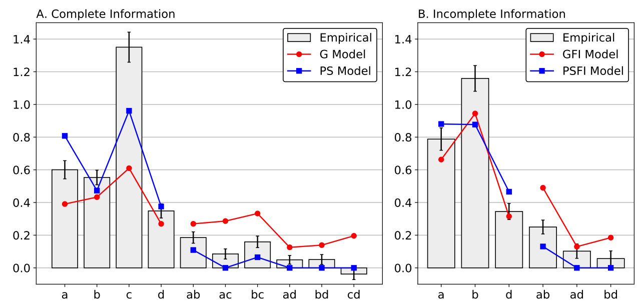

Figure 1: Regression weights for individual features and interactions of Exp. 1. Fits for the PSM (blue) and the GM (red) are superimposed on the data. Error bars are standard errors.

In accordance with our predictions for the main effects, orthogonal planned comparisons showed that the average coefficient for C was significantly greater than the average of coefficients for A and B ($F(1, 65)$=40.19, MS_e=.98, p<.001, ηp^2=.38, power>.99); the average coefficient for D was significantly lower than the average of coefficients for A, B and C ($F(1, 65)$=74.87, MS_e=.21, p<.001, ηp^2=.54, power>.99); and there was no difference between coefficients for A and B (F<1). This pattern of results for main effects suggests that subjects did indeed take information about p(effect|cause) as cue to properties inferential value, and that being in a causal structure increases inferential value beyond that of probabilistically related variables (property D).

Regarding the interaction coefficients, our predictions for the causal chain model were that the average of the AC interaction coefficients would be lower than the average of the AB and BC interaction coefficients (A and C are not directly connected in the causal graph); and that the average of the AB and BC interaction coefficients (directly connected properties) would be greater than the average of the AD, BD, and CD coefficients (i.e., our baseline conditions). As predicted, two non-orthogonal planned comparisons showed that the AC interaction was significantly smaller than the average of the AB and BC interactions ($F(1, 65)$=4.8, MS_e=.10, p=.03, ηp^2=.07, power=.58), and that the average of AB and BC interactions was significantly greater than the average of the AD, BD, and CD interactions ($F(1, 65)$=30.1, MS_e=1.81, p<.001, ηp^2=.32, power>.99).

Note that the low AC coefficient (which in fact was not significantly different from the interactions found for AD, BD and CD; $F(1, 65)$=3.8, MS_e=.07, p=.06, ηp^2=.06, power=.48), could be interpreted as participants complying with the Markov condition. However, analysis of the incomplete information condition reveals a different story. Under this condition, participants did not comply with Markov, using information about the state of property A (the screened-off property) and of property B (C's direct cause) to make inferences about the state of the missing C central property. Paired samples t tests revealed coefficients for properties A and B increased significantly when comparing the complete information condition with the incomplete information condition (respectively, complete information mean=0.60, incomplete information mean=0.79; $t(65)$=3.09, p=.003; complete information mean=0.55, incomplete information mean=1.16; $t(65)$=7.43, p<.001). In contrast, property D did not show evidence of being used to perform inferences about the state of property C (complete information mean=0.3485, incomplete information mean=0.3447; $t(65)$=.07, p=.94). Thus, data supported our hypothesis that subjects' performance in the complete information condition would mimic adherence to Markov.

Model fitting We fit the PSM to the classification ratings of Exp. 1. For comparison, we also fit the generative model (GM) of causal-based categorization (Rehder, 2003a; Rehder & Kim, 2009) (see Fig. 1). In the GM representation, a category k establishes a set of causal mechanisms. Each mechanism relates a feature j with its parent i operating with probability m_{ij} when i is present. Other background causes of j operate collectively with probability b_j. When j's parents operate independently, j's parents and the background causes produce j in members of category k conditional on the state of j's parents with probability,

$$p_k(f_i \mid Pa_k(f_j)) = 1 - (1 - b_j) \prod_{f_i \in Pa_k(f_i)} (1 - m_{ij})^{ind(f_i)} \quad (5)$$

where $ind(i)$ is an indicator variable that evaluates to 1 when i is present and 0 otherwise. The model assumes that root causes are independent of one another and the probability of each is represented with its own parameter c_j. The GM predicts that categorization judgments are a monotonic function of the joint distribution associated with the category's causal model,

$$p_k(f_{k,i}, \ldots, f_{k,N}) = \prod_{j=1\ldots N} p_k(f_i \mid p_{a_k}(f_i)) \quad (6)$$

