Norms moderate causal judgments in cases of double prevention

Kevin O'Neill
kevin.o'neill@ucl.ac.uk
Department of Experimental Psychology
University College London

Paul Henne
phenne@lakeforest.edu
Department of Philosophy
Neuroscience Program
Lake Forest College

Tadeg Quillien
tadeg.quillien@ed.ac.uk
Department of Psychology
University of Edinburgh

Thomas Icard
icard@stanford.edu
Department of Philosophy
Stanford University

Felipe De Brigard
felipe.debrigard@duke.edu
Department of Philosophy
Department of Psychology & Neuroscience
Duke University

Abstract

If Peter prevents Jack from catching a falling bottle that Mike knocked over, most people would think that Mike caused the spill to a greater degree than Peter. Cases of double prevention like these are famously inconsistent with the idea that causal judgments rely on counterfactual dependence; the spill wouldn't have happened if Mike hadn't knocked the bottle over or if Peter hadn't prevented Jack from catching the bottle. But newer counterfactual models are more flexible, and they assume that people imagine different counterfactuals in proportion with their perceived normality. Following recent work showing that these newer models can account for causal judgments in cases of double prevention, here we find that normality affects such judgments. Specifically, when the productive factor is normal and the double preventer is abnormal, we find that participants preferentially rate either the productive factor or the double preventer as more causal depending on the normality of the possible preventer. Contrary to standard interpretations, then, our results suggest that cases of double prevention are actually more problematic for competing theories of causal judgment than they are for counterfactual theories.

Keywords: causal judgment; counterfactual thinking; double prevention

Introduction

One influential idea about how people make causal judgments is that people have a counterfactual concept of causation. On this kind of view, people judge an event as causing an outcome when the outcome depends on it—i.e., when the outcome would have been different if the candidate cause had been different (Lewis, 1974).

Although such theories capture the basic intuition that causes make a difference to their effects, it is well-known that they have difficulty accommodating cases of double prevention in which a *double preventer* prevents a *possible preventer* from preventing an outcome initiated by a *productive factor*. Counterfactual theories predict that people should treat both the double preventer and the productive factor equally as causes, since the outcome would not have happened in the absence of either factor. But people typically judge the productive factor to be more causal than the double preventer. As a result, many have argued that people use a productive concept of causation in cases of double prevention, focusing on transfers of force or energy between objects (Hall, 2004; Lombrozo, 2010; Wolff & Thorstad, 2017).

But newer counterfactual theories have since developed. Under such theories, people sample a number of possibilities and then evaluate the average counterfactual dependence between the candidate cause and the effect among those possibilities (Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2021; Icard, Kominsky, & Knobe, 2017; Quillien, 2020).

Recent work suggests that these newer theories account for people's causal judgments in cases of double prevention. Henne and O'Neill (2022) found that people are more likely to agree that effects counterfactually depend on productive factors than double preventers, which partially explained their tendency to judge productive factors as more causal. Manipulations to prompt counterfactual thinking about the double preventer also reduced people's preference for productive factors as causes. Finally, when fitted to the data, counterfactual models accounted for participants' judgments (Henne & O'Neill, 2022; O'Neill, Quillien, & Henne, 2022; Quillien, O'Neill, & Henne, 2024). Together, these results highlight that counterfactual concepts of causation are *sufficient* to explain causal judgments of double prevention. Here, we report new experimental results that suggest counterfactual concepts are also *necessary* to explain cases of double prevention.

Normality and counterfactual thought

Our experiments manipulate the *normality* of events in a double prevention scenario. 'Normality' refers to a blend of statistical and normative considerations: normal events happen often or conform to moral norms (Bear & Knobe, 2017). People tend to imagine possibilities where normal events happen (Bear, Bensinger, Jara-Ettinger, Knobe, & Cushman, 2020; Bear & Knobe, 2017; Byrne, 2016; Icard, 2016). Making assumptions in line with empirical findings, counterfactual models explain why people tend to attribute more causal responsibility to abnormal events (Gill, Kominsky, Icard, & Knobe, 2022; Henne, O'Neill, Bello, Khemlani, & De Brigard, 2021; Knobe & Fraser, 2008; Kominsky & Phillips, 2019; Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015; O'Neill, Henne, Pearson, & De Brigard, 2024; Quillien & Barlev, 2022) and predict when this effect reverses (Icard et al., 2017; Morris, Phillips, Gerstenberg, & Cushman, 2019; Quillien & Lucas, 2023).

