Categorizing perceived causal events

Nicolás Marchant[1] (nicolasmarchant@alumnos.uai.cl), **Bonan Zhao**[2] **(**b.zhao@ed.ac.uk**)**, **Neil R. Bramley**[2] (neil.bramley@ed.ac.uk), **Diego Morales**[3] (dimoralesb@gmail.com) & **Sergio E. Chaigneau**[1,4] (sergio.chaigneau@uai.cl)

[1] Center of Cognitive and Social Neuroscience, Universidad Adolfo Ibáñez, Santiago, Chile
[2] Department of Psychology, The university of Edinburgh, Edinburgh, UK
[3] Center of Cognitive Science Research, Universidad de Talca, Talca, Chile
[4] Center of Cognitive Research (CINCO), Universidad Adolfo Ibáñez, Santiago, Chile

Abstract

Over the last few decades, Causal Model Theory (CMT) has become a dominant framework for human causal-based reasoning, including categorization and inference. CMT prescribes how people should reason about probabilistic events in terms of causal models. In typical causal-based categorization experiments, subjects are provided with verbal descriptions of causally linked features, generally including probabilistic information. Another line of research focuses on perceived or experienced causal events, rather than on verbal descriptions. In this work we asked whether effects which are consistent with CMT, and that have been obtained with verbal descriptions, generalize to visually perceived events. In two experiments, we presented subjects with videos of a 3D A→B causal event rather than verbal descriptions. In Exp. 1, we found that subjects who saw the causal event did not show the coherence effect in categorization (i.e., subjects tend to rate the null ¬A¬B event as a category member). However, subjects who did see the null event during training did show the effect. In Exp. 2, we ruled out the possibility that Exp. 1's results were simply an effect of how frequently events were experienced during training. We conclude that a one-shot perceived causal event is not sufficient for people to show causal-based reasoning as CMT predicts.

Keywords: Categorization; Bayesian reasoning; Launching effect; Conceptual coherence.

Introduction

The idea that conceptual representations encode causal relationships between features has gained in popularity in parallel with the emergence of CMT in cognitive psychology (Malt & Smith, 1984; Waldmann, Holyoak, & Fratianne, 1995; Wisnieski, 1995; Hampton, Storms, Simmons, & Heussen, 2009; Rehder, 2017; Zhao, Lucas, & Bramley, 2021). In causal categorization and inference experiments, people are typically presented with descriptions of novel concepts that include causally structured feature information. They are then asked to judge category membership of new cases based on their feature values, or to make inferences about their unobserved features from their observed ones. To illustrate with a simple example, consider the following description (adapted from Rehder, 2003a): "*Kehoe ants* have thick blood, which frequently causes them to become immobile during cold winters." Subjects might then be presented with a description of an ant with thin blood that becomes immobile during cold winters, and asked to rate its membership in the *Kehoe ant* category (i.e., a categorization task). Alternatively, subjects may be asked to estimate the probability that a *Kehoe ant* with thin blood will become immobile during cold winters (i.e., an inference task). Using verbally described concepts, research has shown that people assess category membership in ways that are broadly consistent with CMT, exhibiting characteristic patterns such as coherence effects, explaining away, and a causal status bias over features (Kahneman & Tversky, 1982; Lombrozo, 2010; Walsh & Sloman, 2011; Marchant & Chaigneau, 2020; Rehder, 2003a; 2003b).

But causal inference extends far beyond verbal description. People often perceive causal relationships directly when observing physical interactions such as collisions (Michotte, 1946/1963; Scholl & Tremoulet, 2000; Blakemore et al, 2001; Wolff, 2008; Rips, 2011), even when the perceived causality is in fact illusory or coincidental (Bechlivanidis, Buehner, Tecwyn, Lagnado, Hoerl, & McCormack, 2021). In a typical perceptual causality experiment, subjects observe a clip in which an object A starts moving, connects with an object B, whereupon object B starts moving – a so called "launching event" (Gordon, Day, & Stecher, 1990), and subjects typically report seeing A *cause* B to move. This perception of physical causality seems to be strong enough to trump other types of information (e.g., Buehner, & Humphreys, 2010; Bechlivanidis, Schlottmann, & Lagnado, 2019). In another causal learning paradigm, Blicket detector experiments also demonstrate that people readily draw on objects' interactions and perceptual features to infer categories based on causality, and make inferences driven by causal categories (Gopnik & Sobel, 2000; Kemp, Goodman, & Tenenbaum 2010; Sim & Xu 2017). In Blicket experiments, subjects learn that certain objects can make a machine activate (i.e., light up and play a sound). Children as young as two use "blicketness" to categorize novel objects (Gopnik & Sobel, 2000), and adults have been shown to draw on perceptual features of objects as well as interaction evidence in complex ways to impute causal categories and functional forms that in turn guide causal predictions (Kemp, Goodman, & Tenenbaum, 2010).

