Dissecting causal asymmetries in inductive generalization

Zeyu Xia*
School of Informatics
University of Edinburgh

Bonan Zhao*
Department of Psychology
University of Edinburgh

Tadeg Quillien
School of Informatics
University of Edinburgh

Christopher G. Lucas
School of Informatics
University of Edinburgh

Abstract

Suppose we observe something happen in an interaction between two objects A and B. Can we then predict what will happen in an interaction between A and C, or between B and C? Recent research, inspired by work on the "causal asymmetry", suggests that people use cues to causal agency to guide object-based generalization decisions, even in relatively abstract settings. When object A possesses cues to causal agency (e.g. it moves, remains stable throughout the interaction), people tend to predict that what happened will probably also occur in an interaction between A and C, but not between B and C. Here we replicate and extend this work, with the goal of identifying the cues that people use to determine that an object is a causal agent. In four experiments, we manipulate three properties of the agent and recipient objects. We find that people anchor their inductive generalizations around the agent object when that object possesses all three cues to causal agency, but removing either cue abolishes the asymmetry.

Keywords: causal reasoning; generalization; inductive biases; intuitive physics; causal asymmetry

Introduction

Suppose you add honey to your tea and find that the tea tastes sweeter. How would you generalize this newly-found knowledge? Should you infer that putting anything in your tea will make your tea sweeter, or that honey makes things sweeter in general? Generalization is a difficult problem when reasoning about interactions between objects: You may observe interactions between objects A and B, and later encounter interactions involving A and C, or one involving B and C, which features of the first interaction do you expect to generalize? Should you attribute the effect of the first interaction to properties of object A, B, or both?

In the honey and tea case, we can turn to pre-existing causal knowledge for help: we know that being sweet is a property of food items, and that sweetness can transmit to things it adds to. Hence, we should infer that honey makes things sweeter, and not that anything we put into tea makes our tea sweeter. In general, causality is a powerful guide to inductive generalization (Gelman, 2003; Rehder & Hastie, 2001), limiting the vast space of possibilities to a handful of possible ones (Griffiths & Tenenbaum, 2009; Kemp et al., 2010; Lagnado & Sloman, 2006).

But what structures causal domain knowledge? People seem to naturally impose causal roles onto objects based on how they interact, construing one object as a causal "agent" and another as a passive "recipient" (Mayrhofer & Waldmann, 2015; White, 2006), even in situations where science would not single out either of them as special. For instance, when billiard ball A collides with ball B, people tend to say that A caused B to move, even though from the point of view of Newtonian mechanics, it would be equally valid to say that ball B caused ball A to stop moving (Michotte, 1963). White (2006) summarizes such inductive biases under term "causal asymmetry", and we argue that the cues that compel people to make such causal judgments in perceptual singular causal events also serve as inductive biases that guide people's generalization predictions.

In particular, we are interested in identifying the exact cues that people use to decide causal anchors (Hafri et al., 2018; Mayrhofer & Waldmann, 2014; White, 2014). A recent study by Zhao et al. (2022) finds evidence that people anchor causal generalization predictions with respect to the agent object only, but their design was not fine-grained enough to identify what factors about these objects that lead people to treat one as the agent, and hence the special anchor for the purpose of generalization. In this paper, we present four experiments that aim to isolate these factors. Our approach brings together two strands of research: computational models of categorization (Anderson, 1991; Goodman et al., 2008; Navarro et al., 2006; Nosofsky, 2011; Rehder & Hastie, 2001), as well as research on how people construct causal representations of singular events (Lagnado et al., 2013; Mayrhofer & Waldmann, 2014; Michotte, 1963; White, 2006, 2007, 2014), from the perspective of how object interactions probe causal anchoring.

Causal asymmetries in object-based categorization

We derive our predictions from a recent computational model of object-based categorization (Zhao et al., 2022), which builds upon work on Bayesian theories of categorization and the idea of program induction (Anderson, 1991; Goodman et al., 2008; Kemp et al., 2010, 2012; Piantadosi et al., 2016).

