



Cognitive Science 46 (2022) e13127

© 2022 Cognitive Science Society LLC

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13127

Double Prevention, Causal Judgments, and Counterfactuals

Paul Henne,^a  Kevin O'Neill^b 

^a*Department of Philosophy, Neuroscience Program, Lake Forest College*

^b*Center for Cognitive Neuroscience, Duke University*

Received 20 May 2021; received in revised form 4 January 2022; accepted 4 March 2022

Abstract

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. In double-prevention cases like these, philosophers and nonphilosophers alike tend to judge that Mike knocking into the bottle caused the beer to spill and that Peter knocking into Jack did not cause the beer to spill. This difference in causal judgment is a difficult puzzle for counterfactual theories of causal judgment; if each event had not happened, the outcome would not have, yet there is a difference in people's causal judgments. In four experiments and three supplemental experiments, we confirm this difference in causal judgments. We also show that differences in people's counterfactual thinking can explain this difference in their causal judgments and that recent counterfactual models of causal judgment can account for these patterns. We discuss these results in relation to work on counterfactual thinking and causal modeling.

Keywords: Causal judgment; Double prevention; Counterfactual thinking; Experimental philosophy

1. Introduction

Mike, Jack, and Peter were at a party, and there was an open bottle of beer on the table. Mike turned around and accidentally knocked against the bottle. Jack saw that the bottle was about to fall. Since it was easily within his reach, Jack was just about to catch the bottle when Peter accidentally knocked against him, making Jack unable to catch the bottle. Jack did not manage to grab the bottle, and it fell to the ground and spilled. In double-prevention scenarios like these (Fig. 1), philosophers and nonphilosophers alike tend to judge that Mike knocking into the bottle—that is, the productive factor—caused the beer to spill and that Peter

Correspondence should be sent to Paul Henne, Department of Philosophy, Lake Forest College, 555 North Sheridan Road, Lake Forest, IL 60045, USA. E-mail: phenne@lakeforest.edu

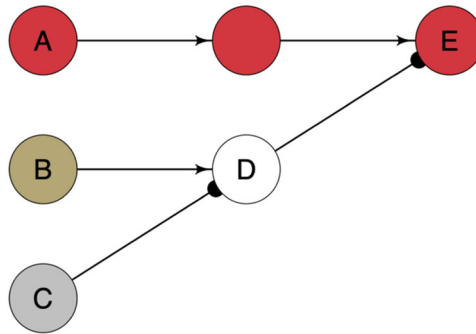


Fig. 1. Neuron diagram representing a double-prevention scenario. Each neuron represents a token-level event. Colored neurons represent events happening, and uncolored neurons represent events not happening. Arrows represent stimulatory connections between neurons, lines with black dots at the end represent inhibitory (prevention) connections between neurons. A represents the productive factor, B represents the possible preventer, C represents the double preventer, D represents the actualization of the possible preventer, and E represents the outcome. Red represents the stimulatory pathway of the productive factor, gold represents the stimulatory pathway of the possible preventer, and gray represents the inhibitory pathway of the double preventer.

knocking against Jack—that is, the double preventer—did not cause the beer to spill (Chang, 2009; Lombrozo, 2010; Moore, 2009; Paul & Hall, 2013). But why?

One of the most prominent ways to explain causal judgments is by employing counterfactual theories, on which causal judgments depend on counterfactual thinking (Gerstenberg, Goodman et al., 2021; Gerstenberg, Peterson et al., 2017; Hitchcock, 2007; Lewis, 1974; Paul, 2010). On such views, when someone wants to know if an event caused an outcome, they imagine a counterfactual alternative where the event did not happen and ask if the outcome still would have happened in that alternative scenario. For instance, if someone wanted to know if Mike knocking into the bottle caused the beer to spill, they consider a counterfactual alternative where he did not knock into the beer and ask if the beer still would have spilled. In that imagined alternative, the beer would not have spilled, so Mike knocking into the beer made a difference to the outcome—his knocking into the bottle caused the beer to spill.

While counterfactual theories explain a wide array of causal judgments (Gerstenberg, Goodman et al., 2021), cases of double prevention are a major challenge for our best theories of causal judgment (Bernstein, 2017; Hall, 2000, 2004; Hitchcock, 2010; Paul, 2010; Woodward, 2012); counterfactual theories fail to explain why people tend to accept the productive factor and reject the double preventer as the cause of the outcome in double-prevention scenarios. When someone imagines the counterfactual where Mike had not knocked into the bottle, the beer would not have spilled in this alternative scenario—*ceteris paribus*. As such, it makes sense to claim that Mike knocking into the bottle caused the beer to spill. But suppose someone imagines the counterfactual alternative where Peter had not knocked into Jack. In this scenario, too, the beer would not have spilled. However, it does not seem right to claim that Peter knocking into Jack caused the beer to spill—at least not to the same extent that Mike knocking into the bottle did—as counterfactual theories predict. For most

counterfactual theories, there is no way to explain why people tend to accept the productive factor as a cause yet deny the double preventer as a cause—although some argue that counterfactual accounts may be able to explain this tendency (Lombrozo, 2010; Woodward, 2006).

This major problem—along with some others—encourages many theorists accept that there is a productive concept of causation—where an event is a cause when it physically produces some outcome or transfers energy to it—in addition to a counterfactual concept (Hall, 2004; Lombrozo, 2010). On these kinds of pluralist theories, sometimes people employ the productive concept, and other times they employ the counterfactual concept. The productive factor in double-prevention scenarios, for instance, is a cause of the outcome because it physically produced the outcome. Mike knocking the beer physically produces the spilling of the beer, while Peter knocking into Jack has no physical impact on the spilling of the beer, so Mike knocking the beer is a productive cause, while Peter knocking Jack is not. In other kinds of scenarios—like cases of omission where there is no transfer of force—people employ the counterfactual concept of causation (Henne et al., 2017). While some argue against these accounts (Dinh & Danks, 2021; Strevens, 2013), pluralist accounts garner support from a range of empirical studies (Schlottmann & Shanks, 1992), and have been growing in popularity (Godfrey-Smith, 2009), and being used to resolve long-standing problems in philosophy (List & Menzies, 2017).

Some of the strongest empirical support for causal pluralism comes from work on double-prevention scenarios. In a series of groundbreaking studies, Lombrozo found that people tended to accept the productive factor as the cause and reject the double preventer as the cause when agents performed their actions accidentally (Lombrozo, 2010). When she described the agents as acting intentionally, however, people judged both the productive factor and the double preventer to be causes of the outcome (Lombrozo, 2010; Rose et al., 2021). Lombrozo took these findings to support a kind of causal pluralism; people use distinct concepts of causation in different cases. When, for instance, people think about unintentional action, they think about the actual mechanisms that brought about the outcome, so double preventers are not causes, whereas productive factors are. When people think about intentional actions, however, they think about agential goals, so intentional double preventers, as well as productive factors, are causes when the agents bring about the outcome intentionally (Lombrozo, 2010).

While this evidence in favor of pluralism is compelling, there may be another way for counterfactual accounts to explain the difference in causal judgments between the productive factor and the double preventer without appealing to causal pluralism—as some have suggested (Lombrozo, 2010; Woodward, 2006). Surely, we should favor a unified theory of causal judgment if it can explain causal judgments just as well as or better than a pluralistic one. Given the recent advances in counterfactual models of causal judgment, we wanted to know if the differences in causal judgments between the productive factor and the double preventer depends on differences in counterfactual thinking. People, for instance, may be more inclined to agree with a counterfactual statement about the productive factor than they would be to agree with one about the double preventer. For example, it definitely seems true that if Mike had not knocked into the beer, then the beer would not have spilled. But it seems less clear that if Peter had not knocked into Jack, the beer would not have spilled. For instance,

if Peter had not knocked into Jack, Jack still could have failed to catch the beer. If there is a difference in agreement about the counterfactuals, then this might explain the difference in causal judgments about the productive factor and the double preventer.

In series of experiments, we tested this counterfactual explanation for the double-prevention effect, that is, where people tend to agree that the productive factor is more causal than the double preventer in double-prevention scenarios. In vignette-based and video-based studies, we asked participants for their agreement with the counterfactual and causal statements about the productive factor and the double preventer, and we investigated whether differences in people's counterfactual thinking explained the differences in their causal judgments. Using a vignette-based paradigm in Experiment 1, we found that people agree more with the counterfactual and causal statements about the productive factor relative to the double preventer. Using a video-based paradigm in Experiment 2, we found again that people agree more with the counterfactual and causal statements about the productive factor relative to the double preventer. In both experiments—and in three supplemental experiments—we also found that the difference in the counterfactual judgments explained the difference in causal judgments. In Experiment 3, we manipulated the double-prevention scenario so that the counterfactual about the double preventer was just as certain as the productive factor, and we found no difference between causal judgments about the productive factor and the double preventer. In Experiment 4, we directly manipulated people's tendency to consider the counterfactual alternative to the double preventer, and we found that this manipulation decreased the size of the double-prevention effect. We then compared the results from Experiment 4 to an extension of a recent computational model of causal judgment. Together, these findings suggest that counterfactual models can accommodate causal judgments in double-prevention cases and weakens some of the motivation for causal pluralism.

