Modulating the coherence effect in causal-based processing

Nicolás Marchant (nicolasmarchant@alumnos.uai.cl)
Center for Social and Cognitive Neuroscience, School of Psychology, Universidad Adolfo Ibáñez.
Av. Presidente Errázuriz, 3328, Las Condes, Santiago de Chile.

Sergio E. Chaigneau (sergio.chaigneau@uai.cl)
Center for Social and Cognitive Neuroscience, School of Psychology, Universidad Adolfo Ibáñez.
Av. Presidente Errázuriz, 3328, Las Condes, Santiago de Chile.

Abstract

Causal-based cognition is thought to be relevant for human beings because it allows inferring the unfolding of events. Theories of causal-based cognition offer researchers a way to understand inter-feature relations, above and beyond the purely associative relations posited by similarity theories. In the causal-model theory (a.k.a. the Generative Model), people are thought to categorize an exemplar depending on how likely its particular feature combination is, given the category's causal model. This mechanism predicts the coherence effect (i.e., when people categorize, features interact). This effect has been widely reported in the literature. In the current experiment, we sought to specify conditions that modulate the coherence effect. To that end, we implemented a between-subjects manipulation where participants had to judge either category membership or category consistency. Our results show that subjects exhibit a larger coherence effect in consistency condition. We discuss our results' relevance for causal-model theory and for the possibility of distinguishing causal-based from similarity-based processing.

Keywords: causal reasoning; coherence effect; causal-based categorization; similarity; exemplar models

Introduction

Attention to causal cognition has burgeoned during the last 20 years. Presumably, understanding causal relations in the world allows humans to infer their actions' effects (Holyoak & Cheng, 2011). Causal cognition offers researchers an alternative to associationist and similarity-based theories (Waldmann, Hagmayer, & Blaisdell, 2006). The contrast between causal-based and associationist explanations has played itself out repeatedly in the literature.

Currently, the most general proposal regarding mechanisms by which causal knowledge becomes relevant for categorization is Rehder's causal-model theory (a.k.a. the Generative Model, Rehder, 2003a; Rehder & Hastie, 2001). In causal-model theory, people classify exemplars as category members to the extent that the pattern of causes and effects they exhibit is likely given the category's causal model. A crucial aspect of causal-model theory is the prediction of the coherence effect, which serves as focus for the current work (Rehder, 2017; Rehder & Kim, 2006, 2010).

Consider that subjects know that in a given category A causes B (e.g., in *tropical frogs*, being poisonous causes them to have brightly colored skin). Imagine, also, that those subjects are shown all possible present and absent cause and effect combinations (i.e., AB, ¬AB, A¬B, ¬A¬B), and asked to rate each combination's category membership. Causal-model theory predicts that, given that if a cause (i.e., A) is not observed, then its effect (i.e., B) is also likely not to be observed, people should judge an exemplar showing the ¬A¬B pattern (e.g., a *tropical frog* that is not poisonous and does not have brightly colored skin) to be a good category exemplar because it preserves the learned causal structure (i.e., A → B) better than the ¬AB or A¬B feature combinations (Rehder, 2017).

There is abundant evidence in the literature showing the coherence effect (Rehder, 2017; Rehder & Kim, 2006; 2010; Hampton, Storms, Simmons & Heussen, 2009). Moreover, the coherence effect is a hallmark of causal-model theory. Note, however, that causal-model theory is compatible with a large coherence effect (i.e., the ¬A¬B combination is judged a better category exemplar than the ¬AB or A¬B combinations) as well as with a small one. A coherence effect could be small and still compatible with the causal-model theory (i.e., the theory only requires that people judge the ¬A¬B combination to be a better exemplar than would be predicted solely based on the A and B main effects). However, note that models that use a multiplicative similarity metric (e.g., the multiplicative exemplar model, Nosofsky, 1984; 1986) can also predict a coherence effect, albeit a small one. In such models, because similarity decreases logarithmically with the number of absent properties, the ¬A¬B exemplar may still be judged to be weakly similar to its category, resulting in a small interaction.

From our discussion above, it should be clear that distinguishing whether a coherence effect found in a causal categorization study is due to similarity processing or due to causal-based processing, may not be a trivial enterprise. In fact, in a recent model (Rehder, 2018), the author presents a model (the *beta-Q* model) that explains categorization judgments by computing a joint distribution of judgments coming from a causal and a similarity based process as a way of accounting for independence violations in human judgments. In Fig. 1, the reader can find illustrations of the kind of interaction that would unambiguously signal causal-based processing as predicted by causal-model theory, and the kind of interaction that might be accounted for by both, similarity and causal-based processing.