According to the model, people conclude different causal laws upon observing different interactions between objects (Figure 1). Perceptual features are materials for the content of causal effects, and the way objects interact gives rise to causal roles such as being an agent or recipient, which decides which object's features to focus on when constructing causal laws. In the honey and tea example, taking honey as the agent object and tea as recipient leads one to conclude that things

*These authors contributed equally to this work

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

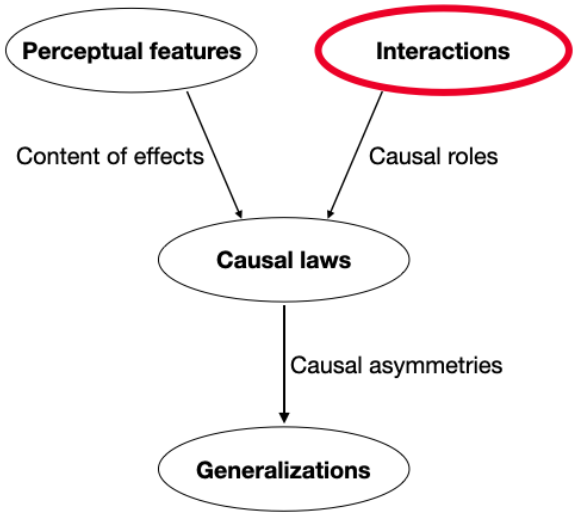

Figure 1: Schematic representation of the object-based causal categorization model (Zhao et al., 2022). The model uses perceptual features to construct the content of causal laws, and decides focus on causal roles according to how objects interact. The constructed causal laws are then applied to make generalization predictions, which exhibit causal asymmetries when the right causal cues are present. Here we are interested in how the way that objects interact influences the assignment of causal roles (red circle).

like honey transmit their taste to the recipient, and hence one may generalize that honey can make water taste sweeter too.

This model operates on a feature similarity-based categorization process and singular event causal induction mechanism. The categorization process is agnostic about which features should be taken into account—they consider all the possible perceptual features equally important. However, the model puts different weights on causal agent/recipient roles with a focus parameter, that later on can be determined by looking at how people make generalizations.

Here is how. Suppose people anchor their categorization process with the agent object, i.e., they only consider the agent object's features when deciding how general a causal law should be. If we allow participants to observe six interactions that always involve the same agent object along with a range of different recipient objects (Figure 2A, 'Fixed-L' panel), people will be biased to assume that all these interactions are ruled by the same causal law, and hence use all of the six observations to infer what this causal law consists of. Contrast to this case, if we allow participants to observe six interactions where the agent object is different every time and the recipient object is the same (Figure 2A, 'Fixed-R' panel), then people will (by assumption) infer that each interaction is determined by its own causal law, and only have one trial's worth of evidence to infer the content of each causal law.

The prediction, then, is that people in the first condition (where the agent object is always the same) should agree with each other more when they make predictions about what should happen in novel interactions, because they have abundant evidence to infer a single causal law, while people in the second condition (where the agent object is varied) are more uncertain and show less agreement with each other during generalization. If people are not biased toward anchoring on the agent object, however, then we should not observe a difference in inter-participant agreement between the two conditions. [1]

Zhao et al. (2022) reported a causal asymmetry in an experiment using a design similar to the one described above, and their results supported the bias toward anchoring on the agent object in participants' generalizations. However, their model presumes clear-cut causal roles: one object is the agent and the other object is the recipient. In fact, the agent and recipient objects in their study differed on many dimensions, making it impossible to tell how people decide on the causal focus in the first place. Here, we adapt their experimental design to narrow down on the cues that people use to anchor the categorization process (Figure 1, red circle).

Possible cues to causal roles

Figure 3A illustrates the original animation in Zhao et al. (2022). In their setup, the agent objects (on the left) differed from the recipient objects along three dimensions: the agent object was marked by a glowing yellow border; it moved toward the recipient object, which had no border. When the agent object touched the recipient object, the recipient object would change into the result form, while the agent object remained unchanged. As illustrated in Figure 1, the way objects interact with each other indicate their causal roles, and causal roles then influence the focus parameter in the categorization process.