2. Experiment 1

We had three aims in this experiment. First, we wanted to conceptually replicate the finding from earlier work that people are more inclined to agree that the productive factor was the cause of an outcome than they are to agree that the double preventer was the cause (Chang, 2009; Lombrozo, 2010). Second, we wanted to determine if people are more inclined to agree with the counterfactual for the productive factor than with that of the double preventer. Third, we wanted to see if this difference in considering the counterfactual explained the difference in causal judgments.

In this between-participants experiment, we gave participants a simple vignette about a double-prevention scenario that was based on a vignette used in earlier work on preempted prevention (Walsh & Sloman, 2011). In the scenario, all agents acted accidentally. We manipulated whether participants saw a counterfactual and then a causal statement about the productive factor or a counterfactual and then a causal statement about the double preventer, and then we measured participants' agreement with these statements.

Table 1
Statements used in Experiment 1

	Productive Factor	Double Preventer
Counterfactual	If Mike had not knocked against the bottle, the beer would not have spilled.	If Peter had not knocked against Jack, the beer would not have spilled.
Causal	The beer spilled because Mike knocked against the bottle.	The beer spilled because Peter knocked against Jack.

2.1. Sample size

We based the sample size on Lombrozo (2010). In Experiment 1 of her work, she reported a difference between the productive factor and the double preventer where both actions were accidental. From the reported statistics, we calculated an effect size of $d = .7$. We then calculated the number of participants required for $d = .7$ for $\alpha = .05$ at .95 power. We required 108 participants in total. Expecting a low drop-out rate, we aimed to recruit 120 participants.

2.2. Participants

We recruited all participants in all experiments in this manuscript on *Prolific*. All participants were U.S. nationals, were born in and resided in the United States, spoke English as their first language, and had a 99% approval rating on *Prolific*. A total of 120 participants completed the experiment that was programmed in Qualtrics. One participant reported not paying attention, so they were excluded. We analyzed data from the remaining 119 participants ($M_{\text{age}} = 32$, $SD = 11.1$, $\text{range}_{\text{age}} = [18-70]$, 54% female). After completing the survey, participants were compensated \$0.25.

2.3. Materials and procedure

After they consented to participation, participants were randomly assigned to one of two conditions (productive factor, double preventer) in a between-participants design. All participants read the following vignette:

Mike, Jack, and Peter were at a party, and there was an open bottle of beer on the table.

Mike turned around and accidentally knocked against the bottle.

Jack saw that the bottle was about to fall. It was easily within his reach. Jack was just about to catch the bottle when Peter accidentally knocked against him, making Jack unable to catch the bottle.

Jack didn't manage to grab the bottle, and it fell to the ground and spilled.

They then answered the counterfactual question and then causal question that were displayed on the same page as the vignette (Table 1). For each question, participants were asked

for their level of agreement with the statement on a -50 to 50 scale [-50 = strongly disagree, 0 = neutral, 50 = strongly agree]. Participants were then asked for basic demographic information and to respond to one explicit attention check that was used in all experiments in this manuscript (<https://osf.io/shpwq/>).¹

All research in this manuscript was approved by the Human Subjects Review Committee at Lake Forest College. All data collection was completed in all experiments in this manuscript prior to any analysis by the authors. All materials, Supporting Information, data, preregistrations, analysis code for all experiments in this manuscript are available at <https://osf.io/9fq5/>.

2.4. Results

Participants were more inclined to agree with the counterfactual statement about the productive factor ($M = 42.76$, $SD = 11.67$, $n = 59$) than with the counterfactual statement about the double preventer ($M = 20.50$, $SD = 23.20$, $n = 60$) ($t(87.41) = -6.62$, $p < .001$, $d = -1.42$, $CI [-1.88, -.95]$) (Fig. 3a).

Participants were more inclined to agree with the causal statement about the productive factor ($M = 41.57$, $SD = 11.98$, $n = 59$) than with the causal statement about the double preventer ($M = -17.65$, $SD = 24.20$, $n = 60$) ($t(86.61) = -16.95$, $p < .001$, $d = -3.64$, $CI [-4.33, -2.95]$) (Fig. 3b).

To investigate the relationship between counterfactual thinking and causal judgments in these double-prevention cases, we then conducted mediation analyses. The analysis examined the average causal mediation effect (ACME) of the agreement with counterfactual and the average direct effect (ADE) of condition on the causal judgments. There was a mediation effect (ACME = 6.09 , $p < .001$, $CI [1.77, 10.93]$) and there was a direct effect (ADE = 53.06 , $p < .001$, $CI [45.44, 60.74]$), suggesting that participants' agreement with the counterfactual statement partially mediated the relationship between condition (productive factor, double preventer) and their agreement with the causal statement (proportion mediated = $.10$, $p < .001$, $CI [.02, .18]$).

2.5. Discussion

In this experiment, participants judged the productive factor to be more causal than the double preventer. This result conceptually replicates earlier findings on double prevention cases (Chang, 2009; Lombrozo, 2010). Interestingly, participants were also more inclined to agree with the counterfactual about the productive factor than the counterfactual about the double preventer. While some have investigated the extent to which participants consider the counterfactual alternative to the double preventer (Gerstenberg et al., 2021), we are the first to investigate this contrast between the productive factor and the double preventer. Moreover, our mediation analysis revealed that participants' consideration of the counterfactual alternatives partially explained the difference in causal judgments between the productive factor and the double preventer.

At this point, we had a few concerns about the findings. First, most work on double-prevention cases had used within-participants designs and a dependent measure that asked about simple causal constructions using the verb "caused" rather than causal explanations

using the conjunction “because” (Lombrozo, 2010). While people distinguish between causal explanation and causal judgments, these different kinds of judgments only come apart in specific situations (Henne et al., 2017; Livengood & Machery, 2007). And we had no predictions about them coming apart in this kind of situation. Nonetheless, we wanted to ensure that people would make the same causal judgments when asked about the simple causal construction rather than the causal explanation. In Supplementary Experiment 1, we used the same materials as Experiment 1 for a within-participants design. We also modified the dependent measure such that we asked participants for their level of agreement with simple causal constructions rather than causal explanations. As we expected, we found exactly the same pattern of results as we did in Experiment 1 (proportion mediated = .20, $p = .004$, CI [.06, .40]). As such, we used the simple causal construction for our dependent measures throughout.

Second, we wondered if something about participants’ perception of the agents and their unintentional action, rather than anything about counterfactual thinking, affected people’s causal judgments. In Experiment 1, we used a vignette that involved a series of unintentional actions—in line with previous work on prevention cases (Lombrozo, 2010; Walsh & Sloman, 2011)—because we were not concerned about intentionality for this project. Nonetheless, people’s judgements shift when the actions are intentional rather than unintentional; in fact, people’s causal judgments are nearly the same for the productive factor and the double preventer when the actions are both intentional (Lombrozo, 2010; Rose et al., 2021). Because of the influence of perceived mental states on people’s causal judgments, we were worried that unintentional action could affect causal judgments in our original vignette. To investigate this concern, we developed a vignette of a double prevention case that involved no agents at all (Supplemental Experiment 2). Again, in this within-participants experiment, we found the same pattern of results (proportion mediated = .27, $p = .002$, CI [.12, .46]). As such, while intentionality affects causal judgments in double-prevention cases (Lombrozo, 2010; Rose et al., 2021), agents’ intentions were not driving this difference in causal judgments between the productive factor and the double preventer. We further explore this issue in Experiment 2.

Third, we worried that the number of counterfactual changes could have affected people’s causal judgments, rather than just their level of agreement with the counterfactual about the potential causal factor. In Experiment 1, for instance, the counterfactual statement about the productive factor is “If Mike had not knocked against the bottle, the beer would not have spilled” and the counterfactual statement about the double preventer is “If Peter had not knocked against Jack, the beer would not have spilled.” When thinking about the double preventer, people might have been thinking, however, about an embedded counterfactual like “If Peter had not knocked against Jack, then if Jack caught the bottle, the beer would not have spilled.” If people thought about this counterfactual, then they would be considering more alternative possibilities about the double preventer than the counterfactual about the productive factor.