Participants ratings were predicted as follows:

$$rating^{PSM}(o_i) = s_k(o_i; p_A, p_B, p_C, p_D)/\beta$$
$$rating^{GM}(o_i) = 6 p_k(o_i; c_A, b_B, b_C, b_D, m_{AB}, m_{BC})^{\gamma}$$

where β and γ are free parameters. We fit both models by searching for the parameter values that minimized the squared difference between the predicted ratings and the empirical ones. In the complete information condition both models achieved a high correlation with the ratings: r_{PSM} = .85, r_{GM} = .90. We used the Akaikes information criterion (AIC[1]) to compare the degree of fit of both models controlling for the different number of parameters. The bigger AIC for the GM (15.2) in comparison to the PSM (12.1) indicates that, in fact, the PSM provides a slightly better characterization of the data

[1]AIC = $ln(SEE/n) + 2(p+1)$ where SSE is the sum of squared error for a participant, n is number of data points fit, and p is the model's number of parameters.

in this condition. The best-fitting parameters for the PSM were: p_A = 0.325, p_B = 0.243, p_C = 0.775, p_D = 0.238, β = 0.094 and for the GM: c_A = 0.871, b_B = 0.802, b_C = 0.951, b_D = 0.765, m_{AB} = 0.558, m_{BC} = 0.327, γ = 0.565. Note that while both models achieve a similar level of fit to the data, the GM achieves this by assigning values to the causal relation parameters lower than participants were taught during training (0.75).

In the incomplete information condition (see Fig. 2), we adjusted both the PSM and the GM to take into account the unknown state of C. We did this by inferring the probability of C being present given the state of its parent B using the GM equations: $p(E=1 \mid C=1) = 1-(1-m_{CE})(1-b_E)$ and $p(E=1 \mid C=0) = b_E$. We treated this probability as the state of C and then proceeded as before for both models. In this condition the models achieved a high correlation with the ratings: r_{PSM} = .89, r_{GM} = .92. Again, we obtained a bigger AIC for the GM (15.0) in comparison to the PSM (12.7). The best-fitting parameters for the PSM were: p_A = 0.347, p_B = 0.297, p_D = 0.264, b_C = 0.463, m_{BC} = 0.392, β = 0.083 and for the GM: c_A = 0.865, b_B = 0.884, b_C = 0.933, b_D = 0.687, m_{AB} = 0.552, m_{BC} = 0.617, γ = 0.505. Note that the causal relation parameter for the relation between B and C was higher in this condition.

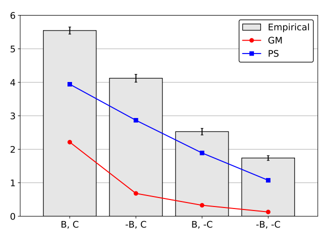

Figure 2: Average ratings for objects with different combinations of states (present or absent) for features B and C in Exp. 1, complete information condition. Fits for the PSM (blue) and the GM (red) are superimposed on the data. Error bars are standard errors.

Two things are noteworthy from these results. First, as shown in Fig. 1, the GM consistently overestimates the magnitude of the coherence effect (i.e., the 2-way interactions), while the PSM shows clearly better fits. Additionally, as shown in Fig. 2, the PSM is better able to predict the consequences of inconsistent information on participants ratings, as compared to the GM. This relates to similarity gradients implied by eq. (3).

Experiment 2

Because results like those of Exp. 1 are difficult to reconcile with causal-model theory, in particular the lack of a coherence effect, Rehder (2017) proposed that small property interactions in results like those of Exp. 1, occur because instructions and materials emphasized a single almost defining property (property C in Exp. 1). Had traditional category labels been used (i.e., a category name, such as "dog"), large interactions would emerge, as expected by causal-model theory. To test Rehder's (2017) hypothesis, in Exp. 2 we used the causal chain model, but subjects were not told that there was a central property that gave the category its name. Instead, an arbitrary category label was provided.

As there should be no inferential processes in this task, we predicted that all properties in the causal model would show about the same weight, and on average they would produce a greater regression weight than the isolated property D. Regarding the interactions, the PSM predicts that, because of the piecemeal strategy, directly connected properties (AB, BC) would exhibit a larger regression weight than the indirectly connected properties (AC), and that the AB and BC terms would show a higher regression weight than the interactions of not connected properties (AD, BD, CD).

Design and Participants Exp. 2's design was identical to that of Exp. 1. Subjects (N = 64) were Adolfo Ibáñez University undergraduates (males = 21) who participated for course credit.

Materials and Procedures Materials were identical to those used in Exp. 1. However, arbitrary names were used to label categories, and no property was described as central or described the category's function. Except for the arbitrary category name, procedures were identical to Exp. 1.