Interactions as shown in Figure 3A convolute three possible factors: movement, change of state, and nominal indicator. Each of these factors has theoretical reasons to induce the observed asymmetry in generalizations.

Movement As shown in Michotte (1963)'s famous launching effect, movement seems to be a fundamental factor in causal perception. People watching simple physical interactions between two objects report that the moving object *causes* the state-change of the other object (Michotte, 1963; Scholl & Tremoulet, 2000). In Figure 3A, the fact that the object on the left moved might have led participants to consider that the object on the left was causally responsible for what happened to the object on the right.

Stability Change of state is another possible marker for introducing a causal asymmetry. Soo & Rottman (2018) discovered that in time series data, people are more likely to think that the object that remains stable is the cause and the object that changes is the effect. Again, in Figure 3A, the

[1] The above is an intuitive explanation for why the effect should arise – see Zhao et al. (2022) for rigorous exposition of how this prediction follows from a computational model of object-based categorization.

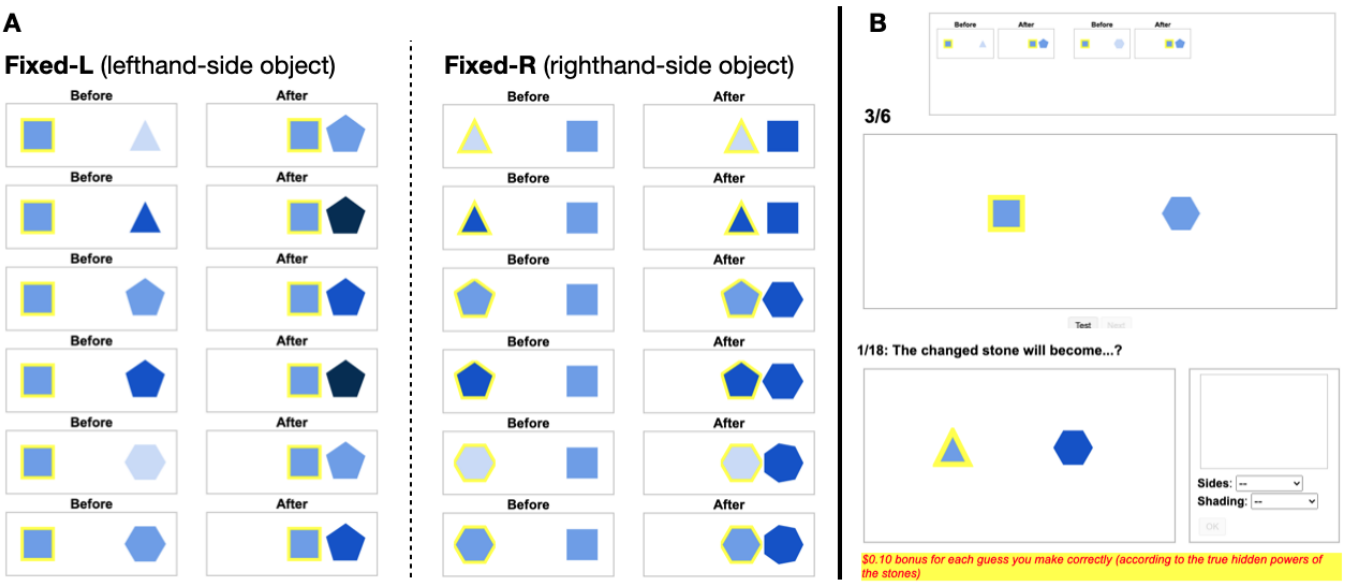

Figure 2: A. Learning pairs for the original condition. Objects marked with yellow glowing borders are the agent object. B. Task interface. Top box: visual summaries for tested learning pairs. Middle: place to play animations, triggered by clicking the "test" button. Bottom: interface for generalization predictions.

agent object does not change during the interaction, while the recipient object changes after contact with the agent object. This asymmetry in state change (stability) may be another reason for biasing the focus parameter toward the agent.