To rule out this confound, we designed another experiment (Supplemental Experiment 3). In this experiment, we used a vignette similar to Experiment 1, and we added another event between the productive factor and the outcome’s occurrence. This design change allowed for the same number of counterfactual changes for each potential causal factor. Again, in this

within-participants experiment, we found the same pattern of results (proportion mediated = .35, $p < .001$, CI [.17, .62]). As such, we had no evidence that the number of counterfactual changes drove this difference in people's counterfactual and causal judgments.

We had one remaining concern: our vignette-based paradigm potentially left a lot up to participants' imagination. We could not be sure that they were definitely imagining these complicated causal structures in double-prevention scenarios. In order to better control the stimuli, we used video stimuli in Experiment 2.

3. Experiment 2

Our aims in this experiment were the same as those in Experiment 1 and in Supplementary Experiments 1–3. Here, we also aimed to conceptually replicate these findings using video stimuli.

3.1. Sample size

We based the sample size on Supplemental Experiment 2, a nonagential vignette. As such, we calculated the number of participants required for $d = .5$ for $\alpha = .05$ at .95 power. We required 53 participants in total. Expecting a low drop-out rate, we aimed to recruit 55 participants.

3.2. Participants

All participants were U.S. nationals, were born in and resided in the United States, spoke English as their first language, and had a 99% approval rating on Prolific. A total of 55 participants completed the experiment that was programmed in Qualtrics. One participant reported not paying attention, so they were excluded. We analyzed data from the remaining 54 participants ($M_{\text{age}} = 36$, $SD = 13.8$, $\text{range}_{\text{age}} = [18-70]$, 29% female). After completing the survey, participants were compensated \$0.30.

3.3. Materials and procedure

After they consented to participation, participants were introduced to the orientation and components of the video they would watch with an image of the beginning of the video. A red arrow identified the goal for participants, and they were told that if a ball entered the goal, it would score. In a within-participants design, all participants watched a single 15 s video of a double prevention scenario (Fig. 2a; for video, see: <https://osf.io/g7kus/>). Participants were told that they could watch the video as many times as they wanted and refer back to it as they responded to the questions.

Then they responded to all measures that were displayed on the same page as the video in fixed order. On the same scale as Experiment 1, participants answered the counterfactual questions about the productive factor and then the double preventer, and then they answered the causal questions about the productive factor and then the double preventer (Table 2).

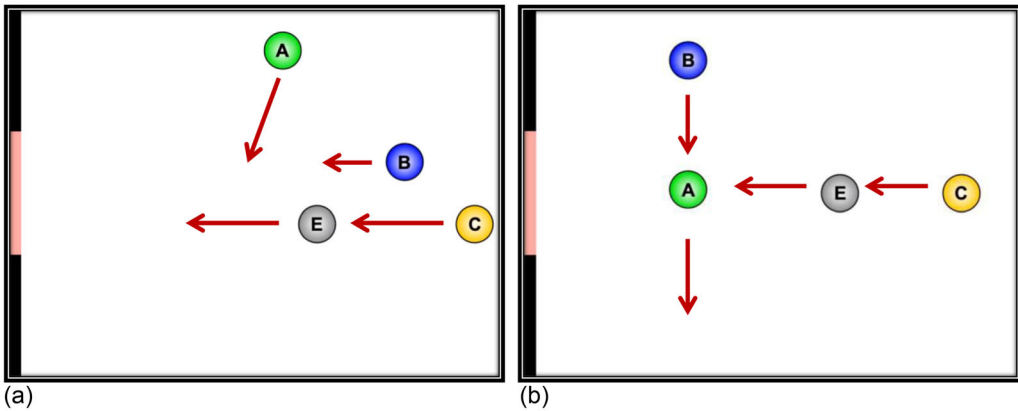


Fig. 2. (a) Video used in Experiment 2, 3 (control condition), and 4 at 0:04 s into the video. (b) Video used in Experiment 3 (manipulation condition). In each video, ball C hitting ball E was the productive factor, ball A was the possible preventer, and ball B hitting ball A was the double-prevention event.

Table 2
Statements used in Experiment 2

	Productive Factor	Double Preventer
Counterfactual	If ball C had not hit ball E, ball E would not have scored.	If ball B had not hit ball A, ball E would not have scored.
Causal	Ball C hitting ball E caused ball E to score.	Ball B hitting ball A caused ball E to score.

Participants were then asked for basic demographic information and to respond to the explicit attention check that was used in Experiment 1.

3.4. Results

Participants were more inclined to agree with the counterfactual statement about the productive factor ($M = 37.88, SD = 29.12, n = 54$) than with the counterfactual statement about the double preventer ($M = 12, SD = 44.73, n = 54$) ($t(91.09) = -3.56, p < .001, d = -.37, CI [-.58, -.16]$) (Fig. 3c).

Participants were more inclined to agree with the causal statement about the productive factor ($M = 38.87, SD = 22.74, n = 54$) than with the causal statement about the double preventer ($M = -4.16, SD = 40.86, n = 54$) ($t(82.96) = -6.76, p < .001, d = -.74, CI [-.98, -.50]$) (Fig. 3d).

To investigate the relationship between counterfactual thinking and causal judgments in these double prevention cases, we again conducted a mediation analysis. There was a mediation effect (ACME = 13.36, $p < .001, CI [6.21, 22.07]$) and there was a direct effect (ADE = 29.57, $p < .001, CI [19.18, 40.02]$), suggesting that participants agreement with the

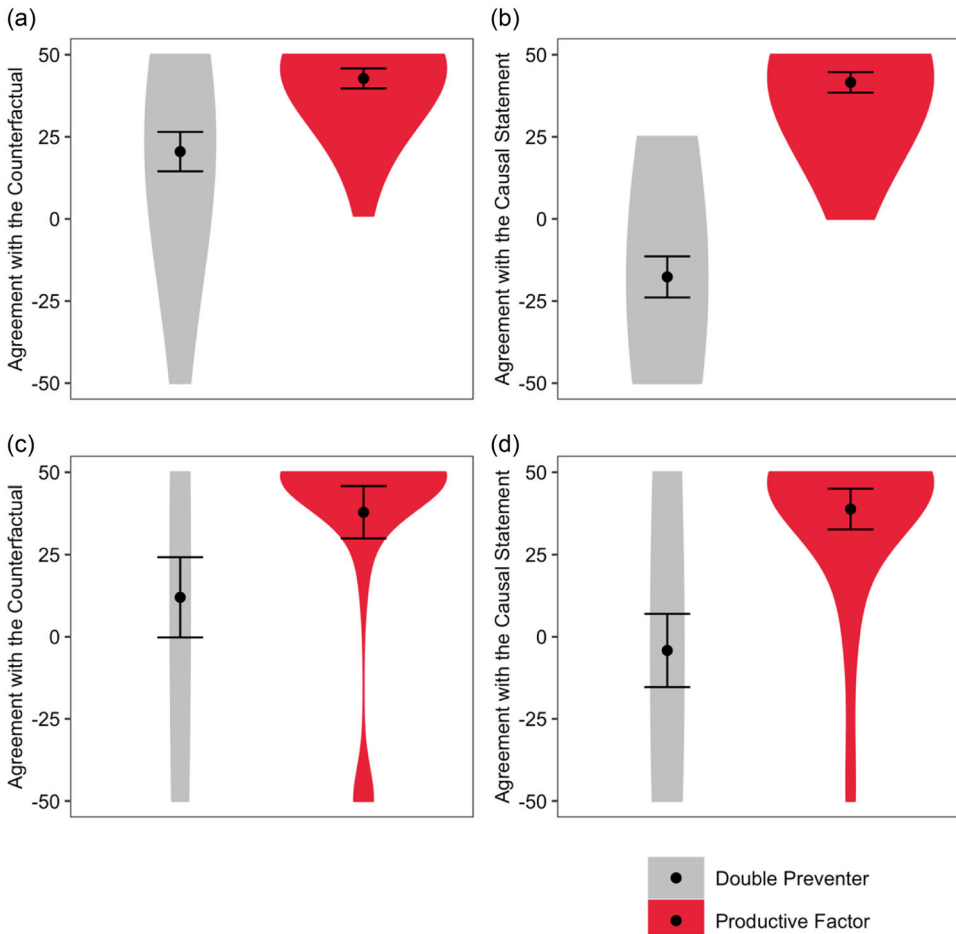


Fig. 3. (a) Mean agreement with the counterfactual statement in Experiment 1 as a function of condition. (b) Mean agreement with the causal statement in Experiment 1 as a function of condition. (c) Mean agreement with the counterfactual statement in Experiment 2 as a function of condition. (d) Mean agreement with the causal statement in Experiment 2 as a function of condition. Error bars indicate 95% confidence intervals.

counterfactual statement partially mediated the relationship between condition and their agreement with the causal statement (proportion mediated = .30, $p < .001$, CI [.16, .48]).