We are interested in the difference between verbal descriptions and perceptual experiences of causally structured concepts: In most categorization experimental settings, the verbal descriptions often contain probabilistic information, indicating that the events being described have

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.* ©2022 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

a certain probability of occurring in the presence and/or absence of their causes; perceptual experiences, on the other hand, are inherently single-shot, providing evidence about the potential alternative realizations of the system only indirectly. Therefore, we explore whether perceiving a sequence of physically realistic causal events yields some of the same types of categorical reasoning that verbal descriptions of causal structure afford.

The coherence effect

Under CMT, a new observation's probability of category membership is driven by its likelihood under the category's generative causal model (e.g., Rehder, 2003a; 2003b). *Ceteris paribus*, this is most likely when the feature values are *coherent*, in the sense of being a plausible manifestation of the generative model. In the coherence effect, people show sensitivity to this principle. For example, if the conceptual causal relationship A→B is understood to be strong and generative, people will take observations in which both A and B occur, and where neither A and B occur as more compatible with the concept – hence more likely to be produced by a category member – than observations in which A but not B occurs, or where B but not A occurs (Hampton, Storms, Simmons & Heussen, 2009; Malt & Smith 1984; Marchant & Chaigneau, 2020, 2021; Murphy & Wisniewski, 1989; Rehder, 2017; Rehder & Kim, 2006, 2010; Wisniewski, 1995). If people further believe that the effect has a low base rate, they may find the case where only the effect is present to be particularly incompatible with the category. To illustrate with the *Kehoe ant* concept: the coherence effect occurs when people judge that an ant with thin blood that does not become immobile in cold winters (¬A¬B) is more likely to be a Kehoe ant than one that has thick blood and does not become immobile during cold winters (¬AB), or than an ant that has thin blood and becomes immobile during cold winters (A¬B). This reasoning pattern has been replicated reliably in the causal categorization literature and is taken to show people consider consistency between the causal structure of the concept and the evidence to be as or more important than the presence of characteristic features.

Hypothesis

As discussed above, the evidence of coherence effects in categorization comes from experiments using verbal descriptions of causal events. This often includes some probabilistic information, either in the form of specific parameters for the generative causal model (e.g., for P(A), P(B), P(B|A)), or using adverbs that confer degrees of reliability for the connections (e.g., "sometimes", "frequently", "often", etc.). Recall that one manifestation of such a coherence effect is considering that the scenario where the causal event does not occur (¬A¬B) to be relatively consistent with an A→B causal model. If being led to think about this alternative event as part of a causal model is necessary for the coherence effect to obtain (Mayrhofer, & Rothe, 2012), then, experiencing a known category member producing the causal event will not automatically afford that type of reasoning because it is a one-shot event with no probabilistic information. Note that though there is evidence suggesting that people can take alternative or counterfactual events into account when analyzing perceived events (Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2021), in that study subjects were explicitly asked about making inferences in alternative perceptual scenarios (e.g., would the same result occur if the cause were not present?).

In the current work, we test whether and under what conditions observations of a causal perceptual event lead to coherence effects, taking this as evidence that people form mental categories according to the CMT. This issue is important for establishing the generality of CMT beyond verbal descriptions. Categorization is a central cognitive ability, critical for generalization (Zhao, Lucas, & Bramley, 2021) and symbolic cognition in general (Piantadosi, Tenenbaum & Goodman, 2016). As such, any account of it must be able to interface with different modes of learning and inference (Ashby & Maddox, 2005; 2011).

Experiment 1

We conducted an experiment in which subjects learned about a simple physical causal mechanism involving two salient launching acts A and B. In one condition, subjects first watched a video clip of the mechanism in action in which both acts occurred (AB *causal event* and ¬A¬B *null event*, Phase 1). In a second condition subjects first watched a video clip in which only the causal event occurred (AB, *causal only*). Later, subjects in both conditions rated whether four events (AB, A¬B, ¬AB, ¬A¬B) depicted the same mechanism. To minimize the influence of specific background knowledge, we used a novel label for the artificial category. Concretely, it was referred to as a *"Self-retracting Mechanism"*. If subjects showed a classic coherence effect, then they should rate the ¬A¬B null event at least as likely to depict a self-retracting mechanism than the events depicting A¬B and ¬AB. Alternatively if subjects' judgements are purely based on featural similarity to the category, we would expect them to rate ¬A¬B null event as less probable category member than A¬B and ¬AB events.