Two counterfactual theories, the Necessity-Sufficiency (NS; Icard et al., 2017) and Counterfactual Effect Size (CES; Quillien, 2020) models, make a novel prediction about the effect of normality in double-prevention cases. To understand this prediction, recall that people usually prefer to say that the productive factor is the cause (Henne & O'Neill, 2022; Lombrozo, 2010; Thanawala & Erb, 2024). The counterfactual models predict that this preference should be influenced by the normality of the *possible preventer*. Specifically:

1. When the possible preventer is abnormal, people should generally judge the productive factor as more causal than the double preventer.

2. When the possible preventer is normal, this preference should be weaker or even reversed: people should judge that the double preventer is almost as causal as (or even more causal than) the productive factor.

Counterfactual models make this prediction because the double preventer only makes a difference to the effect in possibilities where the possible preventer happens. When the possible preventer is abnormal, people will mostly imagine possibilities where the possible preventer does not happen, meaning that they should disagree that the double preventer caused the outcome. In contrast, when the possible preventer is normal, people will consider many possibilities where the possible preventer happens, and so they should agree that the double preventer caused the outcome.

The present study

Below we report two experiments that find exactly this pattern of effects. These results add pressure on theories of causal judgments relying on productive concepts (Wolff, 2007; Wolff & Thorstad, 2017). In previous findings of normality effects, the event being manipulated directly interacts with the effect (Icard et al., 2017; Knobe & Fraser, 2008; Kominsky & Phillips, 2019; Morris et al., 2019). Such normality effects could potentially be explained as a cognitive bias in which abnormality highlights the actual interaction between cause and effect. For instance, people might judge the productive factor to be more causal when it is abnormal, since the norm violation draws attention to the fact that the productive factor generated the effect. But unlike the productive factor, the possible preventer only *could have* interacted with the effect. So, it would be difficult to explain how the normality of the possible preventer could affect causal judgments without reference to the consideration of alternative possibilities.

To test our predictions, we manipulated the normality of the possible preventer in cases where the productive factor is normal and the double preventer is abnormal, since model simulations predicted our manipulation to have the strongest effect in this setting. In Experiment 1, for consistency with previous literature, we used a vignette modified from Henne and O'Neill (2022) and included a 'base' condition in which all events were normal. In Experiment 2, we designed five new vignettes affording stronger manipulations of normality.

In a crowded bar, Mike accidentally knocked against a bottle, which happens all the time. Seeing that the bottle was about to fall, Jack decided to try to catch the bottle		
Base: since it was within his reach. After starting to reach for the bottle,	*Normal:* since he was a very skilled juggler and it was easily within his reach. As it was extremely easy for him to reach the bottle,	*Abnormal:* even though he was slow and it wasn't easily within his reach. Despite the fact that it was extremely difficult for him to reach the bottle,
he was just about to grasp it when Peter		
Base: accidentally knocked against	*Normal:* intentionally pushed	*Abnormal:* intentionally pushed
him, making Jack unable to catch the bottle. Jack did not grab the bottle, and it fell to the ground and spilled.		
To what degree do you agree with the following statement?		
Productive Factor: Mike knocking into the bottle caused the bottle to spill.	*Double Preventer:* Peter [knocking into]/[pushing] Jack caused the bottle to spill.	

Table 1: Vignette for Experiment 1.

To foreshadow, we found in both experiments that the normality of the possible preventer moderates causal judgments in line with the predictions of counterfactual models.

Experiment 1

Methods

Participants Based on simulated power analyses using data from Henne and O'Neill (2022), we recruited 630 participants (309 male, 303 female, 18 other) via Prolific. All participants resided in the United States, were fluent in English, had a minimum Prolific approval rating of 99%, were be paid $0.50 for completing the study, and provided informed consent in accordance with Duke University IRB. Four participants (0.63%) were excluded after reporting not having paid attention to the task, leaving a final sample of 626 participants (308 male, 300 female, 18 other).

Materials We used the vignette in Table 1 modified from Henne and O'Neill (2022). In this vignette, Mike accidentally knocks over a bottle (the productive factor), Jack tries to catch it (the possible preventer), but Peter prevents Jack from catching it (the double preventer), resulting in a spill. In all conditions, the productive factor was relatively normal. In our two primary conditions of interest, the double preventer was abnormal and we manipulated whether the possible preventer was normal or abnormal. To compare our findings with previous results in the literature, we also included a *base* condition in which all three factors were normal.

This experiment was preregistered (link), and all materials, experiment code, and analysis code are available via the Open Science Framework.