Results Results Individualized regression coefficients were submitted to a mixed 2 (domain: living things, artifacts) $\times$ 10 (coefficients) mixed ANOVA (see Fig. 3). The mixed ANOVA showed there was no effect of the domain factor (F<1) and it did not interact with the coefficients factor (F<1). Consequently, for all subsequent analyses we collapsed this factor.

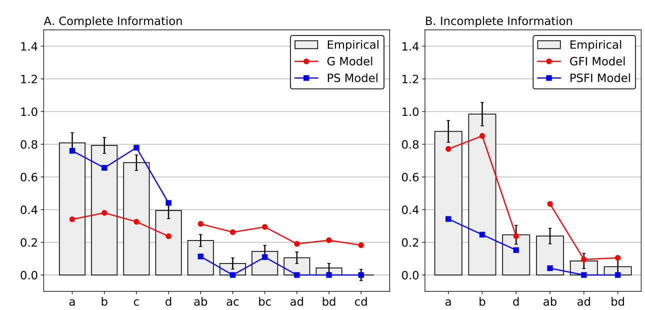

Figure 3: Regression weights for Exp. 2's individual features and interactions. Fits for the PSM (blue) and the GM (red) are superimposed on the data. Error bars are standard errors.

As predicted by the PSM, orthogonal planned comparisons showed that the average coefficient for D was significantly lower than the average of coefficients for A, B and C (F(1, 63)=44.72, MS_e=.19, p<.001, ηp^2=.42, power>.99); but that there were no significant differences between C versus A and B (F(1, 63)=3.46, MS_e=.24, p=.068, ηp^2=.05, power=.45), or A versus B (F<1). This pattern of results for main effects suggests that our subjects did indeed judge all properties to be about equally central. This result contrasts with Exp. 1, where inference induced differential property weights. However, as in Exp. 1, property D was judged to be less central than properties belonging to the causal model. Again, this shows that participants are sensitive to causal information and are not disregarding it by using a pure associative strategy.

As predicted by the PSM, two non-orthogonal planned comparisons showed that the AC interaction was significantly smaller than the average of the AB and BC interactions (F(1, 63)=7.02, MS_e=.42, p=.01, ηp^2=.10, power=.74), and that the average of AB and BC interactions was significantly greater than the average of the AD, BD, and CD interactions (F(1, 63)=15.05, MS_e=2.52, p<.001, ηp^2=.19, power=.97).

As in Exp. 1, the low AC coefficient suggests that participants are complying with the causal Markov condition. At odds with Exp. 1, participants did not use property A (the screened-off property) to infer the state of property C (complete information mean=0.82, incomplete information mean=0.89; t(63)=1.2, p=.24), but used property B (Cs direct cause) (complete information mean=0.79, incomplete information mean=1.0; t(63)=3.95, p<.001). Furthermore, in the incomplete information condition, participants relied less on the isolated property D to make inferences (complete information mean=0.38, incomplete information mean=0.25; t(63)=2.16, p=.04). These results are broadly consistent with the hypothesis that using an arbitrary category label would promote causal classification.

Model fitting We fit the PSM and the GM to the classification ratings of Exp. 2 in the same ways as in Exp. 1 (Fig. 3). In the complete information condition both models achieved a high correlation with the ratings: r_{PSM} = 0.83, r_{GM} = 0.84. The bigger AIC for the GM (15.8) in comparison to the PSM (11.9) indicates that the PSM provides a slightly better characterization of the data in this condition. The best-fitting parameters for the PSM were: p_A = 0.529, p_B = 0.455, p_C = 0.577, p_D = .293, β = 0.096 and for the GM: c_A = 0.921, b_B = 0.925, b_C = 0.887, b_D = 0.801, m_{AB} = 0.537, m_{BC} = 0.158, γ = 0.908. As in Exp. 1, while fits for both models are similar, the GM achieves this by assigning values to the causal relation parameters lower than participants were taught.

For the incomplete information condition, we adjusted the PSM and the GM as in Exp. 1. The models achieved a high correlation with the ratings: r_{PSM} = 0.88, r_{GM} = 0.90. Again, we obtained a bigger AIC for the GM (14.6) in comparison to the PSM (13.0). The best-fitting parameters for the PSM were: p_A = 0.316, p_B = 0.204, p_D = 0.236, b_C = 0.527, m_{BC} = 0.312, β = 0.230 and for the GM: c_A = 0.889, b_B = 0.847, b_C = 0.937, b_D = 0.640, m_{AB} = 0.439, m_{BC} = 0.592, γ = 0.431. Note that, as in Exp. 1, the causal relation parameter for the relation between B and C was higher in this condition.