Method

Participants: Forty-eight subjects (31 female) aged 18 to 44 (mean = 25.44, *SD* = 6.0) were recruited online through Prolific Academic (https://www.prolific.co/) and received monetary compensation according to Prolific rules (at a rate of £7.56 per hour). The task took around 5 minutes. Two subjects were excluded from analysis because they answered the attention check question incorrectly, leading to a final sample of forty-six subjects.

Design: We implemented a 2 conditions (causal-only, causal + null event) x 4 event types (AB, A¬B, ¬AB, ¬A¬B) mixed design, with repeated measures in the event types factor.

Materials and Procedure: We created clips of a 3D physical scene using Blender (Community, B. O., 2018). Each scene

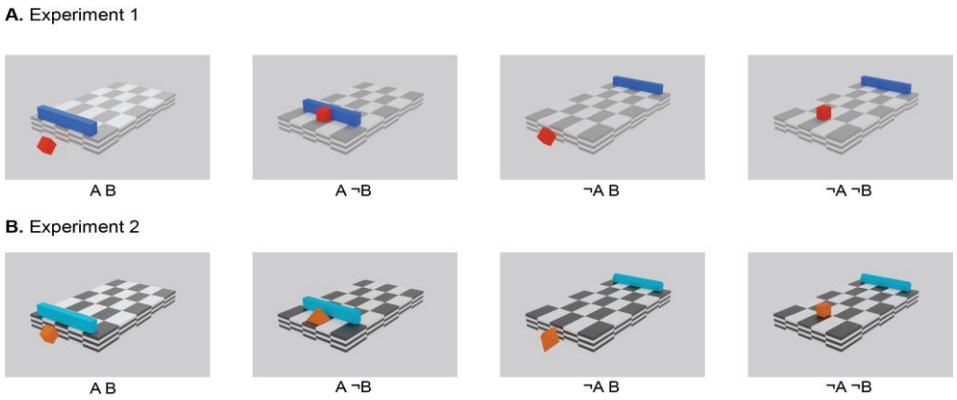

Figure 1: Freeze frames of events used in Experiments 1 and 2, captured towards the end of each event.

depicts simple objects involving in a dynamic interaction, totaling four 8-second videos (see Fig. 1, and available at: https://osf.io/a5pwz/). In each scene, a blue elongated cuboid starts at rest, on a checkered surface. A red cube falls to the surface. In the AB event, the blue object starts moving and collides with the red square box (act A). At the moment of collision, the blue rectangle slows down, and the red cube begins to move along the same trajectory. The red cube then falls from the surface (act B). The other three videos were created to show the same objects, but with one or both acts absent. For the A¬B event, the blue object starts moving and collides with the red object. However, after collision, the red object stops short of the surface's edge and does not fall from the checkered surface. For the ¬AB causal event, the blue object does not move, while the red object starts moving independently and falls from the surface. For the ¬A¬B event, both objects remain at rest.

The experiment had two phases: In Phase 1, subjects watched the example mechanism in action. Subjects in *causal-only* condition watched only the AB event (Figure 1A). Subjects in *causal + null event* condition watched both AB and ¬A¬B events (Figure 1A). After this, they proceeded to Phase 2, where they were asked to classify four events AB, A¬B, ¬AB and ¬A¬B. In each task, subjects were asked "Is this video a Self-retracting Mechanism?" and responded using a scale ranging from 0 (Definitely is not a "Self-retracting Mechanism") to 100 (Definitely is a "Self-retracting Mechanism"). The slider allowed increments in steps of size 5 and the thumb was initialized at the middle of the scale at the start of each trial.

Subjects were instructed that they would see a video illustrating a *"Self-retracting Mechanism"* and then make judgments about whether several other events depict the same mechanism. Additionally, the instructions indicated that the experiment consisted of two phases. Subjects were randomly assigned to one of the two experimental conditions. During Phase 1, subjects had to watch the video(s) once and then pressed the "Next" button to continue to Phase 2. Immediately after Phase 1, subjects answered an attentional check question in which they had to choose the correct category name out of three possible alternatives. Subjects that incorrectly responded to this question continued to Phase 2 but were removed from the analyses.