Procedure In a 3 (Condition: base/normal/abnormal) by 2 (Factor: productive factor/double preventer) between-participants design, participants were presented with a single vignette. Next, to reduce demand characteristics, they were presented with a causal statement regarding either the productive factor or the double preventer (see Table 1) and were asked to rate the extent to which they agreed with the

statement on a scale from "strongly disagree" (coded as 0) to "strongly agree" (coded as 1), with a midpoint of "neither agree nor disagree" (coded as .5).

Analyses Since causal judgments on slider scales exhibit boundary effects at both ends of the scale (O'Neill, Henne, Bello, Pearson, & De Brigard, 2022), we analyzed participants' causal judgments using Bayesian regression modeling the mean and variance parameters of an Ordered Beta distribution, fit using the `cmdstanr` interface to Stan (Gabry, Češnovar, Johnson, & Bronder, 2024; Kubinec, 2023; Stan Development Team, 2024). We used Student-$t(3,0,2.5)$ priors on the intercepts, $\mathcal{N}(0,1)$ priors on all other coefficients, and $\mathcal{N}(0,10)$ priors on the cut-points. All contrasts were made on the inverse logit scale to preserve linearity. We used a Bayesian analog of the p-value computed from the probability of direction to test for effect existence, and we used Bayes Factors to test for effect significance (Makowski, Ben-Shachar, Chen, & Lüdecke, 2019).

Computational modeling We generated quantitative predictions for the Necessity-Sufficiency model (Icard et al., 2017) and the Counterfactual Effect Size model (Quillien, 2020; Quillien & Lucas, 2023), two counterfactual theories. For space constraints, we refer the reader to the original references for description of these models (see also supplementary information for derivations of model predictions). One of the inputs to these models is the sampling probability of each event, which by hypothesis reflects the perceived normality of that event. Because we manipulated normality only qualitatively, we inferred these parameters from the data.

For these model-based analyses, we excluded the base condition to fix the sampling probabilities of the productive factor and the double preventer across conditions. This constraint was necessary to make the models identifiable, since each model had four free parameters (the sampling probabilities of the productive factor, the double preventer, and the possible preventer in the normal/abnormal conditions) to fit the four means. We assumed a normal likelihood on the same scale as the participant judgments for both models and we used uniform priors over all sampling probability parameters. We compared model performance using Bayesian R^2 and approximate leave-one-out cross-validated expected log pointwise predictive density (Gelman, Goodrich, Gabry, & Vehtari, 2019; Vehtari, Gelman, & Gabry, 2017).

Results

We present mean causal judgment by condition and factor in Figure 1. In the base condition, there was strong evidence that causal judgments were higher for the productive factor (M = .80, 95% CI = [.75, .84]) compared to the double preventer (M = .44, 95% CI = [.38, .49], β = .31, 95% CI = [.27, .37], $P = 0$, $BF > 10000$). When the double preventer was abnormal but the possible preventer was normal, there was only weak evidence that judgments of the productive factor (M = .69, 95% CI = [.63, .75]) were higher than the double preventer (M = .60, 95% CI = [.54, .65], β = .08, 95% CI = [.01, .15], $P = .03$, $BF = 2.99$). When both the double preventer and possible preventer was abnormal, the productive factor (M = .80, 95% CI = [.75, .84]) also had higher causal judgments than the double preventer (M = .53, 95% CI = [.47, .59], β = .24, 95% CI = [.17, .30], $P = 0$, $BF > 10000$). Critically, there was strong evidence that the preference for the productive factor was stronger when the possible preventer was abnormal compared to when it was normal (β = .16, 95% CI = [.07, .25], $P < .001$, $BF = 81$).

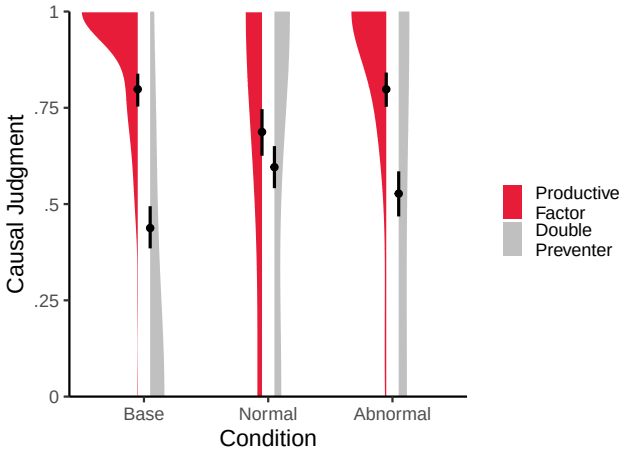

Figure 1: Mean causal judgments and 95% credible intervals by condition (base, normal, or abnormal) and factor (productive factor or double preventer) for Experiment 1.