3.5. Discussion

Using video stimuli in Experiment 2, we found that (1) participants judged the productive factor to be more causal than the double preventer and (2) participants were more inclined to agree with the counterfactual about the productive factor than the counterfactual about the double preventer. Our mediation analysis revealed that participants' consideration of the counterfactual alternatives partially explained the difference in causal judgments between the

productive factor and the double preventer. These results conceptually replicate our findings from Experiment 1 and our supplemental experiments using visual stimuli.

While we found support for a counterfactual explanation, our results at this point were correlational. We had not manipulated the salience of the counterfactual alternatives to see if this affected causal judgments of each factor. We do this in Experiment 3.

4. Experiment 3

In this experiment, we aimed to manipulate the perceived counterfactual alternatives such that people would be more inclined to agree with the counterfactual statement about the double preventer, thus they would be more inclined to agree with the causal statement about it. To this end, we indirectly manipulated the counterfactual about the double preventer between participants by adjusting the video stimuli. In the control condition, we used the same video from Experiment 2. In the counterfactual-manipulation condition, we used a modified double-prevention video in which the angles between the billiard balls were simpler, indirectly highlighting the counterfactual about the double preventer (i.e., it highlighted that if the double preventer had not prevented the preventer, the outcome would not have occurred). Using the same measures as in Experiment 2, we used the counterfactual statement as a manipulation check. Our counterfactual explanation predicted that the double-prevention effect—that is, that people judged the productive factor as more causal than the double preventer—would be weaker or completely eliminated in the counterfactual-manipulation condition.

4.1. Sample size

Doubling our sample size from Experiment 2, we aimed to recruit 110 participants.

4.2. Participants

All participants were U.S. nationals, were born in and resided in the United States, spoke English as their first language, took the survey on a desktop or laptop, and had a 99% approval rating on Prolific. A total of 111 participants completed the experiment that was programmed in Qualtrics. One participant reported not paying attention, one participant reported that the video did not properly display, and four participants reported that they did not take the study on a desktop or laptop, so they were excluded. We analyzed data from the remaining 105 participants ($M_{\text{age}} = 34.2$, $SD = 12.7$, $\text{range}_{\text{age}} = [18-67]$, 56% female). After completing the survey, participants were compensated \$0.35.

4.3. Materials and procedure

After they consented to participation, participants were introduced to the orientation and components of the video just as in Experiment 2. Participants were randomly assigned to a control condition or a manipulation condition. In the control condition, participants watched the 15-s video from Experiment 2 (Fig. 2a; for videos, see: <https://osf.io/8a7fj/>). In the

manipulation condition, participants watched a single 15 s video of a different double prevention scenario, where we made salient the counterfactual in which ball B does not hit ball A and ball E does not score (Fig. 2b; for videos, see: <https://osf.io/cdnwg/>). In this new double-prevention video, ball A was motionless in the path of ball E. Ball B hit ball A, preventing it from preventing ball E from scoring. In this case, it was much clearer that had ball B not hit ball A, ball E would not have scored. Participants were told that they could watch the video as many times as they wanted and refer back to it as they responded to the questions. Then they responded to all measures that were displayed on the same page as the video in fixed order. The statements and scales were identical to those used in Experiment 2 (Table 2). Participants were then asked for basic demographic information and to respond to the explicit attention check that was used in Experiment 1.

4.4. Results

We analyzed agreement with the counterfactual statements and the causal statements using a 2 (Measure: productive factor vs. double preventer) \times 2 (Condition: control vs. manipulation) ANOVA.

For the counterfactual statement, there was a main effect of condition ($F(1,102) = 12.46$, $p < .001$, $\eta^2 p = .11$, CI [.02, .23]), no main effect of measure ($F(1,102) = 1.85$, $p = .17$, $\eta^2 p = .02$, CI [.00, .10]), and an interaction between the two ($F(1,102) = 5.35$, $p = .02$, $\eta^2 p = .05$, CI [.00, .15]). To decompose the interaction, we looked at the effect of measure (productive factor, double preventer) at each level of condition (control, manipulation). In the control condition, participants were more inclined to agree with the counterfactual statement about the productive factor ($M = 32.15$, $SD = 32.38$, $n = 26$) than with the counterfactual statement about the double preventer ($M = 9.15$, $SD = 47.39$, $n = 26$) ($t(102) = -2.60$, $p = .01$, $d = -.72$, CI [-1.28, -.16]) (Fig. 4a). In the manipulation condition, there was no evidence that participants were more inclined to agree with the counterfactual statement about the productive factor ($M = 39.66$, $SD = 26.71$, $n = 27$) than with the counterfactual statement about the double preventer ($M = 45.29$, $SD = 9.59$, $n = 27$) ($t(102) = .65$, $p = .51$, $d = .17$, CI [-.36, .71]) (Fig. 4a).

For the causal statement, there was a main effect of condition ($F(1,102) = 5.16$, $p = .02$, $\eta^2 p = .05$, CI [.00, .15]), a main effect of measure ($F(1,102) = 10.07$, $p = .001$, $\eta^2 p = .09$, CI [.01, .21]), and an interaction between the two ($F(1,102) = 12.32$, $p < .001$, $\eta^2 p = .11$, CI [.02, .23]). To decompose the interaction, we looked at the effect of Measure (productive factor, double preventer) at each level of Condition (control, manipulation). In the control condition, participants were more inclined to agree with the causal statement about the productive factor ($M = 39.26$, $SD = 22.80$, $n = 26$) than with the causal statement about the double preventer ($M = 0.69$, $SD = 46.22$, $n = 26$) ($t(102) = -4.72$, $p < .001$, $d = -1.31$, CI [-1.89, -.73]) (Fig. 4b). In the manipulation condition, there was no evidence that participants were more inclined to agree with the causal statement about the productive factor ($M = 32.18$, $SD = 21.14$, $n = 27$) than with the causal statement about the double preventer ($M = 33.74$, $SD = 19.81$, $n = 27$) ($t(102) = .19$, $p = .84$, $d = .05$, CI [-.48, .59]) (Fig. 4b).

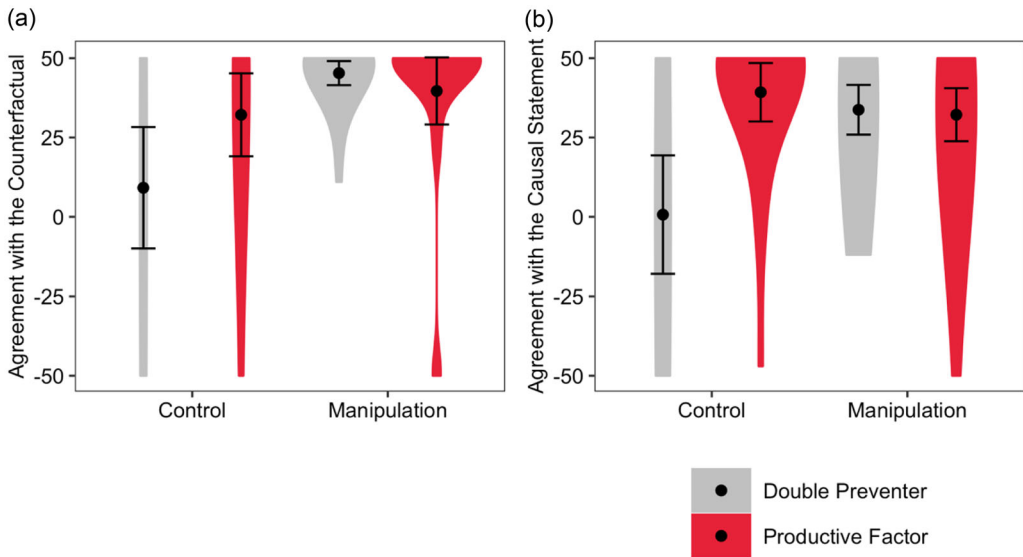


Fig. 4. (a) Mean agreement with the counterfactual statement in Experiment 3 as a function of condition and measure. (b) Mean agreement with the causal statement in Experiment 3 as a function of condition and measure. Error bars indicate 95% confidence intervals.

4.5. Discussion

In the control conditions of this experiment, we replicated the results of Experiment 2: people agreed more with the counterfactual and causal statements about the productive factor relative to the double preventer. The counterfactual manipulation that we implemented, however, weakened this double-prevention effect; in the manipulation condition, we found no evidence that people agreed more with the counterfactual and the causal statements about the productive factor relative to the double preventer. These findings bolster a counterfactual explanation for the double-prevention effect.