During Phase 2, subjects viewed all four events twice. And each and every time had to rate each one's probability of category membership. In Phase 2, subjects always observed the AB causal event first, to promote the correct rating scale use. The other seven videos were shown in random order. Subjects were not allowed to go back to check previous responses, nor could they modify their responses once an answer was submitted.

Results

Because subjects watched each event video twice, we computed mean rating for each event (i.e., the average of each type of video's first and second presentation). Fig. 2A shows the mean and standard errors of ratings for each event in both conditions. Mean ratings were submitted to a 2 (condition: *causal-only*; *causal + null event*) x 4 (event: AB, A¬B, ¬AB, ¬A¬B) mixed ANOVA, with the last being the repeated measure factor. The analysis revealed a significant main effect of event $F(3,132) = 34.28$, $MSe = 863.30$, $p < .001$, $\eta_p^2 = .44$, power > .99, a non-significant main effect of condition ($F(1,44) = 0.04$, $p = .85$), and a significant interaction ($F(3,132) = 6.17$, $MSe = 863.30$, $p = .001$, $\eta_p^2 = .12$, power =

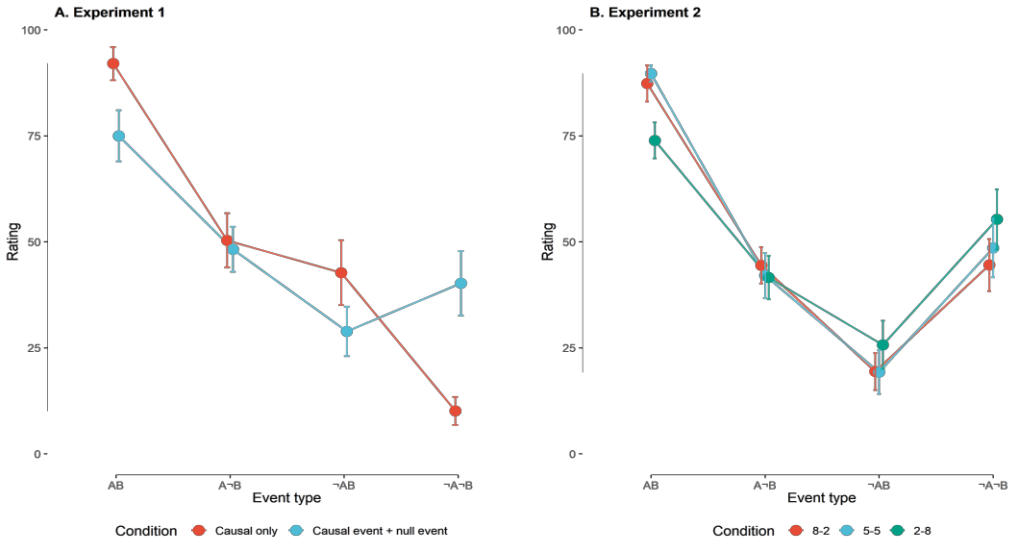

Figure 2: Mean ratings in both conditions for each event type on Exps. 1 and 2. *Note:* Error bars are ∓ 1 *SE* of the mean.

.96). Following interaction and as it is illustrated in Fig. 2A, subjects in *the causal-only* condition gave higher ratings to the AB event than subjects in *causal + null event* condition $F(1,44) = 5.41$, $MSe = 616.66$, $p = .025$, $\eta_p^2 = .11$, power = .62). However, the opposite effect was found for the ¬A¬B event, in which subjects in the *causal-only* condition provided ratings that were lower ($F(1,44) = 12.42$, $MSe = 837.33$, $p = .001$, $\eta_p^2 = .22$, power = .93). For events A¬B and ¬AB we did not find any significant difference between conditions.