Indicator In Zhao et al. (2022)'s original experiment, the agent object was marked by a glowing yellow border, and participants were instructed that a glowing yellow border means the object is "active", and that active stones can change the other inactive stones. If participants assume that such instructions are relevant (Grice, 1975; Sperber & Wilson, 1986), they might have constructed causal laws that are anchored in the objects labeled as being active. In addition, the glowing yellow border might also have led participants to pay more visual attention to the agent objects.

Experiments

Our aim is trying to identify which of these cues are responsible for the causal asymmetry Zhao et al. (2022) found in their data. We test one of the three factors (movement, change of state/stability, and visual-nominal indicator) separately, and measure their effect on the level of inter-person agreement in causal generalizations using a similar "keeping one object constant" design (Figure 2A). As explained above, if the agent object (the object on the left in the original design) embodies cues that people use to anchor categorization, then we should observe higher inter-participant agreement in the condition where the agent object remains the same across interactions than in the condition where it varies.[2]

[2]While we could in principle also look at the proportion of participants' correct responses (i.e. responses that match the ground truth used to generate the training examples), this information is less helpful because there are many possible hypotheses, in addition to the ground truth rule, that are consistent with the data.

Methods

Participants Two-hundred-and-two participants were recuited from Amazon Mechanical Turk (82 females, $M_{\text{age}} = 37.6 \pm 10.1$). Twenty-eight participants were excluded from analysis because they failed to provide task-relevant responses in free-text inputs, leading to one-hundred-and-seventy-four participants in total. Participants were paid both for their time and a performance-based bonus. The task took 12.5 ± 10.1 minutes.

Materials and design Objects in this experiment are composed of a shading feature, ranging in {light, medium, dark, very dark} shades of blue, and number of edges, ranging from three (triangle) to seven (heptagon). The ground truth causal relationship we used to generate the training examples is the same across four experiments: the recipient object becomes one shade darker than itself and gains one more edge than the agent object (Figure 2 & 3). Note that the final state of the recipient object is a function of both its own features and those of the agent object. Therefore, the ground truth used in generating training examples does not pre-suppose asymmetries.

Experiment 1 is a replication of Zhao et al. (2022), in which we used the original animation (Figure 3A) as in their paper and code. Here, the object on the Left is intuitively seen as a causal agent, and the object on the Right is intuitively seen as the causal recipient.

Experiment 2 aims to dissect the movement factor from the original animation. We designed an animation as in Figure 3B, where the Left object remains static while the Right object moves. When the Right object touches the Left

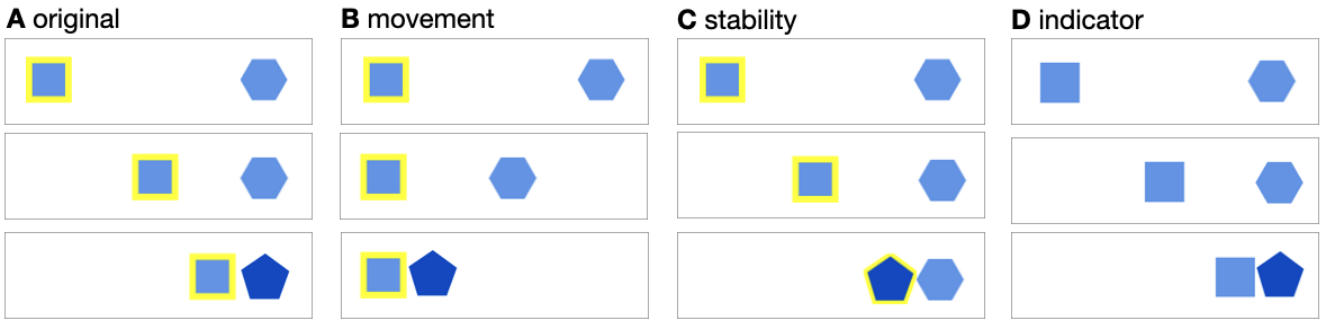

Figure 3: Illustration of various animations for the same underlying causal relationship. Objects with glowing yellow borders were told to be "active" in instructions. Full stimuli are available on OSF.

object, the moving object Right changes according to the ground truth causal relationship, while the Left object stays unchanged as in the original animation. By this animation, we removed the movement cue from the Left object.