At this point, a reviewer noted that our counterfactual manipulation was indirect; in the manipulation condition, we intended to make the counterfactual to the double preventer more salient by manipulating the arrangement and movement of the objects in the video. This indirect technique varied from more direct manipulation techniques that encourage participants to directly imagine the counterfactual alternative directly (Henne, Kulesza et al., 2021; Kominisky & Phillips, 2019). In Experiment 4, we aimed to directly manipulate people's counterfactual considerations by asking them to imagine the counterfactual to the double preventer.

5. Experiment 4

In this experiment, we aimed to manipulate people's tendency to consider the counterfactual to the double preventer and determine if this manipulation affected their causal judgments

about it. To this end, we asked people to imagine a counterfactual where that event did not happen at all (Henne, Kulesza et al., 2021; Kominsky & Phillips, 2019). We used the same video from Experiment 2 (Fig. 2a), and we asked participants either to imagine the double preventer not occurring (manipulation condition) or to describe the events that happened in the video (control condition). We used the same measures used in Experiments 2 and 3. For this direct manipulation, our counterfactual explanation predicted that the double-prevention effect—that is, that people judged the productive factor as more causal than the double preventer—would be weaker or completely eliminated in the counterfactual-manipulation condition.

5.1. *Sample size*

Expecting a smaller effect for this direct manipulation technique, we ran a power analysis for this experiment (included in the preregistration). According to this, we aimed to recruit 420 participants.

5.2. *Participants*

All participants were U.S. nationals, were born in and resided in the United States, spoke English as their first language, took the survey on a desktop or laptop, and had a 99% approval rating on Prolific. A total of 421 participants completed the experiment that was programmed in Qualtrics. Eight participants reported not paying attention, one participant reported that the video did not properly display, three participants reported that they did not take the study on a desktop or laptop, and one participant ignored the task, so they were excluded. We analyzed data from the remaining 408 participants ($M_{\text{age}} = 34.9$, $SD = 12.2$, $\text{range}_{\text{age}} = [18-71]$, 52% female). After completing the survey, participants were compensated \$0.55.

5.3. *Materials and procedure*

After they consented to participation, participants were introduced to the components in the video and their orientation just as in Experiment 3. We added further instructions to clarify that there were no bumpers on the top or the bottom of the video and that the black bumpers were bumpers and not goals (Fig. 2a). All participants watched the 15 s video from Experiment 2 that was also the control condition in Experiment 3 (Fig. 2a; for video, see: <https://osf.io/8a7fj/>). Participants were told that they could watch the video as many times as they wanted and refer back to it as they responded to the questions. Participants were randomly assigned to the control condition or to the manipulation condition. All instructions and measures were displayed on the same page as the video. In the control condition, we asked participants to describe what happened in the video. In the manipulation condition, we asked participants to imagine a counterfactual alternative where the double preventer did not hit the possible preventer and the possible preventer hit the ball that scores (Table 3). Then participants responded to the same measures used in Experiments 2 and 3 in the same fixed order. The statements and scales were identical to those used in Experiment 2 and 3 (Table 2). Participants were then asked for basic demographic information and to respond to the explicit attention check that was used in Experiment 1.

Table 3
Counterfactual manipulation instructions used in Experiment 4

Control	Counterfactual Manipulation
Before you answer the questions below, we would like you to consider what happened in the video you just watched. Please describe what actually happened in the video. (Please write at least two or three sentences.)	Before you answer the questions below, we would like you to imagine that ball B did not hit ball A at all so that ball A hit ball E. Please describe what you imagine. (Please write at least two or three sentences.)

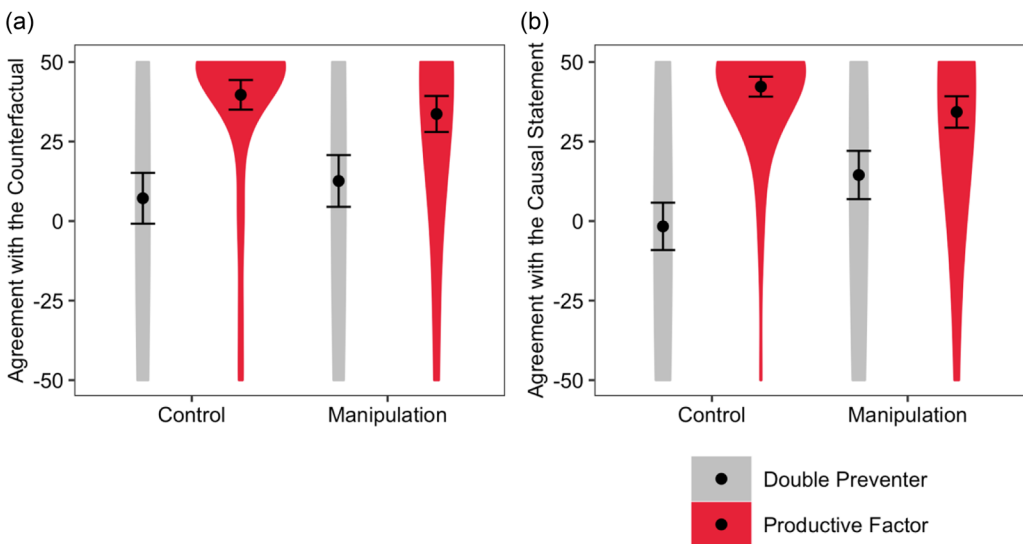


Fig. 5. (a) Mean agreement with the counterfactual statement in Experiment 4 as a function of condition and measure. (b) Mean agreement with the causal statement in Experiment 4 as a function of condition and measure. Error bars indicate 95% confidence intervals.

5.4. Results

We analyzed the data just as we did in Experiment 3.

For the counterfactual statement, there was no main effect of condition ($F(1,404) = 0.08, p = .76, \eta^2p = .00, CI [.00, .01]$), but there was a main effect of measure, indicating a double prevention effect for the counterfactual statements ($F(1,404) = 61.23, p < .001, \eta^2p = .13, CI [.08, .19]$). Notably, there was no interaction between condition and measure such that we had no evidence that the double-prevention effect for counterfactual judgments was weaker in the manipulation condition relative to the control condition ($F(1,404) = 2.78, p = .09, \eta^2p = .00, CI [.00, .03]$) (Fig. 5a).

For the causal statement, there was no main effect of condition ($F(1,404) = 3.38, p = .06, \eta^2p = .00, CI [.00, .03]$), but there was a main effect of measure ($F(1,404) = 108.21,$

$p < .001$, $\eta^2 p = .21$, CI [.15, .28]). Critically, there was an interaction between condition and measure such that the double-prevention effect for causal judgments was weaker in the manipulation condition relative to the control condition ($F(1,404) = 15.37$, $p < .001$, $\eta^2 p = .04$, CI [.01, .08]) (Fig. 5b). To decompose the interaction, we looked at the effect of Measure (productive factor, double preventer) at each level of Condition (control, manipulation). In the control condition, participants were more inclined to agree with the causal statement about the productive factor ($M = 42.24$, $SD = 15.63$, $n = 98$) than with the causal statement about the double preventer ($M = -1.66$, $SD = 38.99$, $n = 108$) ($t(404) = -10.15$, $p < .001$, $d = -1.41$, CI [-1.70, -1.12]). In the manipulation condition, participants were also more inclined to agree with the causal statement about the productive factor ($M = 34.28$, $SD = 25.65$, $n = 106$) than with the causal statement about the double preventer ($M = 14.47$, $SD = 37.39$, $n = 96$) ($t(404) = -4.53$, $p < .001$, $d = -.63$, CI [-.91, -.35]).

5.5. Discussion

Notably, the double-prevention effect for people's causal judgments was weaker in the manipulation condition relative to the control. As such, this direct manipulation of the people's tendency to consider certain counterfactuals affected their causal judgments. However, we had no evidence that people's counterfactual judgments varied as a result of this direct manipulation. Combined with the finding that agreement with the counterfactual statements only partially mediated the double prevention effect in Experiments 1 and 2, one may be tempted to argue that counterfactuals cannot explain the double prevention effect after all. Alternatively, we hypothesized that there are two different mechanisms for counterfactual thinking to affect causal judgment: the agreement with the counterfactuals and the tendency to consider the counterfactuals. We explore one way for these two mechanisms to account for our results in the next section.

6. Modeling the double prevention effect

In Experiment 4, we found that encouraging people to imagine the counterfactual where the double preventer does not happen increased causal judgments about the double preventer, thereby reducing the double-prevention effect. We also found no evidence that this manipulation increased participants' agreement with the statement about the counterfactual that we encouraged them to imagine. These findings highlight the possibility that counterfactual accounts have two interacting mechanisms that account for causal judgments: the tendency to imagine a given counterfactual and the agreement with the imagined counterfactual. Although there are other counterfactual accounts that could account for this effect (Gerstenberg et al., 2021), one was derived with the distinction between the tendency to consider a counterfactual and the agreement with the counterfactual in mind: the necessity-sufficiency model (Icard et al., 2017).