Just as for Exp. 1, in Exp. 2 the GM consistently overestimates the magnitude of the coherence effect (i.e., the 2-way interactions in Fig. 3, particularly in Panel A), while the PSM shows clearly better fits. Finally, as shown in Fig. 4, the PSM is better able to predict the consequences of inconsistent information on participants ratings, as compared to the GM.

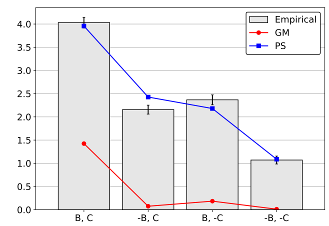

Figure 4: Average ratings for objects with different combinations of states (present or absent) for features B and C in Exp. 2. Fits for the PSM (blue) and the GM (red) are superimposed on the data. Error bars are standard errors.

General Discussion

In our experiments, the PSM was able to predict the pattern of results for main effects and interactions. Importantly, in both experiments the size of interaction coefficients remained low, and were not as high as those of main effects, as predicted by the GM. Furthermore, as would be expected if subjects were using associative mechanisms, the PSM was better able to predict ratings for objects with inconsistent information (Figs. 2 and 4). However, our results were not as clear regarding the mimicking Markov hypothesis. Exp. 1 produced data that is consistent with it, but Exp. 2 did not. In this latter experiment, participants appear to have complied with Markov both in the complete (i.e., low interaction coefficient for conditionally independent features) and in the incomplete information condition (i.e., appropriate screening-off of the conditionally independent distal cause). This pattern of results is consistent with the hypothesis that using an arbitrary category label enhances causal classification (Rehder, 2017). However, as in neither experiment did we obtain coherence effects, evidence for this hypothesis is mixed.

Prior research has found coherence effects in conditions similar to ours (e.g., in Rehder, 2003b, Fig. 4). The question then arises of how to account for these different results. In our experiments, we strove to use procedures as close as possible to those used by other researchers, so we tend to believe that differences do not lie in materials and procedures. Instead, we think it is possible that there are differences in how different populations handle causal information for categorization as well as for other tasks. Recently, using a causal inference task, (Rehder, 2018) found substantial variability in how individuals perform inferences (i.e., a single model was not able to account for the pattern of inferences of all participants, with a substantial minority behaving close to the predictions of an associative model). In a similar vein, we believe that no current model of causal cognition comfortably handles this variability and that future research should look to identify parameters that characterize tasks, individuals and populations in such a way that they are able to account for differences in causal categorization, and causal cognition in general.

References

Barsalou, L. W., Sloman, S. A., & Chaigneau, S. E. (2005). The hipe theory of function. In L. Carlson & E. van der Zee (Eds.), *Representing functional features for language and space: Insights from perception, categorization and development* (pp. 131–147). Oxford, England: Oxford University Press.

Carrara, M., & Mingardo, D. (2013). Artifact categorization. trends and problems. *Review of Philosophy and Psychology*, *3*(4), 351–373.

Chaigneau, S. E., Barsalou, L. W., & Sloman, S. A. (2004). Assessing the causal structure of function. *Journal of Experimental Psychology: General*, *133*(4), 601.

Johnson, S. G., & Ahn, W. (2015). Causal networks or causal islands? the representation of mechanisms and the transitivity of causal judgment. *Cognitive science*, *39*(7), 1468–1503.

Lombrozo, T., & Rehder, B. (2012). Functions in biological kind classification. *Cognitive psychology*, *65*(4), 457–485.

Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (pp. 149–167). Hillsdale, NJ: Erlbaum.

Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the markov property in causal reasoning. *Cognitive Psychology*, *67*(4), 186–216.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York, NY, USA: Cambridge University Press.

Puebla, G., & Chaigneau, S. E. (2014). Inference and coherence in causal-based artifact categorization. *Cognition*, *130*(1), 50–65.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science*, *27*(5), 709–748.

Rehder, B. (2003b). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1141.

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive psychology*, *72*, 54–107.

Rehder, B. (2017). Concepts as causal models: Classification. In M. R. Waldmann (Ed.), *The oxford handbook of causal reasoning* (pp. 347–375). New York, NY, USA: Oxford University Press.

Rehder, B. (2018). Beyond markov: Accounting for independence violations in causal reasoning. *Cognitive psychology*, *103*, 42–84.

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

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: the effects of causal beliefs on categorization, induction, and similarity. *Journal of experimental psychology. General*, *130*(3), 323-60.

Rehder, B., & Kim, S. (2009). Classification as diagnostic reasoning. *Memory & Cognition*, *37 6*, 715-29.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237 4820*, 1317-23.

Sloman, S. A., & Lagnado, D. A. (2015). Causality in thought. *Annual review of psychology*, *66*, 223-47.