A Counterfactual Model of Causal Judgments in Double Prevention

Kevin O'Neill (kevin.oneill@duke.edu)
Department of Psychology & Neuroscience, Center for Cognitive Neuroscience, Duke University
417 Chapel Dr, Durham, NC 27708 USA

Tadeg Quillien (tadeg.quillien@gmail.com)
School of Informatics, University of Edinburgh
10 Crichton Street, Edinburgh EH8 9AB, Scotland, United Kingdom

Paul Henne (paul.henne@lakeforest.edu)
Department of Philosophy, Neuroscience Program, Lake Forest College
555 North Sheridan Road, Lake Forest, IL 60045 USA

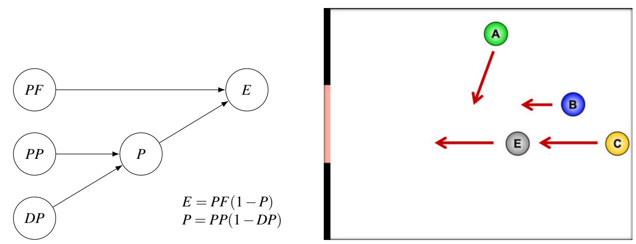
Abstract

In cases of double prevention–when one event prevents another from preventing an outcome initiated by a productive factor–people tend to judge the productive factor as causal but the double preventer as non-causal. Recent work demonstrated that this tendency can be explained by appealing to people's agreement with and tendency to consider counterfactuals: asking people to imagine the absence of the double preventer decreased their tendency to view the productive factor as more causal than the double-preventer. These effects were well-explained by the Necessity-Sufficiency (NS) model, which instantiates a particular counterfactual account. Here we asked whether another model, the Counterfactual Effect Size (CES) model, could predict the same effects. We found that the CES model indeed predicted these effects, suggesting that the ability of counterfactual theories to predict causal judgments in cases of double prevention is not unique to the NS model.

Keywords: causal judgment; counterfactual thinking; Bayesian modeling

Introduction

Mike accidentally knocked against a bottle. Seeing that the bottle was about to fall, Jack was just about to catch it when Peter accidentally knocked against him, making Jack unable to catch it. Jack did not grab the bottle, and it fell to the ground and spilled (Henne & O'Neill, 2022). In cases of double prevention such as this, people tend to judge that the productive factor (Mike) caused the bottle to spill but that the double preventer (Peter) did not (Moore, 2009; Paul & Hall, 2013). This result is usually taken as evidence against counterfactual models of causal judgment, which propose that people make causal judgments by imagining what would have happened if the cause were absent (Chang, 2009; Lombrozo, 2010). Such theories would predict that both the productive factor and the double preventer are causal, since the outcome would not have occurred had either event been absent (Lewis, 1974; Paul, 2009). However, recent work has shown that a counterfactual model known as the necessity-sufficiency model (Icard, Kominsky, & Knobe, 2017) can account for this pattern of judgments by allowing for people to imagine and endorse certain counterfactuals more often than others (Henne & O'Neill, 2022). Here we demonstrate that this explanatory power is not unique to the NS model by showing that the counterfactual effect size (CES) model (Quillien, 2020) can explain the same phenomenon.



(a) (b)

Figure 1: (a) Causal graph of the double prevention causal structure. The effect E occurs if the productive cause PF occurs and the prevention P does not occur. The prevention occurs if the possible preventer PP occurs and the double preventer DP does not occur. (b) Stimulus used in Henne and O'Neill (2022). A: possible preventer, B: double preventer, C: productive factor, E: effect.

The Counterfactual Effect Size Model

The CES model assumes that when making causal judgments, people imagine alternative worlds by sampling from a prior probability distribution over the occurrence of events. For each imagined world, people mutate the value of the candidate cause and check whether the effect has changed accordingly. Averaged across all imagined worlds and placed onto a standardized scale, this process approximates the correlation coefficient between the candidate cause and effect under the assumption of no confounders (Quillien, 2020). Since this assumption holds for double prevention (Figure 1a), the CES model predicts that the causal strength $\kappa_{C\to E}$ of each variable is simply the correlation coefficient between it and the effect. For the case of Bernoulli variables, it is

$$\kappa_{C\to E} = \frac{\sigma_C}{\sigma_E} b_{C,E} = \sqrt{\frac{P(C)(1-P(C))}{P(E)(1-P(E))}} \left[P(E|C) - P(E|\neg C)\right]$$

where $b_{C,E}$ is the regression coefficient. So, to predict causal judgments of PF and DP, we need only determine the conditional probabilities $P(E|PF)$, $P(E|\neg PF)$, $P(E|DP)$, and $P(E|\neg DP)$. Substitution in the causal graph yields $E = PF\ (1 - PP\ (1 - DP)) = PF\ (1 - PP) + (PF)(PP)(DP)$. From here it is easy to see that $P(E|PF) = 1 - P(PP) + P(PP)\ P(DP)$, $P(E|\neg PF) = 0$, $P(E|DP) = P(PF)$, and $P(E|\neg DP) = P(PF)\ (1 - P(PP))$, giving us $b_{PF,E} = 1 - P(PP) + P(PP)\ P(DP)$ and $b_{DP,E} = P(PF)\ P(PP)$. Intuitively, the productive factor makes a difference to the outcome when PP is absent or when PP is present but DP is also present; the double-preventer makes a difference to the outcome when both PF and DP are present.