Experiment 3 uses an animation that removes the stability cue from the Left object (Figure 3C). While keeping the indicator and movement factors identical to the original animation, it is now the Left object, rather than the original Right object, that changes into the result form after the interaction.

Experiment 4 removes the glowing yellow border from the Left object (Figure 3D), while keeping the movement and stability factors as identical to the original animation.

Following Zhao et al. (2022), for all four experiments, we manipulated (between-subjects) whether the Left object stayed the same across the six interactions while the Right object varied (fixed-L condition), or whether the Left object varied across interactions while the Right object stayed the same (fixed-R condition).

Procedure All four experiments followed the same procedure. After reading instructions and passing a comprehension quiz, participants proceeded to a learning phase, where they were invited to test causal interactions for six pairs of objects, by clicking a "test" button and watching the subsequent animated outcomes (Figure 2B, middle panel). A visual summary of each tested pair was shown after the test on top of the screen, and remained visible until the end of the experiment (Figure 2B, top panel). Next, participants were asked to write down their best guesses about the causal relationship between those objects. After that, participants went into the inductive generalization phase, where they made sixteen generalization predictions about novel pairs of objects. Each generalization task was presented sequentially and in random order. Participants composed their predictions by selecting from two drop-down menus, one for the shading feature and another for shape (Figure 2B, bottom panel).

Results

Systematic generalization Our key dependent measure is the inter-participant agreement in generalization, which we measure using Cronbach's alpha over how many participants in a given condition agree with each other in their predictions:

$$\rho_\tau = \frac{k}{k-1}\left(1 - \frac{kp(1-p)}{\sigma_X^2}\right) \tag{1}$$

In Equation 1, k is the number of participants in one condition, p is the probability of choosing an object if randomly answered ($p = 1/16$), and X is the selection vector for a generalization task. The outcome consistency measure ranges from -Inf (indicating uniform spread across all selections) and 1 (indicating a perfect agreement between participants; note that ρ_τ can only approach 1 when k tends to infinity).

For a total $4 \times 2 \times 16 = 128$ generalization tasks, the mean consistency $\rho_\tau = 0.80 \pm 074$ with max = 0.91 and min = 0.39, demonstrating a high level of agreement between participants. Fisher's exact test confirmed that for all eight between-subject conditions, participants' generalizations are not random, $p < 0.001$. Therefore, we conclude that participants made systematic generalization predictions in all eight conditions, even though there were just six data points, no strict ground truth, and potentially misleading animation types.

Causal asymmetry in generalizations Figure 4A summarizes task-wise consistency measures aggregated per condition. Experiment 1 (original) replicates the causal asymmetry in Zhao et al. (2022): participants in the fixed-L condition (original fixed-agent) made more homogeneous predictions across 16 generalization tasks ($M_{\rho_\tau} = 0.83 \pm 0.06$), and those in the fixed-R condition (original fixed-recipient) made more diverse predictions ($M_{\rho_\tau} = 0.76 \pm 0.11$), $t(15) = 1.92, p = .04$.

However, none of the other three experiments exhibits any causal asymmetry (Figure 4A). In Experiment 2 (movement, $p = .29$), Experiment 3 (stability, $p = .64$), and Experiment 4 (indicator, $p = .18$), mean consistency measures are at similar levels between fixed-L and fixed-R conditions, and no significant difference was detected. This indicates that all three factors contribute together to the original causal asymmetry effect, and removing any one of them from the Left object leads people to treating both the agent and recipient equally in generalizations.