To test for the presence of the coherence effect, we followed the significant two-way interaction with planned comparisons at each level of the condition factor. We filtered our data by condition and performed planned contrasts in the repeated measures factor. We found that subjects in the *causal-only* condition rated the ¬A¬B clip as the least probable category member, lower than both the A¬B clip ($F(1,21) = 28.56$, $MSe = 1246.37$, $p < .001$, $\eta_p^2 = .58$, power > .99) and the ¬AB clip ($F(1,21) = 17.84$, $MSe = 1311.59$, $p < .001$, $\eta_p^2 = .46$, power = .98). This pattern is not in line with the expected coherence effect. For the *causal + null event* condition, we found no evidence for a difference between ¬A¬B and the A¬B clip ($F(1,23) = .58$, $MSe = 2680.42$, $p = .46$, $\eta_p^2 = .02$, power = .11), nor with the ¬AB clip ($F(1,23) = 1.11$, $MSe = 2778.79$, $p = .30$, $\eta_p^2 = .05$, power = .17). Thus, in the *causal + null event* condition, subjects judged the null event (¬A¬B) as a similarly plausible category member as A¬B and ¬AB events, which is potentially consistent with the coherence effect.

Interim discussion

The coherence effect predicts that people see the null event depicting an absent cause and an absent effect as a plausible observation of a mechanism with an A→B relation. This did not seem to be the case when subjects had learned about the mechanism only from a positive exemplar: AB *causal-only*. However, when subjects learned about the causal relation by witnessing both the causal and the null event, their judgments were more in line with the classic coherence effect pattern.

However, these results are not completely unambiguous. It is possible that subjects' relatively higher ratings for ¬A¬B in the *causal + null event* condition than the *causal-only* condition was due to the fact that this event was shown as an exemplar of the mechanism during training. That is, subjects might have responded based on recognition of a behavior known to be producible by the mechanism. On this view it is more curious that ratings of ¬A¬B were lower than AB since both were presented once each during training. Exp. 2 was designed to test this deflationary similarity-based judgment hypothesis. Another issue of concern is that our evidence for the coherence effect in the causal + null event condition comes from not obtaining significant differences when comparing the ¬A¬B event against the A¬B and ¬AB events.

Experiment 2

To reduce concerns regarding whether responses in Exp. 1 reflect similarity rather than a causal-model based coherence effect, we ran a second experiment in which subjects again viewed clips of the mechanism but where we systematically manipulated the frequencies of which the AB event and ¬A¬B events were experienced during training so as to provide probabilistic evidence. Using an observational learning paradigm (similar to Lagnado & Sloman, 2004; Park & Sloman, 2013 Exp. 3), we tested whether the similarity-based explanation might be correct. If this were the case, we would expect that ratings for the AB and ¬A¬B events should change as a function of the frequency in which both events were experienced during training.

Method

Participants: Ninety-six subjects (60 female) aged 18 to 44 (mean = 27.02, SD = 8.27) were recruited online through Prolific Academic and received monetary compensation (at a rate of £7.50 per hour). The task took 6 minutes. No subjects were removed before analyses.

Design: We implemented a 2 (order) x 3 (condition: 8-2; 5-5; 2-8) x 4 (event types: AB, A¬B, ¬AB, ¬A¬B) mixed design experiment with repeated measures in the last factor. Conditions are explained next.

Material and procedures: We used similar materials and followed the same procedures as in Exp. 1. In Phase 1, subjects saw causal event scenarios of a "Self-retracting Mechanism" as shown in Fig. 1, and in Phase 2 they were asked to decide whether a possible event is a member of the "Self-retracting Mechanism" category using a rating scale. However, in Exp. 2, we implemented an observational learning paradigm during Phase 1. We used three conditions in which, (1) the AB event was experienced on 80% of the trials and the ¬A¬B event on 20% of the trials (the 8-2 condition); (2) the AB event was experienced on 50% of the trials and the ¬A¬B event on 50% of the trials (the 5-5 condition); (3) the AB event was experienced on 20% of the trials and the ¬A¬B clip on 80% of the trials (the 2-8 condition). The three observational learning conditions consisted of passively viewing ten events. Because there were ten videos but only four types of events, we created two different shapes for each individual object. For act A (blue object moving), we introduced two shapes: an elongated cuboid and a pentagonal prism. For act B (red object falling off the surface), we also used two different shapes: a cube and a pyramid (Fig. 1B). Subjects received different shape/ event combinations; hence they would not watch the same video twice. To control for order effects during training, we created two different orders randomly (between subjects). After completing the observational learning phase, subjects received the attentional check question and then continued to Phase 2, where they had to rate all 16 possible event combinations (4 events x 4 shapes = 16 different events) using the same category membership rating scale used in Exp. 1. All sixteen different events were randomized with the exception of a single AB event (i.e., the rectangle that moves and collides with the cube causing it to fall off), which was always presented first to promote correct use of the rating scale.