In the current work we put forth the idea that part of the difficulty in ascribing interaction results to similarity-based

versus causal-based processing, comes from different ways in which people conceptualize the task they are faced with. Generally, in categorization research the dependent variable is a category membership judgment (e.g., Rehder & Hastie, 2001; Rehder, 2003b; Marsh & Ahn, 2006; Rehder & Kim, 2010). To illustrate the different strategies that people may follow, let's go back to our *tropical frog* running example. In that example, people could reason that a frog that is not poisonous and is not brightly colored, is not a *tropical frog* because it does not show any of the features that are characteristic of *tropical frogs* (i.e., focusing only on A and B but not on their causal relation). Alternatively, people could reason that a *tropical frog* that is not poisonous, should not be expected to have brightly colored skin, and thus, the ¬A¬B exemplar is a perfectly good category member. Note that the second interpretation of the task should produce a large coherence effect because the exemplar with the ¬A¬B feature combination should be judged to be an even better category member than an exemplar with either the A¬B or the ¬AB feature combinations (both of which violate the causal relation, if the causal relation exists).

To show that this second strategy embodies the assumptions behind causal-model theory, we devised a different dependent variable to the one typically used in categorization studies. In a between-subject experiment we implemented two different conditions: *categorization* and *consistency*. In the categorization condition we used a typical categorization rating procedure, which could be approached via any of the two strategies described above. In the consistency condition, we asked participants to rate if the presented category exemplar was to be expected given the category's causal model. We claim that this is the process that causal-model theory assumes that people use. In this second condition, we predicted that focusing on the expected pattern given the received causal information would produce a large interaction effect just as the generative model predicts (Rehder, 2003a; Rehder & Kim, 2010). In contrast, the potential for different strategies in the traditional category membership rating question, would result in a lower coherence effect.

Method

Design

We set up a 2 (Condition: categorization and consistency) x 4 (feature combination: AB, ¬AB, A¬B and ¬A¬B) mixed design experiment. Participants learned about a simple A→B causal model and then used a rating scale (from 1 to 7) to categorize all possible feature combinations.

Participants in the categorization condition had to rate if each exemplar was a member of the studied category. Participants in the consistency condition had to rate if each exemplar was consistent with the causal model of the studied category. Note that though, we expected to obtain a coherence effect in both conditions, we predicted a larger coherence effect in the consistency condition, and a smaller one in the categorization condition, for reasons already discussed.

Participants

Forty-eight undergraduate students (32 female) aged 18 to 47 (mean = 25.43, *SD* = 6.15), agreed to voluntarily participate in the experiment. They were randomly assigned to one of the experimental conditions (consistency and categorization), and control conditions (see below) to the constraint that an equal number of participants were in each cell. Participants received a randomly assigned booklet, and it took them on average 5 minutes to complete the task.

Materials and Procedure

Our materials described a type of rock and a type of language disorder, category names were arbitrary, in the sense of not being related to a specific feature, but rather to the category as a whole (i.e., just like the label "dog" is not related to any particular feature, see Table 1). Additionally, we used an extremely simple A→B model to facilitate subjects' understanding of our materials. Importantly, if participants did not understand causal relations (due to training or other reasons) or if they only relied on stimulus similarity (e.g., to reduce cognitive effort), then they would show the same pattern in both between-subjects conditions (i.e., always respond based on similarity, producing always a small coherence effect).

Table 1: Between-subject condition question manipulation by material.

Condition	"Logodisplenic Disorder"	"Metamorphic Rock"
Consistency	Given what you learned about A causing B in Logodisplenic Disorder, would you say that this case was to be expected?	Given what you learned about A causing B in Metamorphic Rocks, would you say that this rock was to be expected?
Categorization	Given what you learned about A causing B in Logodisplenic Disorder, would you say that this patient belonged to the Logodisplenic Disorder category?	Given what you learned about A causing B in Metamorphic Rocks, would you say that this rock belonged to the Metamorphic Rock category?

To control for other possible factors, during learning, half of our participants received scenarios with causes being described first, and the other half received scenarios with effects being described first (i.e., our training scenario control). Finally, participants always rated AB exemplar first to promote correct rating scale use. Other exemplars were presented in one of three possible latin-square orders as a means of controlling for order effects (i.e., our exemplar order control).

We produced forty-eight booklets with a total of seven pages each. The first page contained the informed consent that every participant read and signed. The second page contained a cover story that described a category along with its causal model (i.e., *Logodisplenic Disorder* was a type of language disorder, *Metamorphic Rock* was a type of rock). Both stories described a simple causal model with one cause and one effect, where there was a conditional probability of 0.70 of the effect being present given the cause (see Table 2). The third page contained instructions regarding the rating scale. In the fourth to seventh pages, booklets presented exemplars with different feature combinations, making a total of four cases. Each exemplar was presented by describing a specialist (respectively, a neurologist, a geologist) who found and described the exemplar, and participants were asked to report their judgments by using a seven-point rating scale. Scenarios and exemplars were presented in writing and the corresponding causal model (i.e., A→B) remained always in view.