The necessity-sufficiency model assumes that when people make causal judgments, they might consider several different alternative possibilities. When they consider a possibility in

which the potential cause does not occur, people focus on necessity. When they consider a possibility in which the potential cause does occur, they instead focus on sufficiency. The model also assumes that causal judgments are a weighted average of necessity and sufficiency. The necessity-sufficiency model thus captures a natural separation between effects of agreement with a counterfactual statement (i.e., necessity and sufficiency strength) and the tendency to imagine the counterfactual (i.e., the weights of necessity and sufficiency strength). If the necessity-sufficiency model can account for the double prevention effect and for its reduction under our counterfactual manipulation in Experiment 4, then we will have evidence that existing counterfactual accounts can explain causal judgments in double-prevention cases.

6.1. Methods

To determine whether the necessity-sufficiency model can account for the double-prevention effect, we derived the predictions of this model for the causal structure depicted in Fig. 1 (see derivation on OSF: <https://osf.io/4z52p/>), and then we fit this model to the causal judgments from Experiment 4. There are three free variables in the model for each of the two conditions: the probability that people imagine the counterfactuals to the productive factor, the possible preventer, and the double preventer. Given these three probabilities, the model estimates causal judgments of the double preventer and of the productive factor. We used the probabilistic programming language Stan to fit a Bayesian version of the model to our data (Carpenter et al., 2017). We assumed uniform priors for each of the three probabilities and sampled four chains for 10,000 iterations each using Hamiltonian Markov Chain Monte Carlo. To ensure that the model was identifiable, 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. Causal judgments were rescaled to the range [0, 1] to match the scale of the model.

6.2. Results

The model posterior distributions, medians, and 95% credible intervals are depicted in Fig. 6. Critically, we found that the necessity-sufficiency model accounted for a significant amount of variance in causal judgments, $R^2 = .39$, 95% CI = [.36, .41]. As expected, the model showed that people were more likely to imagine the counterfactual to the double preventer in the manipulation condition ($Md = .50$, 95% CI = [.39, .61]) compared to the control condition ($Md = .25$, 95% CI = [.17, .35]), $\beta = .24$, 95% CI = [.14, .35], BF = 3319. The model estimated that although people rarely imagined the counterfactual to the possible preventer ($Md = .07$, 95% CI = [.00, .27]), they often imagined the counterfactual to the productive factor ($Md = .65$, 95% CI = [.56, .73]). In the control condition, participants were as likely to imagine the counterfactual to the double preventer as to the possible preventer, $\beta = -.18$, 95% CI = [-.32, .02], BF = 1.26, and were less likely to imagine the counterfactual to the double preventer than to the productive factor, $\beta = .40$, 95% CI = [.31, .47], BF = 121. In the manipulation condition, they were more likely to imagine the counterfactual to the double preventer than to the possible preventer, $\beta = -.41$, 95% CI = [-.55, -.24],

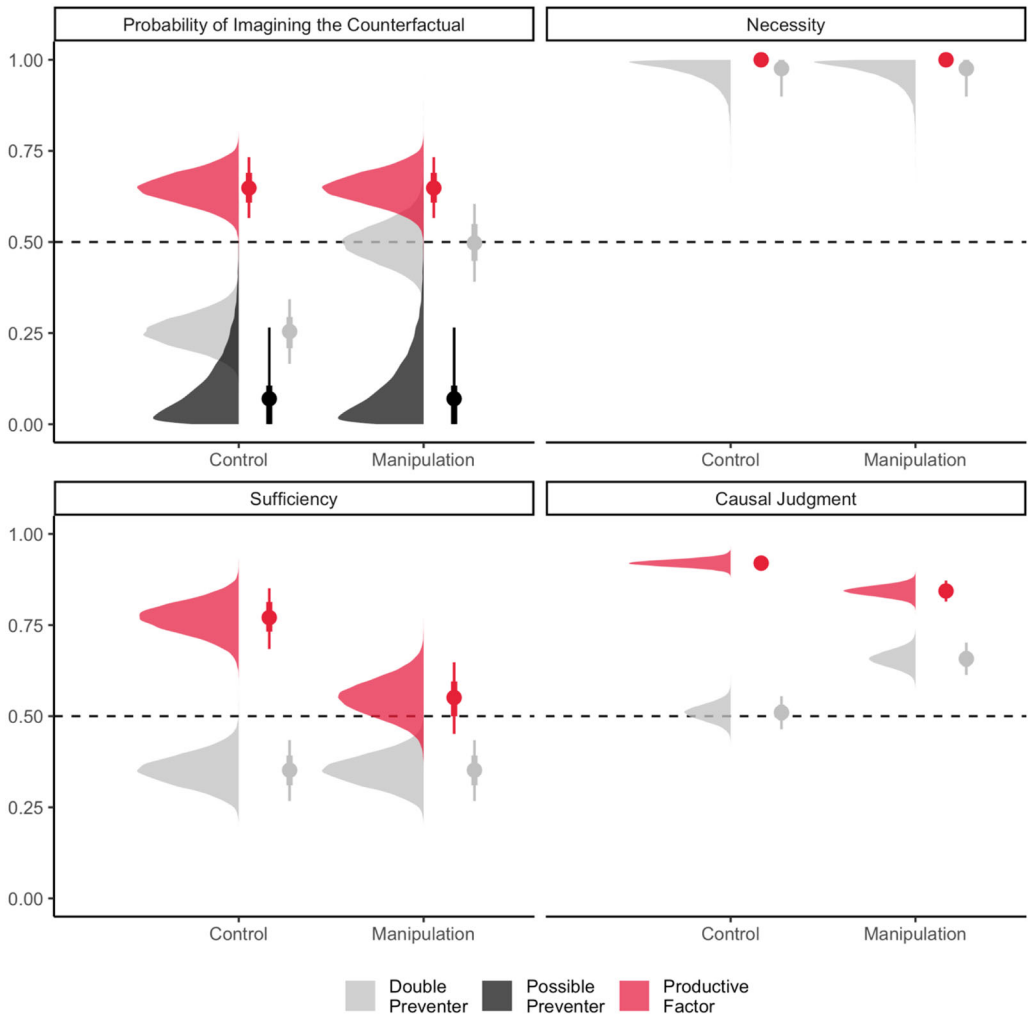


Fig. 6. Posterior distributions of the necessity-sufficiency model for Experiment 4. Points indicate posterior medians, thick error bars indicate 66% credible intervals, and thin error bars indicate 95% credible intervals.

BF = 989, and were about as likely to imagine the counterfactual to the double preventer as to the productive factor, $\beta = .16$, 95% CI = [.04, .25], BF = 3.30.

In terms of necessity and sufficiency strength, the model predicted that the double preventer was highly necessary ($Md = .98$, 95% CI = [.90, 1.0]), but only weakly sufficient ($Md = .35$, 95% CI = [.27, .44]) for the outcome. Because these parameters were functions of only the estimated probabilities that people imagined the productive factor and the possible preventer (which were constant between conditions), the model predicted that the necessity and sufficiency strength of the double preventer would not change between the control and manipulation conditions. Similarly, the model predicted that the productive factor would be

fully necessary for the effect to occur—its necessity strength was fixed to 1 and did not vary between conditions. Interestingly, however, the model predicted that people would view the productive factor as less sufficient in the manipulation condition ($Md = .55$, 95% CI = [.45, .65]) than in the control condition ($Md = .77$, 95% CI = [.68, .85]), $\beta = -.22$, 95% CI = [-.31, -.13], BF = 12128, suggesting that considering the counterfactual to the double preventer brings to mind situations in which the productive factor would not have been sufficient for the effect.

Finally, the model captured the double prevention effect on causal judgments, as well as the effect of our counterfactual manipulation. In the control condition, predicted causal judgments of the double preventer ($Md = .51$, 95% CI = [.47, .56]) were lower than those of the productive factor ($Md = .92$, 95% CI = [.90, .94]), $\beta = .41$, 95% CI = [.35, .47], BF > 40,000. Likewise, in the manipulation condition, predicted causal judgments of the double preventer ($Md = .66$, 95% CI = [.62, .70]) were lower than those of the productive factor ($Md = .84$, 95% CI = [.81, .87]), $\beta = .19$, 95% CI = [.12, .25], BF = 3983. The double-prevention effect was also estimated to be weaker in the manipulation condition than in the control condition, $\beta = -.22$, 95% CI = [-.32, -.14], BF = 7874.

