

Revisiting causal pluralism: Intention, process, and dependency in cases of double prevention

Huseina Thanawala, Christopher D. Erb*

School of Psychology, University of Auckland, 23 Symonds Street, Building 302, Auckland 1010, New Zealand

ARTICLE INFO

Keywords:

Causal pluralism
Causal reasoning
Double prevention
Force theory
Intentionality

ABSTRACT

Causal pluralism proposes that humans can reason about causes and effects in terms of both dependency and process relations, depending on the scenario. Support for this view is provided by responses to double prevention scenarios in which an affector attempts to bring about an outcome, a preventer attempts to prevent the outcome, and a double preventer intervenes to stop the preventer's prevention attempt. Previous research indicates that reasoners award the affector high causal ratings regardless of whether their action was executed intentionally, whereas reasoners only award the double preventer high causal ratings when the double preventer acts intentionally. These results were interpreted as evidence that intentional actions prompt reasoners to differentially weight separately maintained representations of dependency and process relations. The current study presents three challenges to this causal pluralism account by (1) questioning whether intentionality only affects causal ratings by increasing the weighting placed on dependency relations during reasoning (Experiments 1-2), (2) presenting evidence consistent with the possibility that reasoners interpret the double preventer's action in terms of a process relation (Experiments 1-2), and (3) demonstrating that an unintentional double preventer can receive significantly higher causal ratings than an intentional affector when the order in which the characters act is altered (Experiments 3-4). These results underscore the need to revisit fundamental questions regarding how reasoners form, maintain, and reason over representations of causal scenarios featuring intentional actions.

When we look for answers as to why something happened, we look for causes (Wolff, 2014). The ability to learn and reason about causal relations is vital to our ability to behave flexibly and adaptively in the world. Causal relations can range from concrete and deterministic events (e.g., when learning that touching a hot stove top will burn your hand) to abstract and probabilistic events (e.g., when selecting a retirement plan to maximize future savings). Given the importance of causal reasoning for such a wide range of human activities, it is unsurprising that the topic has been explored across many fields, including philosophy (Dinh & Danks, 2021), psychology (Alicke, Mandel, Hilton, Gerstenberg, & Lagnado, 2015; Wolff & Barbey, 2015), economics (Little, 2010), and computer science (Schölkopf, 2022). This widespread interest has resulted in a diverse range of perspectives on how humans reason about cause and effect across different scenarios (Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2021; Lagnado & Channon, 2008; Walsh & Sloman, 2011), age groups (Buchanan & Sobel, 2011; Gopnik et al., 2004), and cultural contexts (Norenzayan & Nisbett, 2000).

Within psychology, two broad theories have been put forward to capture how we represent and reason about causal relations: *dependency*

theories and *process theories* (Chang, 2009). Dependency theories link causes and effects by appealing to logical, modal, or statistical relations. For instance, a dependency theory that appeals to counterfactual reasoning would rule an object or event as a cause of another event if the outcome would not have occurred in the absence of that object or event (Lewis, 1973). A range of dependency theories have been proposed in the literature, including the counterfactual theory of causation (Lewis, 1973), the probabilistic contrast model (Cheng & Novick, 1990), mental model theory (Khemlani, Barbey, & Johnson-Laird, 2014), and causal model theory (Lagnado, Waldmann, Hagmayer, & Sloman, 2007; Sloman, Barbey, & Hotaling, 2009).

In contrast to dependency theories, process theories claim that we interpret and analyze a wide range of causal relations in the same way that we reason about cause and effect in physical systems (Wolff & Barbey, 2015). Such theories share the notion that causal reasoning involves the representation of an exchange or transmission of physical quantities between the cause and the effect (Wolff, Barbey, & Hausknecht, 2010). Prominent process theories include transference theory (Aronson, 1971; Fair, 1979), conserved quantity theory (Dowe, 1992),

* Corresponding author.

E-mail address: christopher.erb@auckland.ac.nz (C.D. Erb).

and force theory (Wolff, 2014).

Dependency and process theories are commonly pitted against each other as competing accounts of causal reasoning. This competitive framing is understandable given that dependency theories are based on the occurrence or non-occurrence of events, without regard to the nature of the processes that produced those events, while process theories are rooted in the exchange or transmission between physical processes. However, despite the clear differences between the two theoretical approaches, it is plausible that both theories reflect psychologically real and distinct ways of thinking about causal relations. That is, humans may think about causal relations in terms of either dependency or process theories depending on a range of factors, including individual characteristics or aspects of the scenario.