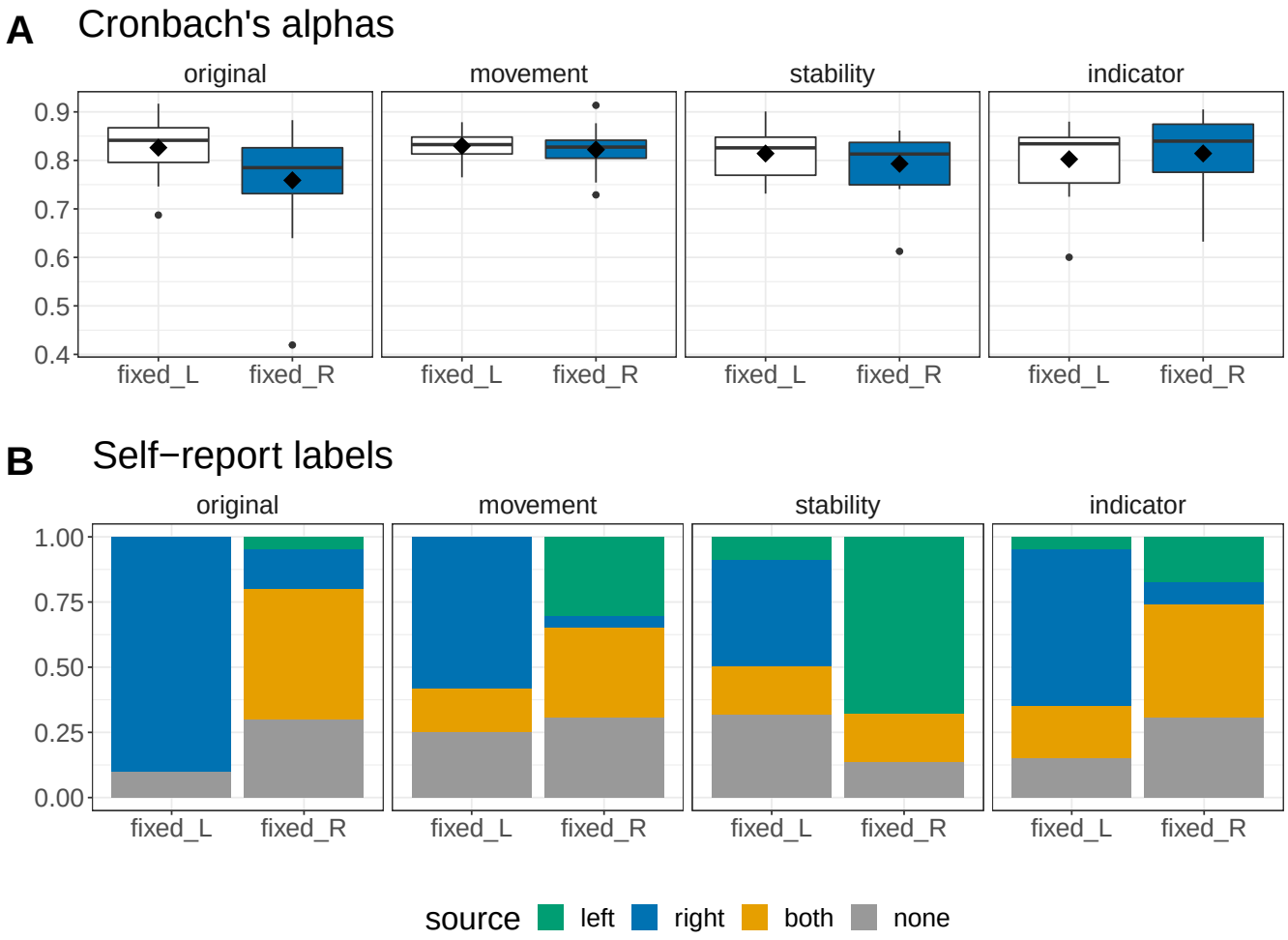

Figure 4: Experiment results. A. Generalization congruency per condition; y-axis is task-wise Cronbach's alpha value. B. Self-report labels with respect to which object's features were mentioned for inference.