6.3. Discussion

Here we tested whether the necessity-sufficiency model, a counterfactual model of causal judgments, could be extended to account for the double-prevention effect and the effect of our counterfactual manipulation observed in Experiment 4. We found that not only could the model successfully account for these effects, but it could also lend further insight as to why such effects occur in the first place. Specifically, the model identified two explanations for the double prevention effect: a difference in the perceived sufficiency of the two causes and a difference in the tendency to imagine the counterfactuals to the double preventer and the productive factor.

One explanation that the necessity-sufficiency model provides for the double prevention effect is that people see the productive factor as more sufficient than the double preventer. Although we did not foresee this prediction, it is intuitive: whereas the double preventer is only sufficient for the outcome when the productive factor occurs, the productive factor is always sufficient for the outcome—except in the case where the possible preventer occurs and the double preventer does not. Similarly, the model suggests that the reason encouraging participants to imagine counterfactual alternatives in which the double-preventer does not happen at all reduces the double prevention effect is that this manipulation encourages participants to consider cases in which the productive factor is not sufficient for the outcome, reducing the difference in sufficiency between the two causes.

The second explanation for the double prevention effect provided by the necessity-sufficiency model is that people are more likely to consider counterfactuals in which the productive factor does not happen than they are to consider counterfactuals in which the double preventer does not happen. Recall that the model proposes that when people imagine a possibility in which a potential cause does not happen, they focus on the necessity of this event for the outcome. If people are more likely to imagine the counterfactual to the

productive factor than not, the model predicts that they are more likely to pay attention to the necessity of the productive factor. Since people would see the productive factor as more necessary, the model predicts that people would judge the productive factor to make a larger difference to the outcome. In contrast, the model estimated that people tend not to imagine the counterfactual in which the double preventer does not happen. Even under a manipulation aimed at increasing people's tendency to consider this counterfactual, the model estimated that people were equally likely to imagine the double preventer occurring or not occurring. So, if people tended to imagine the double preventer happening, the model predicts that they would focus on sufficiency. Since people would see the double preventer as more necessary than it is sufficient, the model predicts that causal judgments would be lower for the double preventer. Under the manipulation, however, the model estimated that people focused equally on necessity and sufficiency, raising causal judgments about the double preventer.

In sum, we found that the necessity-sufficiency model explains the double-prevention effect through two means: through differences in the perceived sufficiency of the productive factor and the double preventer and through differences in people's tendency to weight necessity and sufficiency between the two events. Even though the necessity-sufficiency model was not derived specifically to explain these effects, it was still able to predict them. We take these results as clear evidence that counterfactual accounts of causal judgment have the capacity to explain causal judgments about double-prevention cases.

This model, however, leaves some unanswered questions. While the model estimated that people would perceive both the productive factor and the double preventer as necessary for the effect, we found in Experiment 4—as we did in Experiments 1 and 2—that people were more inclined to agree with the counterfactual statement about the productive factor than the counterfactual statement about the double preventer. Although there are a few potential explanations for this apparent difference, we will focus on one: participants could have had trouble understanding the causal structure in our materials. Since there were four moving objects on the screen, participants could have made noisy judgments about what would have happened in the absence of the double preventer. On this explanation, when people considered, for instance, the counterfactual alternative where ball B had not hit ball A, it was not entirely clear to them that ball E would not have scored. We have some evidence for this explanation in Experiment 4: even when we had participants imagine the counterfactual to the double preventer and we clarified that the outcome would not have happened (Table 3), some participants still described imagining a situation where ball E scored. This explanation suggests that the differences in agreement with the counterfactual statements in our experiments actually represents a difference in the understanding of the causal structure or confidence in the counterfactual alternatives. If this explanation of people's agreement with the counterfactual statements is right, then our results support the necessity-sufficiency model's explanation of the double-prevention effect.

The model also makes some untested predictions that should be investigated in future work. Here, we highlight two key predictions. As discussed above, the model estimated that people should judge the productive factor as more sufficient than the double preventer. While we did not measure sufficiency judgments in the present study, we expect that people's judgments of sufficiency should follow this pattern. A second prediction is that the probability that people

imagine the counterfactual in which the possible preventer does not happen should modulate the size of the double-prevention effect. Although the model estimated that people rarely imagine possibilities where the possible preventer does not happen, the productive factor is sufficient for the outcome and the double preventer is not necessary for the outcome in this alternative possibility. So, explicitly asking participants to imagine this possibility should increase the size of the double prevention effect. If these two predictions are right, then we can be certain that the necessity-sufficiency model, as opposed to other counterfactual models, explains causal judgments in double-prevention cases.

7. General discussion

In four experiments, three supplemental experiments, and a computational model we found support for a counterfactual explanation for the double-prevention effect. In Experiments 1 and 2 and the supplemental experiments, we found that people agreed more with counterfactual statements about the productive factor than those about the double preventer, and this difference in judgments accounted for the double-prevention effect. In Experiment 3, we showed that manipulating the salience of the counterfactual alternative to the double preventer eliminated the double-prevention effect altogether. In Experiment 4, we directly manipulated people's tendency to consider the counterfactual alternative to the double preventer, and we found that this manipulation weakened the double-prevention effect. We then fit these data to a particular counterfactual model—the necessity-sufficiency model—and we found that it could account for the double-prevention effect.

Our initial hypothesis was that if the double-prevention effect arises because of differences in counterfactual thinking, then people would agree more with the counterfactual statements about the productive factor than those about the double preventer and that this difference would account for the difference in causal judgments. In Experiments 1–3, we found that a double-prevention effect in counterfactual thinking accounted for the double-prevention effect in causal judgments and that indirectly manipulating people's agreement with counterfactual statements eliminated the double-prevention effect. These results are consistent with our original hypothesis.

The results of Experiment 4 and our model comparison, however, suggest that there is another way in which counterfactual thinking affects the double-prevention effect. In Experiment 4, directly manipulating people's tendency to consider the counterfactual to the double-preventer decreased the double-prevention effect for causal judgments, but the manipulation had no effect on judgments about the counterfactual statements. People's tendency to consider the counterfactual and their agreement with the counterfactual statement are distinct. So, we extended the necessity-sufficiency model (Icard et al., 2017)—because it accounts for these distinct mechanisms—and we compared it to the data from Experiment 4. The results suggest that the double-prevention effect is not entirely about the agreement with the counterfactual statement—as we initially thought it was. Rather, the results suggest that the double-prevention effect results from people's tendency to consider the counterfactuals of the potential causes and the extent to which they perceive each potential cause as modally sufficient for

the outcome. While more work is required to determine the exact details of the relationship between counterfactual thinking and causal judgment in double-prevention cases, we have laid the groundwork here for a counterfactual explanation of the double-prevention effect.

As we noted in the introduction, some philosophers and cognitive scientists take causal judgments in double-prevention scenarios to be evidence in support of causal pluralism, where there is both a productive and counterfactual concept of causation (Hall, 2004; Lombrozo, 2010). Herein, we provide some evidence that this move may not be warranted by the double-prevention effect alone. We argue that the greater strength of causal judgments about the productive factor does not necessarily suggest that this stronger causal judgment results from a greater perceived transfer of energy or oomphiness of the productive factor. This move might be too hasty. Rather, we show that counterfactual thinking and its relationship to causal judgments can account for this stronger causal judgment about the productive factor.

Nonetheless, more work needs to be done on the ability for counterfactual models to account for the double-prevention effect. While we proposed an extension of the necessity-sufficiency model, there is at least one other counterfactual relation that may explain the double-prevention effect and deserves attention. Woodward has suggested that counterfactual accounts can accommodate judgments in double-prevention cases if the account incorporates a notion of sensitivity (Woodward, 2006). A causal relation is insensitive—that is, invariant, stable, or robust—just in case the relation continues to hold if various background conditions had been different. Billy throws a bottle against a wall and it shatters. This is a very insensitive relation; even if background conditions were slightly different—for instance if the wind or the humidity had been different—the relation would hold. But consider a different case where Suzy drops a paper from the top of a building and causes Janet to get a paper cut. This relationship between dropping the paper and Janet getting a paper cut is more sensitive; if the wind, for instance, had been slightly different, Janet would not have gotten cut.

Woodward uses sensitivity alone to explain the double-prevention effect (Woodward, 2006), and some researchers suggest that this account accords with much of the empirical findings (Lombrozo, 2010). On this view, the productive factor is more insensitive than the double preventer. Consider again the example from the beginning of this article. Mike knocking into the bottle and causing it to spill is a more insensitive relation; if background conditions had been different when Mike knocked into the bottle, it seems that it still would have spilled. Peter knocking into Jack, however, seems more sensitive; Peter might have moved in a slightly different way or Jack might have not seen the bottle falling. If sensitivity affects causal judgments—which it seems to do (Vasilyeva et al., 2018)—then counterfactual accounts can also explain double-prevention cases in this way (Woodward, 2006).