Bearing this in mind, Lombrozo (2010) proposed a causal pluralism account which claims that humans can engage in two distinct modes of explanations that are characterized as involving either teleological or mechanistic modes of thinking. Teleological explanations cite functions or goals, while mechanistic explanations cite causal mechanisms. Generally, intentional, goal-directed actions are understood in terms of teleological explanations. Goal-directed behavior does not place emphasis on the precise mechanisms through which the end state is realized. Hence, there is equifinality between the outcomes where the same outcome can be achieved despite variation in the means (Lombrozo, 2010). This suggests that goal-directed behavior is often understood in mechanism-independent terms and encourages a dependency perspective of causation, as there is no requirement that a particular method of transmission is involved (for a discussion, see Lombrozo & Wilkenfeld, 2019).

By contrast, Lombrozo (2010) proposes that unintentional (accidental) actions are more likely to be understood in terms of mechanistic explanations, which require a transmission or exchange of force to ascribe causality. This implies that unintentional actions support a process perspective on causation. Lombrozo (2010) suggests that both dependence and process perspectives can influence causal reasoning, but these perspectives can be leveraged differently depending on whether the reasoners adopt a teleological or mechanistic mode of explanation. Thus, the central claim of this specific version of causal pluralism proposes that dependence relations may be weighed more heavily in evaluating causal relations from a teleological mode while process relations may be weighed more heavily in evaluating causal relations from a mechanistic mode.

Lombrozo (2010) set out to test this early version of the causal pluralism account by presenting participants with double prevention scenarios featuring three characters: an affector who performs an action that will result in a particular effect, a preventer who attempts to prevent the effect from occurring, and a double preventer who disrupts the preventer's prevention attempt. Crucially, the intentional status of the affector and double preventer was manipulated across different versions of the scenarios. An example of one version of the scenario where both the affector (Alice) and double preventer (Carol) acted intentionally is stated below (Lombrozo, 2010, p. 311):

'Alice, Bob, and Carol have spent the afternoon juggling and listening to music. At the moment, Alice is juggling, and the music is not playing. Alice wants to listen to music, so she deliberately throws a juggling ball, which heads straight for the stereo's 'on' button. But while Alice's ball is in the air, Bob starts pulling on the power cord connecting the stereo to the outlet. If Bob unplugs the cord, it will prevent Alice's ball from turning on the stereo and starting the music. However, Carol wants the music to play, so she deliberately steps on the power cord just before Alice's ball hits the 'on' button, preventing Bob's pull from unplugging the stereo. As a result of these events, the music starts to play.'

In the other versions of this scenario, the Alice and Carol characters were both described as acting unintentionally (e.g., Alice accidentally threw the juggling ball, Carol accidentally stepped on the power cord),

or one of the characters was described as acting intentionally while the other character was described as acting accidentally (e.g., Alice accidentally threw the juggling ball, Carol deliberately stepped on the power cord). Reasoners' judgments of the double prevention scenarios supported the predictions of the version of the causal pluralism account proposed and tested by Lombrozo (2010). Consistent with the notion that intentional actions encouraged reasoners to adopt a teleological mode of explanation and, correspondingly, weigh dependence relations more heavily, participants provided similar causal ratings to the affector and the double preventer when they both behaved intentionally (the A+C+ condition), despite the fact that one character (the double preventer, Carol) did not share a process relation with the effect (Lombrozo, 2010).

In another condition in which the actions of the affector and double preventer were accidental (e.g., Alice accidentally threw the ball and Carol accidentally stepped on the power cord; the A-C- condition), reasoners provided higher causal ratings for the affector than the double preventer. According to causal pluralism, this difference emerged because reasoners were more likely to adopt a mechanistic mode of explanation in the absence of intentional actions by the affector and double preventer, resulting in process relations being weighed more heavily during reasoning. On the assumption that the affector alone shared a process relation with the outcome, this adoption of the mechanistic mode would result in higher causal ratings being awarded to the affector relative to the double preventer.