Methods

To determine whether the CES model can account for causal judgments in cases of double prevention, we fit the model to data from Henne and O'Neill (2022) Experiment 4 using the probabilistic programming language Stan (Carpenter et al., 2017). In this experiment, 408 participants judged their agreement with causal statements of the productive cause and the double preventer in response to a video stimulus (Figure 1b). Before making this judgment, participants were either asked to describe what happened in the video (Control) or imagine what would have happened if the double preventer were absent (Manipulation). Henne and O'Neill (2022) found that this manipulation reduced the difference in judgments of the productive cause and the double preventer by making participants more likely to consider counterfactuals to the double preventer. To fit the model, we assumed uniform priors over the model parameters and sampled four chains for 10,000 iterations each. To ensure identifiability, we assumed that the probability that people imagine the counterfactuals to the productive factor and the possible preventer were constant between the control and manipulation conditions.

Results

The CES model significantly predicted causal judgments (R^2 = .39, 95% CI = [.36, .42]). Model estimates of causal judgments and the probability of imagining counterfactuals for each of the three events are depicted in Figure 2. In the control condition, it predicted that causal judgments of the productive factor (Md = .89, 95% CI = [.86, .91]) were higher than those of the double preventer (Md = .44, 95% CI = [.40, .47], β = .45, 95% CI = [.39, .51], $BF > 10^{16}$). It also predicted in the manipulation condition that judgments of the productive factor (Md = .78, 95% CI = [.74, .81]) were higher than those of the double preventer (Md = .59, 95% CI = [.55, .63], β = .19, 95% CI = [.12, .25], BF = 16517), though this difference was smaller than in the control condition (β = -.26, 95% CI = [-.35, -.18], BF = 454316). These patterns were qualitatively similar to those observed in people's judgments (Henne & O'Neill, 2022). The CES model predicted that people were unlikely to imagine the counterfactual for any of the three events, though it did predict a small increase in this probability for the double preventer due to the manipulation (β = .03, 95% CI = [.004, .06], BF = 27).

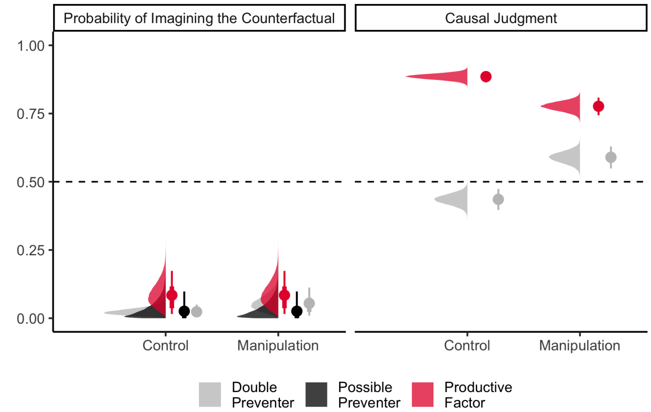


Figure 2: Probability of imagining the counterfactual and causal judgments estimated by the CES model. Thick and thin error bars represent 66% and 95% credible intervals.

Discussion

Henne and O'Neill (2022) recently found that a counterfactual model of causal judgment, the Necessity-Sufficiency model (Icard et al., 2017), can explain people's causal intuitions in cases of double-prevention. In this paper we showed that yet another counterfactual model, the counterfactual effect size model (Quillien, 2020), can predict causal judgments in such cases. These results bolster the argument that cases of double prevention do not provide a counterexample for counterfactual models of causal judgment (Henne & O'Neill, 2022). Specifically, the ability of a counterfactual framework to explain judgments in double-prevention cases does not seem to depend on the assumptions of a specific formal model. Instead, there may be something more general about the counterfactual framework that allows it to account for people's judgments. These considerations reduce the motivation for more complex pluralist accounts of causal judgment (Lombrozo, 2010).

Interestingly, the CES model makes slightly different inferences than the necessity-sufficiency model about which counterfactual possibilities people focus on. While the NS model predicted that people would often simulate counterfactual possibilities where the productive factor is absent (Henne & O'Neill, 2022), the CES model predicted that people would most often simulate counterfactual possibilities where all three events happen just as they did in the actual world. Though this tendency is consistent with work suggesting that the generation of counterfactual possibilities is biased toward what actually happened (Lucas & Kemp, 2015), we leave it to future work to explore the implications of this finding.

References

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1).

Chang, W. (2009). Connecting counterfactual and physical causation. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 31).

Henne, P., & O'Neill, K. (2022). Double prevention, causal judgments, and counterfactuals. *Cognitive Science*, *46*(5), e13127.

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80–93.

Lewis, D. (1974). Causation. *The journal of philosophy*, *70*(17), 556–567.

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

Lucas, C. G., & Kemp, C. (2015). An improved probabilistic account of counterfactual reasoning. *Psychological Review*, *122*(4), 700.

Moore, M. S. (2009). *Causation and responsibility: An essay in law, morals, and metaphysics*. Oxford University Press on Demand.

Paul, L. A. (2009). Counterfactual theories. *The Oxford handbook of causation*, 158–184.

Paul, L. A., & Hall, N. (2013). *Causation: A user's guide*. Oxford University Press.

Quillien, T. (2020). When do we think that x caused y? *Cognition*, *205*, 104410.