Woodward's sensitivity explanation seems to accord with our findings for Experiments 1 and 2. In these experiments, people agreed more with the counterfactual statement about the productive factor than with that of the double preventer. Perhaps this difference in agreement with the counterfactual statements happens because people perceive the relation between the productive factor and the outcome as more insensitive than the relationship between the double preventer and the outcome.

However, we also have some evidence that might challenge this account. In Supplemental Experiment 3, for example, we found a double-prevention effect even when we added in

additional events to the causal sequence. That is, even when we controlled for more counterfactual changes—a potential proxy for sensitivity—people still showed a double-prevention effect. Of course, our experiment was not specifically designed to control for sensitivity and test Woodward's view, so future work should explore this view further.

One way to accommodate Woodward's account is to suggest that people independently evaluate necessity, sufficiency, and sensitivity (perhaps among other relations) when making a causal judgment. In fact, a recent computational model known as the counterfactual simulation model (CSM) does just this (Gerstenberg et al., 2021). The CSM proposes that people's causal judgments are a weighted average of different dependency relations including necessity, sufficiency, and sensitivity. Among other phenomena, the CSM has been shown to account for causal judgments in cases of double prevention (Gerstenberg et al., 2021). However, it has not been shown that sensitivity, conceptualized as a unique aspect of causation, is necessary to account for the double-prevention effect. Since the CSM is so similar to the necessity-sufficiency model, it would be helpful for future work to test discrepancies in predictions between these two promising models. For example, we discussed two unique predictions afforded by the necessity-sufficiency model in Section 6.3. If these predictions do not hold, counterfactual theorists may have reason to prefer the CSM, which can more flexibly capture such results.

There may, however, be a simpler way to accommodate Woodward's account. The results from Experiment 4 and our model fit suggest that people's tendency to consider the counterfactuals to certain events and not others affect double-prevention cases. So, if we assume that people have a greater tendency to consider counterfactual alternatives to events with insensitive relationships—in line with some recent work (Icard et al., 2017; Kominsky et al., 2015)—then we have reason to think that a sensitivity-based explanation is appropriate. In this case, the necessity-sufficiency model seems most able to accommodate this view.

8. Conclusion

We argued that the double-prevention effect is a result of counterfactual thinking. Double-prevention scenarios are one of the primary motivations for causal pluralism (Hall, 2004; Lombrozo, 2010). But we argue—in line with some previous suggestions (Lombrozo, 2010; Woodward, 2006)—that counterfactual theories may be able to do all of the explanatory work in these cases. With growing evidence that counterfactual thinking explains other scenarios that motivate causal pluralism—for instance, in the case of omission (Henne et al., 2019) and late-preemption (Henne, Kulesza et al., 2021)—we think that it is time to return to thinking more carefully about unified counterfactual models of causal judgment and how they might explain patterns of causal judgment. We hope future work investigates these issues using different causal measures (Rose et al., 2021), measures of confidence in causal judgment (O'Neill et al., 2022), and prospective scenarios (Byrne, 2016; Henne, O'Neill et al., 2021). We also hope future work will explore whether this account can explain past work on intentionality in double-prevention cases (Lombrozo, 2010).

Acknowledgment

This research was funded by a startup grant from Lake Forest College to Paul Henne and the Reasons and Decisions Lab at Lake Forest College.

Note

1 Participant with ID number R_9vEiEU13exyEHkJ ignored the task and wrote “Nice video, I think this video has alot of meaning, quite an interesting video.” This participant was excluded from the analysis.

Conflict of interest

The authors declare no conflict of interest.

Open Research Badges



This article has earned Open Data, Open Materials and Preregistered Research Design badges. Data, materials and the preregistered design and analysis plan are available at https://osf.io/9fq5/?view_only=cc7ca21838ea407184a0b6d16918e8e7, <https://osf.io/9fq5/>.

References

- Bernstein, S. (2017). Intuitions and the metaphysics of causation. In D. Rose (Ed.), *Experimental Metaphysics*, Vol. 75. Bloomsbury Publishing.
- Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology*, 67(1), 135–157. <https://doi.org/10.1146/annurev-psych-122414-033249>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32.
- Chang, W. (2009). Connecting counterfactual and physical causation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 31(31), <https://escholarship.org/uc/item/61b6s1j1>
- Dinh, P., & Danks, D. (2021). Causal pluralism in philosophy: Empirical challenges and alternative proposals. *Philosophy of Science*, 88(5), 761–772. <https://doi.org/10.1086/714878>
- Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, 28(5), 936–975.
- Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological Science*, 28(12), 1731–1744. <https://doi.org/10.1177/0956797617713053>
- Godfrey-Smith, P. (2009). Causal pluralism. In H. Beebe, P. Menzies, & C. Hitchcock (Eds.), *The Oxford handbook of causation* (pp. 326–337). Oxford University Press.
- Hall, N. (2000). Causation and the price of transitivity. *The Journal of Philosophy*, 97(4), 198–222.
- Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. Paul (Eds.), *Causation and counterfactuals* (pp. 225–276). MIT Press.

- Henne, P., Kulesza, A., Perez, K., & Houcek, A. (2021). Counterfactual thinking and recency effects in causal judgment. *Cognition*, 212, 104708.
- Henne, P., Niemi, L., Pinillos, Á., De Brigard, F., & Knobe, J. (2019). A counterfactual explanation for the action effect in causal judgment. *Cognition*, 190, 157–164. <https://doi.org/10.1016/j.cognition.2019.05.006>
- Henne, P., O'Neill, K., Bello, P., Khemlani, S., & Brigard, F. D. (2021). Norms affect prospective causal judgments. *Cognitive Science*, 45(1), e12931. <https://doi.org/10.1111/cogs.12931>
- Henne, P., Pinillos, Á., & De Brigard, F. (2017). Cause by omission and norm: Not watering plants. *Australasian Journal of Philosophy*, 95(2), 270–283. <https://doi.org/10.1080/00048402.2016.1182567>
- Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *The Philosophical Review*, 116(4), 495–532.
- Hitchcock, C. (2010). The metaphysical bases of liability: Commentary on Michael Moore's causation and responsibility. *Rutgers LJ*, 42, 377.
- Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, 161, 80–93. <https://doi.org/10.1016/j.cognition.2017.01.010>
- 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. <https://doi.org/10.1016/j.cognition.2015.01.013>
- Lewis, D. (1974). Causation. *The Journal of Philosophy*, 70(17), 556–567. <https://doi.org/10.2307/2025310>
- List, C., & Menzies, P. (2017). My brain made me do it: The exclusion argument against free will, and what's wrong with it. In H. Beebe, C. Hitchcock, & H. Price (Eds.), *Making a difference: Essays on the philosophy of causation*. (pp. 269–285). Oxford University Press.
- Livengood, J., & Machery, E. (2007). The folk probably don't think what you think they think: Experiments on causation by absence. *Midwest Studies in Philosophy*, 31, 107–127.
- Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, 61(4), 303–332. <https://doi.org/10.1016/j.cogpsych.2010.05.002>
- Moore, M. S. (2009). *Causation and responsibility: An essay in law, morals, and metaphysics*. Oxford University Press on Demand.
- O'Neill, K., Henne, P., Bello, P., Pearson, J., & De Brigard, F. (2022). Confidence and gradation in causal judgment. *Cognition*, 223, 105036. <https://doi.org/10.1016/j.cognition.2022.105036>
- Paul, L. A. (2010). *Counterfactual theories* In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), (Vol. 1). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199279739.003.0009>
- Paul, L. A., & Hall, N. (2013). *Causation: A user's guide* (1st ed.). Oxford University Press.
- Rose, D., Sievers, E., & Nichols, S. (2021). Cause and burn. *Cognition*, 207(104517). <https://doi.org/10.1016/j.cognition.2020.104517>
- Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *The Quarterly Journal of Experimental Psychology*, 44(2), 321–342.
- Strevens, M. (2013). Causality reunified. *Erkenntnis*, 78(2), 299–320.
- Vasilyeva, N., Blanchard, T., & Lombrozo, T. (2018). Stable causal relationships are better causal relationships. *Cognitive Science*, 42(4), 1265–1296. <https://doi.org/10.1111/cogs.12605>
- Walsh, C. R., & Sloman, S. A. (2011). The meaning of cause and prevent: The role of causal mechanism. *Mind & Language*, 26(1), 21–52. <https://doi.org/10.1111/j.1468-0017.2010.01409.x>
- Woodward, J. (2006). Sensitive and insensitive causation. *The Philosophical Review*, 115(1), 1–50.
- Woodward, J. (2012). Causation: Interactions between philosophical theories and psychological research. *Philosophy of Science*, 79(5), 961–972. <https://doi.org/10.1086/667850>