Model-based analyses Both counterfactual theories captured a non-trivial portion of the variance in participants' judgments (NS: R^2 = .36, 95% CI = [.33, .40], $ELPD\_LOO = -117.7$, $SE = 16.8$, Figure 2; CES: R^2 = .36, 95% CI = [.32, .39], $ELPD\_LOO = -105.5$, $SE = 13.5$, $\Delta ELPD\_LOO = 12.2$, $SE = 5.6$, Figure 3). Table 2 gives the values of the sampling propensities inferred from the data. For both models, the inferred parameters are consistent with our assumption that people would be more likely to imagine the possible preventer when it was normal than when it was abnormal, although this difference is relatively small for the NS model.

Discussion

In this experiment, we replicated a standard pattern of causal judgments in cases of double prevention: people tend to judge the productive factor as more causal than the double preventer (Hall, 2004; Henne & O'Neill, 2022; Lombrozo, 2010). Importantly, in line with the predictions of counterfactual accounts, we also found that this preference for the productive factor was significantly reduced when the possible preventer was normal compared to when it was abnormal.

Experiment 2

Experiment 2 is a replication of Experiment 1 with five new vignettes permitting stronger manipulations of normality.

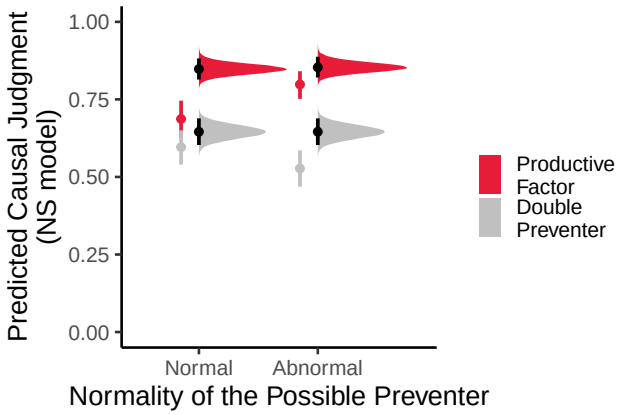

Figure 2: Predictions of mean causal judgment by condition and factor using the Necessity-Sufficiency model. Colored points and errorbars depict empirical means and 95% CIs, black points and distributions depict model predictions.

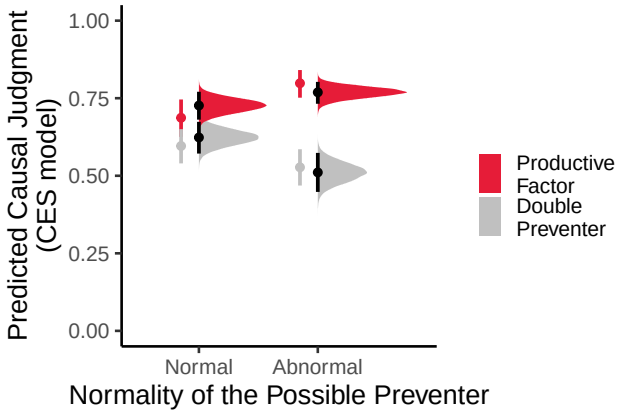

Figure 3: Predictions of mean causal judgment by condition and factor using the Counterfactual Effect Size model. Colored points and errorbars depict empirical means and 95% CIs, black points and distributions depict model predictions.

Model	Factor	Normality	Estimate
NS	Productive Factor		.40 [.31 .50]
	Double Preventer		.60 [.50, .69]
	Possible Preventer	Normal	.96 [.83, 1.00]
		Abnormal	.93 [.76, 1.00]
CES	Productive Factor		.89 [.78, .99]
	Double Preventer		.91 [.81, .99]
	Possible Preventer	Normal	.95 [.84, 1.00]
		Abnormal	.74 [.59, .90]

Table 2: Estimated sampling probabilities and 95% credible intervals by the Necessity-Sufficiency and Counterfactual Effect Size models. Probabilities were fixed across conditions for the productive factor and the double preventer.