Focuses in categorization To understand how people focus their categorization processes under different interaction cues, we analyzed participants' free text self-reports collected at the end of the learning phase. We coded these self-reports using *left, right, both* and *none* to represent which object people referred to when describing a causal relationship. For example, "become darker than itself" is classified as *right* in Experiments 1, 2 and 4, but as *left* in Experiment 3 (see Figure 3); "become darker than the moving stone" would be classified as *right* in Experiments 1, 3 and 4 (and as *left* in Experiment 2). Self-reports that took both objects into account are classified as *both*, such as "becomes one shade darker and converts into a shape with one more side than the active stone". Those that do not refer to objects, not consistent with data, or make no sense are classified as *none*.

Figure 4B visualizes percentages of coded self-reports for all four experiments. In Experiment 1 (original), 90% of participants in the fixed-L/fixed-agent condition reported causal relationships referring to the recipient object's feature only, while those in the fixed-R/fixed-recipient conditions showed a more diverse pattern: 50% mentioned both objects, 15% recipient-only, and 5% referring to just the agent objects' properties. A linear model predicting label *both* using the fixed condition as predictor confirms its significancy, $\beta_{\text{fixed-R}} = 0.5, p < .001$.

Strikingly, only participants in Experiment 1 showed such difference between fixed-R and fixed-L conditions. In all the other three experiments, participants showed no significant difference for label *both* in the two fixing conditions (Experiment 2: $\beta_{\text{fixed-R}} = 0.18, p = .16$; Experiment 3: $\beta_{\text{fixed-R}} = 0, p = 1$; Experiment 4: $\beta_{\text{fixed-R}} = 0.23, p = .11$). Collapsing all four experiments together, we can treat them as a 4 interaction cues $\times$ 2 fixed-L/R mixed design, and fit a multinomial regression model predicting self-report labels with these two factors. Taking label *right* and the *original* interaction cue as baselines, we found that interaction cue is indeed a significant predictor: between original and indicator cues, label *left* differs significantly, $\beta = 2.16, p = .03$; between original and movement cues, label *left* ($\beta = 2.66, p = .008$) and label *none* ($\beta = 2.19, p = .03$) both differ significantly, and be-

tween original and stability cues, all the other three labels *left* ($\beta = 4.33, p < .001$), label *both* ($\beta = 2.51, p < .001$), and label *none* ($\beta = 2.83, p < .001$) differ significantly. Fixed-L/R also appears to be a significant predictor for all three labels *left* ($\beta = -5.16, p < .001$), *both* ($\beta = -4.29, p < .001$), and *none* ($\beta = -3.93, p < .001$), but this is due to the difference between either the left or right object changes in the animations.

In sum, these coded self-reports revealed that removing either factor from the original animation shifts participants' focus to both objects in the causal interaction, resulting in a more symmetric pattern of generalization.

Discussion

In four experiments, we systematically examined what cues in causal interactions shape people's anchor of categorization in generalization. While successfully replicating the causal asymmetry in Zhao et al. (2022), we also found that this asymmetry is sensitive to a mix of factors: object movement, stability in state changes, and visual and nominal causal role indicators. The original causal asymmetry depends on all three factors working together, and removing either one of them will shift the focus of categorization, leading people to assume that the causal law that determines what happens in the interaction is a joint function of both objects.

People's tendency to parse interactions in terms of a causal "agent" and "recipient" is often derided as an irrational bias. For instance, researchers scold lay people for saying that, in a physical collision, it is the moving ball that exerted a force on the static ball, when Newtonian mechanics tell us all forces in the scene are symmetric (White, 2006). We suggest that attributing causal agency to certain objects can actually serve a functional role, and people do take into account multiple factors when making that attribution decision. The disappearance of the causal asymmetry in Experiments 2-4 demonstrates that people can be fully aware of the symmetric causal relationship when they put equal focus toward both objects in the causal interaction. The fact that Experiment 1 replicates the causal asymmetry reinforces that an overly strong causal framing may effectively structure the kind of causal laws that people infer (Gopnik et al., 2004; Griffiths & Tenenbaum, 2009; Lucas & Griffiths, 2010; Mayrhofer & Waldmann, 2015), reflected both in self-report data and inter-participant generalization agreement levels.