If the similarity-based explanation were correct, we should find that subjects are sensitive to frequencies during the observational learning phase. For example, subjects in condition 8-2 should rate the AB event as a better category member, followed by subjects in condition 5-5, and subjects in condition 2-8 should give the lower rating. The opposite pattern of responding should occur for the ¬A¬B null event.

Results

Similarly to Exp.1, we averaged the ratings for each event type (e.g., causal event AB consisted of 4 videos: rectangle-cube; rectangle-pyramid; prism-cube; prism-pyramid). Fig. 2B shows averages of each event type. Data were submitted to a mixed 2 (order) x 3 (condition) x 4 (event type) ANOVA. The ANOVA test showed that order produced no main effect $F(1,90) = 0.65$, $p = .94$ and participated in none of the two-way interactions (order x event type, $F(3,270) = 0.14$, $p = .87$; order x condition, $F(1,90) = 0.73$, $p = .49$, nor in the three-way interaction $F(3,270) = 0.31$, $p = .87$. For this reason, we continued the analysis with the order factor collapsed. We found a significant main effect of event type $F(3,279) = 64.36$, $MSe = 1495.32$, $p < .001$, $\eta_p^2 = 0.41$, power < .99, a non-significant interaction between condition and event type $F(3,279) = 1.25$, $p = .28$, and a non-significant main effect of condition $F(1,93) = 0.06$, $p = .94$. Because of a violation of sphericity ($X^2(5) = 83.74$, $p < .001$), results are reported using the Greenhouse-Geisser correction. The non-significant interaction suggest that subjects were insensitive to event frequencies during observational learning. In the general discussion we return to this idea and provide some possible explanations on why this may occur.

Because we did not find a significant two-way interaction, we collapsed our data by condition and performed planned contrast on event type. After collapsing our data, we found that subjects rated the AB event higher than the A¬B event ($F(1,95) = 191.30$, $MSe = 842.09$, $p < .001$, $\eta_p^2 = .67$, power > .99), the ¬AB event ($F(1,95) = 228.38$, $MSe = 1625.63$, $p < .001$, $\eta_p^2 = .71$, power > .99) and the ¬A¬B event ($F(1,95) = 65.17$, $MSe = 1722.14$, $p < .001$, $\eta_p^2 = .41$, power > .99). This shows that in each condition the AB causal event was rated as the most likely to depict a category member. Furthermore, the A¬B event was rated higher than the ¬AB event ($F(1,95) = 31.13$, $MSe = 1389.14$, $p < .001$, $\eta_p^2 = .25$, power > .99), suggesting that the event where act A is present (i.e., the cause, the blue object moves and hits the red object) contributes more to category membership than the event where act B (i.e., the effect, the red object falling off the surface) is present. Importantly, we found that the null event ¬A¬B was not statistically different from the A¬B event ($F(1,95) = 1.43$, $MSe = 3079.63$, $p = .24$, $\eta_p^2 = .02$, power = .22), but was taken as more likely to indicate category membership than the ¬AB event ($F(1,95) = 22.60$, $MSe = 3329.65$, $p < .001$, $\eta_p^2 = .19$, power > .99). This last result is consistent with the coherence effect prediction and replicates our findings from Exp. 1 but does not rely on a non-significant result.

Interim discussion

Results did not support the similarity-based hypothesis which was offered as an alternative account for Exp. 1. If results in Exp. 1's *causal + null event* condition are dependent on subjects judging that the ¬A¬B was a category member simply because they had seen a known category member exhibit it, then we would expect providing different amounts of experience of the mechanism producing the

¬A¬B event (frequencies = 2, 5 or 8) should have affected subjects' ratings for this type of event at test. In fact, we found that subjects produced a similar level of coherence effect across different training frequency conditions. Interestingly, we found that subjects treated act A (the blue object colliding with the red object) as having a greater influence than act B (the red object falling off the table) on category membership ratings (the A¬B event received, on average, higher ratings than the ¬AB event), in line with the causal status effect (Ahn, 1998; Ahn, Kim, Lassaline, & Dennis, 2000; Mayrhofer & Rothe, 2012) such that causes are weighed more than their effects when judging category membership.