James has a boring office job. Whenever he does paperwork, he always tends to be tired in the afternoon.	
Normal: James knows that coffee keeps him from getting tired. So, he drinks a cup of coffee at a nearby cafe after lunch every day.	*Abnormal:* James knows that coffee keeps him from getting tired. But, he almost never drinks coffee.
One day, James spends the morning doing paperwork. As part of his ordinary routine, he orders a cup of coffee from the cafe after lunch.	One day, James spends the morning doing paperwork. This time, though, he orders a cup of coffee from the cafe after lunch.
However, the barista at the cafe is in a bad mood. This time, he secretly gives James a cup of decaffeinated coffee, which looks and tastes the same as ordinary coffee but doesn't provide energy.	
So, James unknowingly drinks the decaffeinated coffee. As a result, he got very tired later that afternoon.	
To what degree do you agree with the following statement?	
Productive Factor: James doing paperwork in the morning caused him to be tired in the afternoon.	*Double Preventer:* The barista giving James decaffeinated coffee caused him to be tired in the afternoon.

Table 3: Coffee vignette from Experiment 2.

Methods

Participants We recruited 2100 participants via Prolific in a 2 (Normality of Possible Preventer: normal/abnormal) by 2 (Factor: productive factor/double preventer) by 5 (Vignette: allergies/coffee/heartworm/lactose/sunscreen) between-participants design. All participants resided in the United States, were fluent in English, had a minimum Prolific approval rating of 99%, were paid $0.50 for completing the study, and provided informed consent in accordance with Duke University IRB. 114 participants (5.4%) were excluded for reporting not having paid attention to the task in an explicit attention check. Data were analyzed from the remaining 2086 participants (1017 female, 1036 male, 33 other).

Materials Stimuli were five vignettes featuring different instances of double prevention (see Table 3 for an example and the supplementary materials for other vignettes). In all vignettes, the productive factor was normal, the double preventer was abnormal, and the possible preventer was either normal or abnormal. To focus on our main manipulation, we removed the *base* condition from Experiment 1. This experiment was preregistered (link) and all materials, experiment code, and analysis code are available via the Open Science Framework.

Procedure The procedure was the same as in Experiment 1, except that participants saw one of five vignettes.

Analyses We performed the analyses as in Experiment 1, with the exceptions that for this experiment we used tighter $\mathcal{N}(0,1)$ priors for the Ordered Beta cut-points, included vignette-level intercepts and slopes with $\mathcal{N}_{+}(0,1)$ priors for the vignette-level standard deviations and a LKJ(2) prior for vignette-level effect correlations. For the model-based analyses, we allowed the sampling probability parameters to independently vary by vignette and included a shift and scale parameter to linearly map model predictions to the scale of participants' causal judgments.

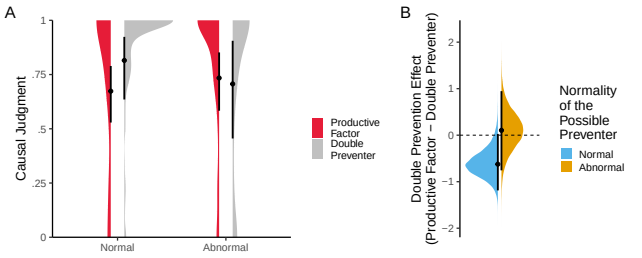

Figure 4: Mean causal judgments by factor and normality (A) and contrasts of mean causal judgment (B) with 95% credible intervals for Experiment 2.

Results

We depict mean causal judgments and contrasts in Figure 4. When the possible preventer was normal, we found weak evidence that causal judgments of the double preventer (M = .82, 95% CI = [.64, .92]) were higher than judgments of the productive factor (M = .67, 95% CI = [.53, .79], β = -.62, 95% CI = [-1.19, .03], $P = .07$, $BF = 3.72$). When the possible preventer was abnormal, there was weak evidence against a difference between judgments of the double preventer (M = .71, 95% CI = [.46, .91]) and the productive factor (M = .73, 95% CI = [.58, .85], β = .11, 95% CI = [-.76, .95], $P = .80$, $BF = .52$). There was moderate evidence that the difference in judgments was larger when the possible preventer was normal (β = .73, 95% CI = [.10, 1.26], $P = .04$, $BF = 6.69$).