The current study explores three challenges to the version of the causal pluralism account of reasoning in double prevention scenarios developed and tested by Lombrozo (2010). However, before turning to these challenges, it is important to emphasize that the causal pluralism account tested by Lombrozo (2010) was a first elaboration of the approach. Lombrozo (2010) readily acknowledged that the account was likely too simple and ultimately endorsed a view called *exportable dependence theory*, which suggests that intentional actions and actions that are linked to an outcome via a direct transference of energy receive higher causal ratings because they are more exportable "in the sense that they are likely to subserve future prediction and intervention" (pg. 327). Although Lombrozo (2010) discusses the potential links between exportable dependence theory and causal pluralism, these links have yet to be formalized and tested directly. Thus, the results of the current study could help to constrain and motivate future research and theorizing on causal pluralism and exportable dependence theory.

The first challenge raised in the current study concerns the various ways in which the intentional status of the characters featured in the scenarios could hypothetically impact reasoning. As we discuss below, this challenge proposes that intentionality might influence causal ratings via a more direct route than the mechanism proposed by the causal pluralism account involving the differential weighting of separately maintained process and dependence representations linking the characters to the outcome. The second challenge concerns whether it is indeed the case that reasoners fail to represent a process relation linking the double preventer and the outcome. To motivate this second challenge, we review recent work on *force theory*, a variant of process theory developed by Wolff and Barbey (2015). As we discuss below, this work lends credence to the possibility that many reasoners do think about the link between a double preventer and an outcome in terms of a process relation even though the double preventer is not connected to the outcome via a direct transference of energy. However, it is unclear at present the extent to which the intentional status of the double preventer might impact the representation of such links. Finally, our third challenge focuses on how altering the sequence in which each of the characters acts in double prevention scenarios impacts causal ratings. This challenge is also motivated by process theories, which emphasize the importance of going beyond the occurrence or non-occurrence of events to consider how events unfold over time, though it is important to highlight that theories within the dependency tradition have also highlighted the importance of sequence effects on causal reasoning (e.g.,

Henne, Kulesza, Perez, & Houcek, 2021; Spellman, 1997).

1. Challenges to causal pluralism

The version of the causal pluralism account developed and tested by Lombrozo (2010) proposes that intentionality influences causal ascriptions in double prevention scenarios by differentially weighting separately maintained process and dependence relations. However, the results of Lombrozo (2010, Experiment 1a) outlined above could be explained by an alternative – and arguably, more parsimonious – account that does not appeal to separately maintained and differentially weighted process and dependence relations. For example, actions that are intentional or that result in a direct transference of energy might receive a boost in causal ratings because reasoners interpret such actions to generate outcomes more reliably.

To address the possibility that the intentionality of action itself increased causal ratings (as opposed to increasing causal ratings via the differential weighting of process and dependence relations), Lombrozo (2010) conducted an experiment using deviant causal chains in which the affector and double preventer were both described as intending to bring about a particular outcome, but the specific actions taken by the characters that led to the outcome were accidental (e.g., Alice wanted to throw a juggling ball at the stereo to turn it on but the thought of doing so made her nervous, resulting in her accidentally throwing a ball that happened to turn on the stereo). Lombrozo reasoned that if intentionality per se influenced causal ratings, then a double preventer who is described as intending to bring about an effect but does so accidentally should nevertheless receive high causal ratings. However, Lombrozo found that the affector received higher causal ratings than the double preventer in scenarios featuring deviant causal links. This finding suggests that describing the double preventer as having the intention to bring about the effect was not sufficient to account for the effects observed in double prevention scenarios featuring non-deviant links.

Although results from the deviant causation conditions lend some support to the notion that intentionality influences the relative weighting of specific causal relations during reasoning, two considerations merit discussion. First, the difference in ratings provided to the affector and double preventer across the deviant and normal double prevention scenarios only reached significance for one of the two conditions tested (Lombrozo, 2010, Experiment 1b), suggesting that intentionality per se may impact causal ratings under some conditions. Second, it is unclear that the deviant causation conditions robustly tested for the effects of interest because the actions in question that linked the characters to the outcome were not themselves intentional, despite the characters having a general intention to bring about the outcome. Consequently, it is unclear that the general intention to bring about the outcome would impact the representation of the specific relation(s) linking the character and the outcome. Thus, a primary aim of the current study is to further investigate how intentionality impacts the causal ratings awarded to the characters featured in double prevention scenarios.