General Discussion

We investigated causal-based categorization using videos of 3D objects, and found that, while observing an exemplar of a single launching mechanism failed to replicate the coherence effect as found in many experiments using verbal descriptions, introducing an observation in which neither act occurs (the null event) did lead to the expected pattern. We tested category membership judgments following a single-shot observation (Exp. 1) and a set of observations indicating frequencies (Exp. 2) and found judgment patterns cannot be easily explained by similarity-based categorization. In other words, people indeed seem to integrate feature information causally to make categorization decisions in our experimental 3D virtual world, but only when provided with the help of null events where both the cause and effect are absent. While previous research has established a coherence effect using verbal descriptions (Hampton, Storms, Simmons & Heussen, 2009; Marchant & Chaigneau, 2020; Rehder, 2017; Rehder & Kim, 2006, 2010; Wisniewski, 1995), ours is the first experiment to explore how these previous findings in categorization generalized to visually perceived scenarios.

One difference between verbal descriptions and visual evidence lies in their ability to communicate probabilistic information, which is critical for predicting the coherence effect (Rehder & Kim, 2006; 2010). In verbal descriptions, probabilistic information can be communicated directly using frequency words, while direct observation provides extra mechanistic richness but only indirect evidence about long-run probabilities. We hypothesized that a single causal launch event would not produce a coherence effect in causal categorization because subjects cannot directly judge the base rates and conditional probabilities involved, and that, in contrast, if we provided subjects experience with the null ¬A¬B event, they would then show the coherence effect. Results from Exp. 1 supported our hypothesis. In Exp. 2 we tested an alternative explanation such that subjects showed a coherence effect because of similarity-based categorization (i.e., direct match of exemplars to generalization cases). To test this alternative hypothesis, we trained subjects on an observational learning paradigm with the AB and ¬A¬B events with different frequencies, and found that subjects were practically insensitive to frequency information, showing a similar level of coherence effect regardless how often they had seen these at training. It seems that having seen the null ¬A¬B event, regardless of its frequency, sufficed to change subjects' response patterns. These results provide evidence that the mere perception of a one-shot causal event does not allow an effect that has been typically shown by using verbal descriptions. Therefore, our results suggest that there may not be a seamless continuity from one type of causal understanding to the other, and that a perceived cause may not automatically allow full-fledged causal-Bayesian reasoning about events. These results may be relevant for theories of development of causal cognition (e.g., Kuhn, 2012), and for theories of comparative cognition (e.g., Blaisdell, Sawa, Leising, & Waldmann, 2006), where the issue of continuity of causal processing across development and across species is important.

Though our results are suggestive, there are at least three limitations that we wish to discuss before closing. The first one, is that it is possible that ¬AB case was rated lower than A¬B because it is a physically surprising event. It has been shown that an object that starts moving without being influenced by another violates our intuitive understanding of physics. Future experiments should address this concern by testing events that are not physically implausible. The second limitation comes from the question regarding why subjects in Exp. 2 were not influenced by frequency information. Other studies have found that subjects are able to learn causal structures and probabilities through observational learning (e.g., Meder, Hagmayer, & Waldmann, 2008; Park & Sloman, 2013). However, in those studies subjects learned by observing frequencies of highly abstracted information (e.g., geometrical symbols co-occurring, sliders indicating the state of variables). It might be that this parsing of events and event structure made the task easier, or perhaps turned it into a reasoning rather than a perceptual task. A third limitation is that naming the category "Self-retracting mechanism" may have given verbal hints about the causal structure of the concept that could have influenced judgements. Future experiments should address this concern by using neutral names for novel causal categories.

In future research, we could consider integrating our paradigm with feedback learning rather than observational learning, as feedback encourages subjects to learn probabilistic information from direct experience (Knowlton, Squire, & Gluck, 1994; Packard & Knowlton, 2002). Another direction is to develop computational models that incorporate counterfactual and contextual information for inference (Gertenberg, Goodman, Lagnado, & Tenenbaum, 2021), and extend its usage to categorization. Recent work also proposed computational modeling framework for object-based causal generalization by constructing causal categories built on perceptual features (Zhao, Lucas, & Bramley, 2021). Such an approach can be adapted to categorization tasks like those reported here, with the right probabilistic information at hand. In sum, our experiments provide a rich testbed for various computational methods to allow us exploring further into causal categorization in the direct visual perceptual domain.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts?: The role of causal status in determining feature centrality. *Cognition, 69*(2), 135-178. doi:10.1016/S0010-0277(98)00063-8.

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*(4), 361-416. doi:10.1006/cogp.2000.0741.

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology*, 56(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, 1224(1), 147–161. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Bechlivanidis, C., Buehner, M. J., Tecwyn, E. C., Lagnado, D. A., Hoerl, C., & McCormack, T. (2021). Human vision reconstructs time to satisfy causal constraints. *Psychological Science*, 09567976211032663.