Given limited evidence for the predicted effect with such a large sample size, we reasoned that there may be vignette-level differences obscuring our results. We indeed found such differences (see Figure 5). All vignettes except the Allergies vignette (β = .26, 95% CI = [-.07, .59], $P = .13$, $BF = .63$) exhibited strong reversal effects when the possible preventer was normal (all $\beta < -.73$, 95% CI = [-1.02, -.44], all $P = 0$, all $BF > 2300$). The abnormal condition was more varied, with some vignettes showing a strong preference for the productive factor (Allergies vignette, β = 1.75, 95% CI = [1.42, 2.09], $P = 0$, $BF > 10000$), others showing no preferences at all (Sunscreen vignette, β = -.19, 95% CI = [-.49, .11], $P = .20$, $BF = .25$), and still others showing a strong preference for the double preventer (Lactose vignette, $\beta = -1.02$, β = -1.02, 95% CI = [-1.32, -.73], $P = 0$, $BF > 10000$). There was strong evidence for an interaction in the Allergies (β = 1.49, 95% CI = [1.02, 1.95], $P = 0$, $BF > 10000$), Coffee (β = 1.23, 95% CI = [0.83, 1.67], $P = 0$, $BF > 10000$), and Sunscreen (β = 0.61, 95% CI = [0.23, 0.98], $P = .003$, $BF = 21.3$) vignettes, moderate evidence in the Heartworm (β = 0.50, 95% CI = [0.12, 0.86], $P = .01$, $BF = 6.17$) vignette, and weak evidence in the Lactose (β = 0.40, 95% CI = [-0.01, 0.80], $P = .055$, $BF = 1.47$) vignette. Overall, participants reliably judged the double preventer as more causal than the productive factor when the possible preventer was normal, but were made more varied judgments when it was abnormal.

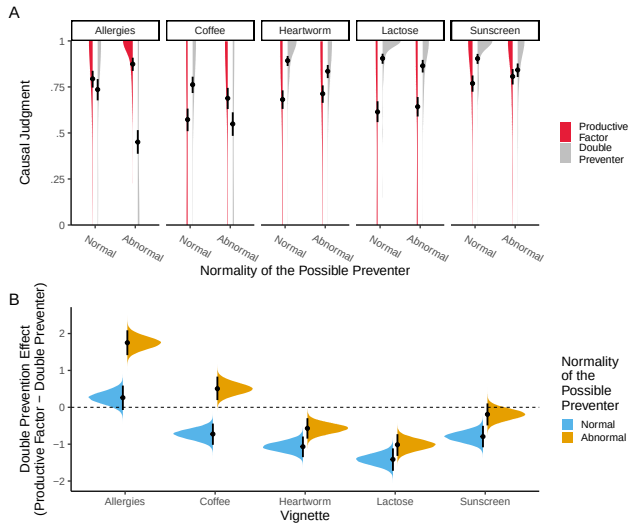

Figure 5: Mean causal judgments by factor, normality, and vignette (A) and contrasts of mean causal judgment (B) with 95% credible intervals for Experiment 2.

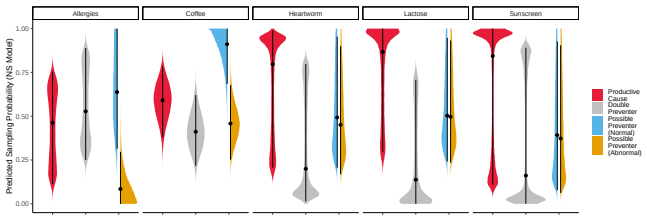

Figure 6: Estimated counterfactual sampling probabilities and 95% credible intervals by the Necessity-Sufficiency model. Sampling probabilities were fixed across conditions for the productive factor and the double preventer.

Model-based analyses An outstanding question is whether this between-vignette variability in causal judgments is consistent with counterfactual models. Under the counterfactual framework, variability across vignettes could be explained by people perceiving events to be more or less normal across different vignettes. To assess this possibility, we fit the models while allowing the normality of events to differ across vignettes. Although this approach introduced many free parameters (likely inflating estimates of model fit), it allowed us to see if our observed variability was *in principle* consistent with counterfactual theories.

Both the NS model (R^2 = .37, 95% CI = [.35, .39], $ELPD\_LOO = -145.4$, $SE = 55.1$) and the CES model (R^2 = .38, 95% CI = [.36, .40], $ELPD\_LOO = -115.6$, $SE = 55.9$) predicted individual causal judgments well. Looking at item-level correlations between model predictions and mean causal judgments, both the NS model ($r(20)$ = .84, 95% CI = [.78, .90]) and the CES model ($r(20)$ = .96, 95% CI = [.92, .98]) explained most of the variation in mean judgments (Figure 8). Model comparison revealed that the CES model made

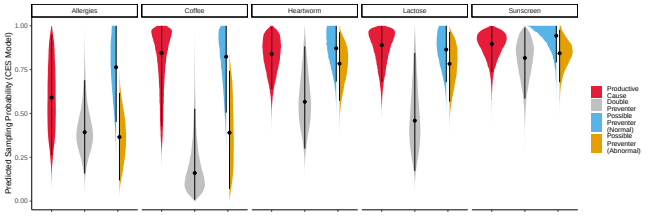

Figure 7: Estimated counterfactual sampling probabilities and 95% credible intervals by the Counterfactual Effect Size model. Sampling probabilities were fixed across conditions for the productive factor and the double preventer.