Humans excel at generalizing from sparse data (Anderson, 1991; Tenenbaum et al., 2011; Lake et al., 2015; Goodman et al., 2008), in part because they use assumptions about causality as inductive biases to guide generalizations (Gelman, 2003; Rehder & Hastie, 2001; Quillien, 2018). We argue that, in intuitive perceptual causality settings, people rely on interaction cues such as whether an object is moving, or remains stable throughout the interaction, to decide whether the object has causal agency, and anchor their future generalizations based on this. Unlike verbal stimuli where the cause and effect can be communicated directly, perceptual causal stimuli need a strong probing of people's intuitive causal perception (Bramley et al., 2018; Gopnik & Sobel, 2000; Ullman et al., 2017). Our results suggest that any study that aims to measure causal reasoning involving animated feature changes needs to take these interaction cues seriously.

However, since our goal was trying to dissect each cue from the original convoluted design of Zhao et al. (2022), we recognize that the new animations we designed here still mixes two factors on one object at a time. Future research could expand on these results by employing a fully factorial design, manipulating the presence or absence of each cue independently. Other kinds of experimental techniques, such as iterated learning (Kirby, 2001; Griffiths et al., 2008; Yeung & Griffiths, 2015) could also provide convergent evidence for the roles that potential cues of agency may play in object-based categorization.

Acknowledgments

T. Quillien and C. G. Lucas are supported by an EPSRC New Investigator Grant (EP/T033967/1).

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological review*, *98*(3), 409.

Bramley, N. R., Gerstenberg, T., Tenenbaum, J. B., & Gureckis, T. M. (2018). Intuitive experimentation in the physical world. *Cognitive psychology*, *105*, 9–38.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press, USA.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review*, *111*(1), 3.

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.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Griffiths, T. L., Christian, B. R., & Kalish, M. L. (2008). Using category structures to test iterated learning as a method for identifying inductive biases. *Cognitive Science*, *32*(1), 68–107.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*(4), 661.

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, *175*, 36–52.

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, *34*(7), 1185–1243.

Kemp, C., Shafto, P., & Tenenbaum, J. B. (2012). An integrated account of generalization across objects and features. *Cognitive Psychology*, *64*(1-2), 35–73.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive science*, *37*(6), 1036–1073.

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 451.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.

Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science*, *34*(1), 113–147.

Mayrhofer, R., & Waldmann, M. R. (2014). Indicators of causal agency in physical interactions: The role of the prior context. *Cognition*, *132*(3), 485–490.

Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: Dispositional intuitions as a guide to causal structure. *Cognitive Science*, *39*(1), 65–95.

Michotte, A. (1963). *The perception of causality*. Oxford, England: Basic Books.

Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006, April). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology*, *50*(2), 101–122.

Nosofsky, R. M. (2011). The generalized context model: an exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (p. 18–39). Cambridge University Press. doi: 10.1017/CBO9780511921322.002

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.

Quillien, T. (2018). Psychological essentialism from first principles. *Evolution and Human Behavior*, *39*(6), 692–699.

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.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, *4*(8), 299–309.

Soo, K. W., & Rottman, B. M. (2018). Causal strength induction from time series data. *Journal of Experimental Psychology: General*, *147*(4), 485.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Citeseer.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, *331*(6022), 1279–1285.

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in cognitive sciences*, *21*(9), 649–665.

White, P. A. (2006). The causal asymmetry. *Psychological Review*, *113*(1), 132–147.

White, P. A. (2007). Impressions of force in visual perception of collision events: A test of the causal asymmetry hypothesis. *Psychonomic Bulletin & Review*, *14*(4), 647–652.

White, P. A. (2014). Singular clues to causality and their use in human causal judgment. *Cognitive science*, *38*(1), 38–75.

Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology*, *76*, 1-29. doi: https://doi.org/10.1016/j.cogpsych.2014.11.001

Zhao, B., Lucas, C. G., & Bramley, N. R. (2022). How do people generalize causal relations over objects? a nonparametric Bayesian account. *Computational Brain & Behavior*, *5*, 22–44.