Table 2: Description of features and category names.

Feature	"Logodisplenic Disorder"	"Metamorphic Rock"
A	Bearer of the FOX1 gene	High concentration of calcium salts
B	Difficulties in developing normal language	Being soft

Results

Ratings were submitted to a 2 (question: consistency, categorization) x 4 (feature combination: AB, ¬AB, A¬B, ¬A¬B) mixed ANOVA, with the last being the repeated measures factor. The analysis produced a main effect of question type ($F(1, 46) = 22.46$, $MSe = .40$, $p < .001$, $\eta_p^2 = .33$, power = .97), a main effect of feature combination ($F(3, 138) = 46.48$, $MSe = 2.68$, $p < .001$, $\eta_p^2 = .50$, power > .99), and a significant interaction ($F(3, 138) = 12.51$, $MSe = 2.68$, $p < .001$, $\eta_p^2 = .21$, power > .99). To follow up on the significant interaction, we performed simple effects analyses between consistency and categorization, at each level of the feature combination factor. Results showed no significant difference for the AB feature combination ($F(1, 46) = 1.47$, $MSe = .91$, $p = .23$, power = .34), a small significant difference for the A¬B combination ($F(1, 46) = 4.38$, $MSe = 1.72$, $p = .04$, power = .66), a non-significant difference for the ¬AB combination ($F < 1$, power = .12), and a highly significant difference for the ¬A¬B combination ($F(1, 46) = 33.29$, $MSe = 3.81$, $p < .001$, power > .99). Lastly, we do not find evidence of significant differences depending on the type of material (i.e., "Metamorphic Rock" and "Logodisplenic Disorder") nor of exemplar order ($p > 0.5$). As Fig. 1 illustrates, most of the difference between conditions is accounted for by the way people responded to the ¬A¬B feature combination. Note that the categorization condition produced a small coherence effect that is compatible with causal-model and with similarity theories, while the consistency condition produced a coherence effect that is unmistakably of causal reasoning origin.

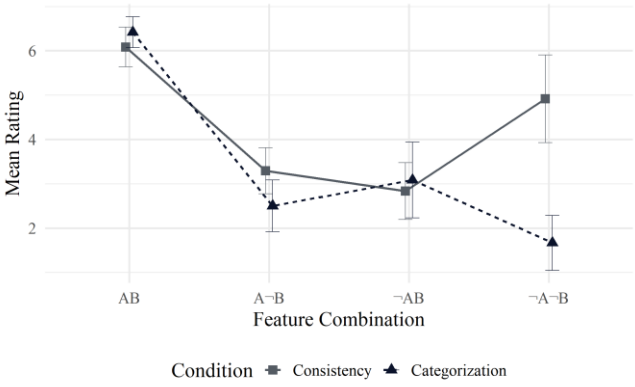

Figure 1: Mean ratings for each feature combination in the Consistency and Categorization conditions. Dashed line = categorization condition, Filled line = consistency condition. The ¬ negation symbol indicates absent features. Error bars are 95% confidence intervals for the mean.

Discussion

In our experiment, we showed that the size of the coherence effect changes depending on the type of question subjects are considering. Participants learned a simple causal model and were then presented with different combinations of absent and present properties, under one of two different between-subjects conditions (*categorization* and *consistency*). In the categorization condition, we asked participants to rate exemplars' category membership. In the consistency condition, we asked participants to rate the presented exemplar as likely or not given their knowledge of the simple causal model. Results showed that the coherence effect tends to be substantially larger in the consistency condition than in the categorization condition.

Results are consistent with our hypothesis that the categorization question can be approached with different strategies, at least under conditions like those of our experiment. Going back to our *tropical frogs* running example, we suspected that if participants were faced with a category membership question, then they might understand that if a *tropical frog* is not poisonous (¬A) and is not brightly

colored (¬B), then this feature combination would entail a weak category membership because the exemplar does not have any of the singular features that makes an exemplar a *tropical frog*. It is important there to clarify that our results do not show an absence of the coherence effect. Rather, we show that the size of the coherence effect can change depending on the dependent variable. In this sense, our results are not in conflict with prior literature that reports a coherence effect (Rehder, 2017; Rehder & Kim, 2006; 2010; Hampton, Storms, Simmons & Heussen, 2009). Rather, our results provide greater precision on how the coherence effect should be interpreted.