Bechlivanidis, C., Schlottmann, A., & Lagnado, D. A. (2019). Causation without realism. *Journal of Experimental Psychology: General*, *148*(5), 785.

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science, 311*(5763), 1020-1022. doi:10.1126/science.1121872.

Blakemore, S. J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A. N., Segebarth, C., & Decety, J. (2001). How the brain perceives causality: An event-related fMRI study. *NeuroReport*, 12(17), 3741–3746. https://doi.org/10.1097/00001756-200112040-00027.

Buehner, M. J., & Humphreys, G. R. (2010). Causal contraction: Spatial binding in the perception of collision events. *Psychological Science*, 21(1), 44–48. https://doi.org/10.1177/0956797609354735

Community, B. O. (2018). *Blender - a 3D modelling and rendering package*. Stichting Blender Foundation, Amsterdam. Retrieved from http://www.blender.org

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review, 128*(5), 936–975. https://doi.org/10.1037/rev0000281

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205-1222. https://doi.org/10.1111/1467-8624.00224

Gordon, I. E., Day, R. H., & Stecher, E. J. (1990). Perceived causality occurs with stroboscopic movement of one or both stimulus elements. *Perception*, 19(1), 17–20. https://doi.org/10.1068/p190017

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.37.8.1150Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). Judgment under Uncertainty: Heuristics and Biases. Cambridge, UK: Cambridge University Press. https://doi:10.1017/CBO9780511809477

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. Cognitive Science, 34(7), 1185-1243. https://doi.org/10.1111/j.1551-6709.2010.01128.x

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1*(2), 106–120.

Kuhn D. (2012). The development of causal reasoning. *Wiley interdisciplinary reviews. Cognitive science*, *3*(3), 327–335. https://doi.org/10.1002/wcs.1160.

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30(4), 856–876. https://doi.org/10.1037/0278-7393.30.4.856.

Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive psychology*, *61*(4), 303-332.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 250–269. https://doi.org/10.1016/S0022-5371(84)90170-1.

Marchant, N., & Chaigneau, S. E. (2020). Modulating the coherence effect in causal-based processing. In S. Denison., M. Mack, Y. Xu, & B.C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 2527 - 2531). Cognitive Science Society.

Mayrhofer, R., & Rothe, A. (2012). Causal Status meets Coherence: The Explanatory Role of Causal Models in Categorization. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, 743–748.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & Review, 15*(1), 75-80. doi:10.3758/PBR.15.1.75.

Michotte, A. E. (1963). The perception of causality. New York: Basic Books. (Original published in 1946).

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 572.

Packard, M.G. & Knowlton, B.J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience, 25*(1), 563-593.

Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the Markov property in causal reasoning. *Cognitive Psychology, 67*(4), 186-216. doi:10.1016/j.cogpsych.2013.09.002

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review*, *123*(4), 392.

Rehder, B. (2003a). Categorization as casual reasoning. *Cognitive Science*, 27(5), 709–748. https://doi.org/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. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 347–376). New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.39

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. https://doi.org/10.1037/0278-7393.32.4.659

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

Rips, L. J. (2011). Causation from perception. *Perspectives on Psychological Science*, 6(1), 77–97. https://doi.org/10.1177/1745691610393525.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0.

Sim, Z. L., & Xu, F. (2017). Learning higher-order generalizations through free play: Evidence from 2-and 3-year-old children. *Developmental psychology*, *53*(4), 642-651. https://doi.org/10.1037/dev0000278.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science, 348*(6230), 91-94.

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal Models and the Acquisition of Category Structure. *Journal of Experimental Psychology: General*, 124(2), 181–206. https://doi.org/10.1037/0096-3445.124.2.181.

Wisniewski, E. J. (1995). Prior Knowledge and Functionally Relevant Features in Concept Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 449–468. https://doi.org/10.1037/0278-7393.21.2.449.

Wolff, P. (2008). Dynamics and the Perception of Causal Events. In T. Shipley & J. Z In T. Shipley & J. Zacks (Eds.). *Understanding Events: From Perception to Action* (pp. 555-589). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195188370.003.0023

Zhao, B., Lucas, C. G., & Bramley, N. R. (2021). How Do People Generalize Causal Relations over Objects? A Non-parametric Bayesian Account. *Computational Brain and Behavior,* 1-23. https://doi.org/10.1007/s42113-021-00124-z.