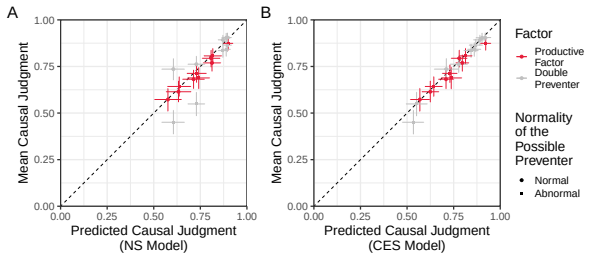

Figure 8: Causal judgments predicted by the Necessity-Sufficiency model (A) and the Counterfactual Effect Size model (B) against mean causal judgment, separately by vignette and factor with 95% credible intervals.

slightly better predictions of causal judgments than the NS model ($\Delta ELPD\_LOO = 29.8$, $SE = 7.1$).

For both models, the inferred parameters were consistent with our assumption that the possible preventer was perceived as more normal when it was in fact normal, though this difference was small for some vignettes (Figures 6 and 7).

Discussion

Similar to Experiment 1, we found that the normality of the possible preventer modulated causal judgments of double prevention. Using a stronger manipulation of normality, we found that this manipulation can reverse participants' tendency to judge productive factors as more causal than double preventers: when the possible preventer was normal, participants often judged the double preventer as more causal. Counterfactual models explained the variability in such effects through differences in perceived normality, which suggests that normality may also explain differences in effect sizes across previous studies (Henne & O'Neill, 2022; Lombrozo, 2010; Thanawala & Erb, 2024).

General Discussion

Cases of double prevention—where a double preventer prevents a possible preventer from preventing an outcome initiated by a productive factor—are a litmus test for theories of causal judgment. In such cases, people tend to judge the productive factor as more causal than the double preventer. But people should, under a classic counterfactual notion of causation, judge the two events equally. Many interpret this as evidence that people sometimes make use of a productive concept of causation in which causes directly interact with effects (Hall, 2004; Lombrozo, 2010; Wolff & Thorstad, 2017).

However, recent work has found that two newer counterfactual theories are able to explain double prevention well (Henne, 2023; Henne & O'Neill, 2022; O'Neill, Quillien, & Henne, 2022; Quillien et al., 2024). In particular, Henne and O'Neill (2022) found that ratings of counterfactual dependence partially explained participants' preference for the productive factor, that manipulations of counterfactual thinking reduced this preference, and that counterfactual models fit to the data reproduced these effects.

Henne and O'Neill (2022) showed that one does not need to appeal to productive concepts of causation to explain causal judgments of double prevention: instead, a counterfactual concept is *sufficient*. Here, we aimed to show support for the stronger claim that counterfactual concepts are also *necessary*. Specifically, counterfactual theories predict that causal judgments should depend on the normality of the possible preventer: people should judge the productive factor as more causal than the double preventer when the possible preventer is abnormal, but this pattern should attenuate or reverse when the possible preventer is normal. In two experiments, we found evidence that participants' judgments are sensitive to normality in this way and that these effects were well-explained by counterfactual models (Icard et al., 2017; Quillien, 2020). Although the CES model made generally more accurate predictions than the NS model, these differences were small. So, we leave it to future work to discriminate between counterfactual models (see also Gill et al., 2022; O'Neill et al., 2024; Quillien & Lucas, 2023). As our manipulations of normality exhibited significant between-vignette variability, future work might also replicate our results in more neutral or physical stimuli (e.g., Gerstenberg & Icard, 2020; Henne & O'Neill, 2022).

To the extent that causal judgments reflect an underlying concept of causation (but see Harding, Gerstenberg, & Icard, 2025; Samland & Waldmann, 2016, for pragmatic theories), the effect of the normality of the possible preventer is difficult to explain under an alternative concept of causation. Past research has shown that people's causal judgments are sensitive to the normality of events which directly interact with an effect. For instance,where two causes contribute to an effect, Icard et al. (2017) found that people judge abnormal events as more causal when both events are necessary for the effect but less causal when each event is individually sufficient. Such effects already pose problems for theories relying on process concepts of causation, since these theories predict that people only consider actual interactions between objects when making causal judgments (Wolff & Thorstad, 2017). Since the possible preventer only *could have* interacted with the effect, our results add further pressure to this difficulty: participants' causal judgments are also sensitive to the normality of events that do not actually interact with the effect.

Acknowledgments

We would like to thank David Lagnado for helpful discussions relating to this project.