A second challenge to the causal pluralism account of reasoning in double prevention scenarios comes from recent research on force theory (Wolff & Barbey, 2015). According to force theory, individual interactions involve two main entities, an affector and a patient. The forces may be physical, psychological, or social. The theory states that causal relations are specified in terms of configurations of forces that are evaluated with respect to an end state. Force theory particularly emphasizes the role of allow and prevent relations. Allow and prevent relations are important within force theory for identifying how individual relations can be combined to form causal chains. Such causal chains are then used to obtain new overarching causal relations. On this view, reasoners can conceptualize the double preventer's link to the outcome in terms of process relations by interpreting the double preventer's action as removing or preventing a force (Wolff & Barbey, 2015).

Wolff and Barbey (2015) used different scenarios involving cars as

examples throughout their study to investigate how force theory can capture causal reasoning across a range of conditions. For example, if A causes B and B causes C, people may derive the overarching causal relation that A causes C. Similarly, if A prevents B and B prevents C, people may derive the overarching causal relation that A causes or allows C. Wolff and Barbey conducted an experiment wherein participants were shown various animations involving car A, car B, and car C. After watching each animation, participants chose the expression that best described the relation between the first and last cars in the chain (e.g., A and C). The results showed that in a double prevention scenario where A prevents B and B prevents C, 50% of participants answered 'A caused C' and 50% of participants answered 'A allowed C.' Force theory predicts that double prevention scenarios are often ambiguous between *cause* and *allow* relations, leading to different proportions of *cause* and *allow* judgments depending on the description of the scenario. For instance, some scenarios are biased towards a cause relation (e.g., a ceiling collapsing after a supporting post is knocked down), whereas other scenarios are biased towards an allow relation (e.g., water flowing through a drain after a plug is removed; for a discussion, see the "Accounting for Multiple Conclusions" section of Wolff & Barbey, 2015).

The results of Wolff and Barbey (2015) suggest that participants in Lombrozo's (2010) study may have interpreted the double preventer to have shared a process relation with the outcome even though the double preventer is not connected to the outcome via a direct transference of energy. Further, it is possible that many of the participants in Lombrozo's study believed that the double preventer *allowed* – as opposed to *caused* – the outcome to occur. Given that the questions used in Lombrozo's (2010) study asked participants to judge how appropriate it was to claim that each of the characters *caused* the outcome, it is possible that the wording of the questions led to an underestimation of the double preventer's role in bringing about the outcome. Consequently, the current study (a) asked participants to judge how appropriate it was to claim that the outcome occurred *because of* each of the characters' actions and (b) investigated the frequency with which reasoners judged the affector and double preventer as having *caused*, *allowed*, or *neither caused nor allowed* the outcome to occur.

Finally, a third challenge to the causal pluralism account of reasoning about double prevention scenarios concerns the extent to which the causal ratings provided to the affector, preventer, and double preventer are impacted by the order with which each character acts. For example, reasoners might judge that a double preventer's action contributed more to the occurrence of an outcome if their action followed rather than preceded the action of the preventer. In the context of the scenario described above, this can be illustrated by contrasting (a) a case in which the preventer begins to pull on the power cord before the double preventer steps on the power cord with (b) a case in which the double preventer steps on the stereo's power cord before the preventer attempted to unplug the stereo. From the standpoint of dependency theories that emphasize logical or counterfactual relations, the occurrence or non-occurrence of the outcome arguably depends on the double preventer's action equally in both cases. However, from the perspective of process theory, the degree of threat posed by the preventer will be impacted by the order with which events unfolded. If, for example, the double preventer's ratings were impacted by altering the order with which each of the characters acted, this would lend credence to the idea that reasoners are thinking about the double preventer in terms of process relations.

It is important to note at this stage that many of the claims made above regarding the predictions of dependency and process theories are open to considerable debate. For instance, although process theories such as force theory have emphasized the importance of allow relations (Wolff, 2014; Wolff & Barbey, 2015), dependency theories have also been developed to account for similar distinctions among cause and enable relations (Beller, Bennett, & Gerstenberg, 2020; Sloman et al., 2009). Similarly, although process theories generally place more emphasis on how events unfold over time, dependence theories can also

emphasize the importance of temporal sequence when considering factors such as counterfactual distance (e.g., if the double preventer is stepping on the power cord before the preventer begins to pull the cord, the counterfactual distance to get to a possible world in which the music doesn't play might be greater than it would have been if the double preventer only stepped on the cord after the preventer began to pull on it). For the purposes of the current study, these considerations are of secondary interest to the primary aim of testing the causal pluralism account of reasoning in double prevention scenarios developed by Lombrozo (2010). That is, we believe that the substance of the challenges posed to the causal pluralism account will remain regardless of whether one ultimately adopts a process perspective, a dependency perspective, or questions the dichotomy altogether.