Relative to this last issue, note that the relatively small size of the coherence effect in our experiment's categorization condition could be accounted for by a similarity-based model, and not only by a causal-model account of categorization, as explained earlier. Unfortunately, our current experiment could not provide us with a definitive answer regarding why the small interaction effect (i.e., if participants were responding based on similarity or causal reasoning). However, our results are suggestive that a large portion of our participants might not have resorted to causal reasoning in the categorization condition (as evidenced by the small confidence interval for the ¬A¬B feature combination in the categorization condition; see Fig 1). In contrast to results in the categorization condition, recall that we predicted that if participants were faced with the consistency question, then they would understand that the combination of not being poisonous and not being brightly colored is consistent with the *tropical frog* causal model. As predicted, in the consistency condition we found a relatively large coherence effect (i.e., high ratings for the ¬A¬B combination) that can only be explained by a causal reasoning process.

To summarize, our results show that the categorization question does not unambiguously lead subjects to use causal reasoning. Judging by the small confidence intervals in Fig. 1, most of them could have resorted to similarity processing. Though we cannot positively assert this, our results should be considered at least as very suggestive. Finally, our results are likely not to be a consequence of sample characteristics (e.g., cultural, social, educational differences with samples used in prior literature), because in the consistency condition, our participants reasoned in close agreement with the causal model theory.

Though we believe our results to be interesting because they may help in clarifying the conditions that modulate the coherence effect, there are several limitations that we want to briefly discuss, and that are currently guiding work in our lab. Problematically, some of these limitations might be working against obtaining a larger coherence effect in the categorization condition. First, in our materials there was no clear indication that features A and B were not by themselves characteristic of the category. This might have led participants to focus on the features themselves, and not on the causal relation, at least in the categorization condition. To solve this, in future experiments we plan to inform participants of feature diagnosticity (i.e., p(category|feature)). Presumably, features with low diagnosticity but with a medium to strong causal relation could produce a larger coherence effect. Also, our materials did not include a mechanistic explanation for the causal link. Lacking this explanation may have led participants to disregard causal information, thus giving greater weight to the individual features (though, note that this did not happen in the consistency condition). This is also something that we are currently working on. Finally, note that our categorization condition is different from the consistency condition in that the former requires categorizing the exemplar, while the latter implies that the exemplar under categorization is already known to be a category member (e.g., it's a *tropical frog* that shows a specific feature combination). This in itself may have led participants frame the question differently, with a focus on the individual features in the categorization condition and a focus on the causal relation in the consistency condition. We are planning future experiments to test if this is also a factor that modulates the coherence effect. On concluding, we are hopeful that the factors we have identified in the current experiment will allow us in the near future to more clearly specify the different variables that modulate the size of the coherence effect in categorization.

Acknowledgements

We appreciate the help and support on data acquisition of Joaquín Migeot, Felipe Toro and Daniela Olivares. This work was financially supported by Fondecyt grant 1190006, from the Chilean government.

References

Hampton, J. A., Storms, G., Simmons, C. L., & Heussen, D. (2009). Feature integration in natural language concepts. *Memory and Cognition*, *37*(8), 1150–1163. https://doi.org/10.3758/MC.37.8.1150.

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, *62*, 135-163. doi:10.1146/annurev.psych.121208.131634.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104–114. https://doi.org/10.1037/0278-7393.10.1.104.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39-57. doi:10.1037/0096-3445.115.1.39.

Marsh, J. K., & Ahn, W. (2006). The role of causal status versus inter-feature links in feature weighting. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science, 27*(5), 709-748. doi:10.1016/S0364-0213(03)00068-5.

Rehder, B. (2003b). A Causal-Model Theory of Conceptual Representation and Categorization. Journal of Experimental Psychology: Learning Memory and

Cognition, 29(6), 1141–1159. https://doi.org/10.1037/0278-7393.29.6.1141.

Rehder, B. (2017). Concepts as Causal Models: Categorization. In M. Waldmann (Ed.), *The Oxford handbook of causal reasoning*. New York, NY: Oxford University Press.

Rehder, B. (2018). Beyond Markov: Accounting for independence violations in causal reasoning. *Cognitive Psychology*, *103*(January), 42–84. https://doi.org/10.1016/j.cogpsych.2018.01.003.

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), 323-360. doi:10.1037/0096-3445.130.3.323.

Rehder, B., & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning Memory and Cognition, 32*(4), 659-683. doi:10.1037/0278-7393.32.4.659.

Rehder, B., & Kim, S. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning Memory and Cognition, 36*(5), 1171-1206. doi:10.1037/a0019765.

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science, 15*(6), 307-311. doi:10.1111/j.1467-8721.2006.00458.x.