References

Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition*, *194*, 104057.

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *cognition*, *167*, 25–37.

Byrne, R. M. (2016). Counterfactual thought. *Annual review of psychology*, *67*(1), 135–157.

Gabry, J., Češnovar, R., Johnson, A., & Bronder, S. (2024). cmdstanr: R interface to 'cmdstan' [Computer software manual]. Retrieved from `https://mc-stan.org/cmdstanr/` (R package version 0.8.1, https://discourse.mc-stan.org)

Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for bayesian regression models. *The American Statistician*.

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.

Gerstenberg, T., & Icard, T. (2020). Expectations affect physical causation judgments. *Journal of Experimental Psychology: General*, *149*(3), 599.

Gill, M., Kominsky, J., Icard, T., & Knobe, J. (2022). An interaction effect of norm violations on causal judgment. *Cognition*.

Hall, N. (2004). Two Concepts of Causation. In *Causation and Counterfactuals.* The MIT Press.

Harding, J., Gerstenberg, T., & Icard, T. (2025). *A communication-first account of explanation.* Retrieved from `https://arxiv.org/abs/2505.03732`

Henne, P. (2023). Experimental metaphysics: Causation. In A. Bauer & S. Kornmesser (Eds.), *The compact compendium of experimental philosophy.* De Gruyter.

Henne, P., & O'Neill, K. (2022). Double Prevention, Causal Judgments, and Counterfactuals. *Cognitive Science*.

Henne, P., O'Neill, K., Bello, P., Khemlani, S., & De Brigard, F. (2021). Norms affect prospective causal judgments. *Cognitive Science*, *45*(1), e12931.

Icard, T. F. (2016). Subjective probability as sampling propensity. *Review of Philosophy and Psychology*, *7*, 863–903.

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

Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. *Moral psychology*, *2*, 441–8.

Kominsky, J. F., & Phillips, J. (2019). Immoral professors and malfunctioning tools: Counterfactual relevance accounts explain the effect of norm violations on causal selection. *Cognitive science*, *43*(11), e12792.

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, *137*, 196–209.

Kubinec, R. (2023). Ordered beta regression: a parsimonious, well-fitting model for continuous data with lower and upper bounds. *Political analysis*, *31*(4), 519–536.

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.

Makowski, D., Ben-Shachar, M. S., Chen, S., & Lüdecke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in psychology*, *10*, 2767.

Morris, A., Phillips, J., Gerstenberg, T., & Cushman, F. (2019). Quantitative causal selection patterns in token causation. *PloS one*, *14*(8), e0219704.

O'Neill, K., Henne, P., Bello, P., Pearson, J., & De Brigard, F. (2022). Confidence and gradation in causal judgment. *Cognition*, *223*, 105036.

O'Neill, K., Henne, P., Pearson, J., & De Brigard, F. (2024). Modeling confidence in causal judgments. *Journal of experimental psychology: general*, *153*(8), 2142.

O'Neill, K., Quillien, T., & Henne, P. (2022). A counterfactual model of causal judgments in double prevention..

Quillien, T. (2020). When do we think that *x* caused *y*? *Cognition*, *205*. doi: 10.1016/j.cognition.2020.104410

Quillien, T., & Barlev, M. (2022). Causal judgment in the wild: Evidence from the 2020 u.s. presidential election. *Cognitive Science*, *56*(2). doi: 10.1111/cogs.13101

Quillien, T., & Lucas, C. G. (2023). Counterfactuals and the logic of causal selection. *Psychological Review*. doi: 10.1037/rev0000428

Quillien, T., O'Neill, K., & Henne, P. (2024, Sep). *A counterfactual explanation for recency effects in double prevention scenarios: commentary on thanawala & erb (2024).* PsyArXiv. Retrieved from `osf.io/preprints/psyarxiv/uv6en` doi: 10.31234/osf.io/uv6en

Samland, J., & Waldmann, M. R. (2016). How prescriptive norms influence causal inferences. *Cognition*, *156*, 164–176.

Stan Development Team. (2024). Stan modeling language users guide and reference manual [Computer software manual]. Retrieved from `https://mc-stan.org/`

Thanawala, H., & Erb, C. D. (2024). Revisiting causal pluralism: Intention, process, and dependency in cases of double prevention. *Cognition*, *248*, 105786.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, *27*, 1413–1432.

Wolff, P. (2007). Representing causation. *Journal of experimental psychology: General*, *136*(1), 82.

Wolff, P., & Thorstad, R. (2017). Force dynamics. *The Oxford handbook of causal reasoning*, 147–168.