2. The current study

The preceding discussion raised four central questions. (1) *Can the intentional status of the affector and double preventer impact causal ratings via a route other than the route posited by the causal pluralism account tested by Lombrozo (2010) involving the differential weighting of process and dependence relations?* If intentionality were to impact causal ratings through some process other than the weighting mechanism proposed and tested by Lombrozo (2010) (e.g., if intentionality boosts causal ratings directly), then one might expect to observe significantly higher causal ratings in conditions in which the affector behaves intentionally (A+ conditions) relative to conditions in which the affector behaves unintentionally (A- conditions). Similarly, such an effect of intentionality could explain why Lombrozo (2010, Experiment 1a) observed descriptively higher causal ratings for the double preventer than the affector in the A-C+ condition in which the double preventer, but not the affector, behaved intentionally. On this view, intentionality might provide a boost to the double preventer's causal ratings that extends beyond the effect of being linked to the outcome via a direct transfer of force enjoyed by the affector. Although observing such effects would not constitute definitive counterevidence to the causal pluralism account, doing so would lend credence to a more parsimonious account of the effect on intentionality that need not posit the differential weighting of separately maintained representations of process and dependence links between the characters and the outcome.

(2) *To what extent do reasoners classify the affector and double preventer as having caused, allowed, or neither having caused nor allowed an outcome?* Relatedly, (3) *to what extent are these classifications impacted by the intentional status of the affector and double preventer?* In light of Wolff and Barbey (2015) results, we might expect that a majority of reasoners would classify the affector as having caused the outcome and the double preventer as having either allowed or caused the outcome. Such a demonstration would lend credence to the notion that reasoners can think about the double preventer in terms of a process relation involving a causal chain, thereby calling into question a central assumption of the causal pluralism account proposed and tested by Lombrozo (2010). However, it is unclear at present how such classifications might be impacted by the intentional status of the affector and double preventer.

Finally, (4) *to what extent are the ratings awarded to the affector, preventer, and double preventer impacted by the order with which they act?* Demonstrating that the temporal sequence of actions described in double prevention scenarios significantly impacts the ratings awarded to the affector, preventer, or double preventer would lend further credence to process accounts of double prevention, including the force theory account. However, if altering the sequence of events described in double prevention scenarios were to result in higher causal ratings being awarded to an unintentional double preventer than an intentional affector, this would pose a particularly strong challenge to the version of the causal pluralism account of double prevention tested by Lombrozo (2010), as the account predicts that an unintentional double preventer should receive lower ratings than an intentional affector.

In the following, we present four experiments designed to address the

questions outlined above. Each of the experiments were pre-registered before data collection commenced, and each experiment was conducted online. The data and analysis files for each experiment are available at https://osf.io/nreqf/?view_only=815a37ead733448d983dd55152032300.

3. Experiment 1

Participants in the current experiment were presented with the double prevention scenario developed by Lombrozo (2010) introduced above in which three friends were juggling and listening to music. Participants were randomly assigned to one of four conditions: one in which both the affector and double preventer behaved intentionally (A+C+), one in which both behaved unintentionally, (A-C-), one in which only the affector behaved intentionally (A+C-), or one in which only the double preventer behaved intentionally (A-C+). We started the questionnaire by asking participants three forced-choice questions regarding whether the affector, preventer, and double preventer caused, allowed, or did not cause or allow the outcome to occur.

In Lombrozo's (2010) study, participants were asked to evaluate causal claims (e.g., "Alice caused the music to play.") on a 6-point scale of appropriateness. We also asked participants to evaluate causal claims on a 6-point scale of appropriateness. However, given the consequential distinction between cause and allow relations highlighted by force theory (Wolff & Barbey, 2015), we avoided the word "cause" in our appropriateness rating statements. Instead, we asked participants to evaluate statements featuring the word "because" (e.g., "The music played because of Alice."). Like Lombrozo's (2010) study, participants were also asked true-or-false comprehension questions to ensure that they had read and understood the double prevention scenarios presented to them.

Before Experiment 1, we conducted a pilot experiment in which participants were recruited online through Amazon Mechanical Turk (see Supplementary Materials). A substantial proportion of participants tested in the pilot study failed to meet our inclusion criteria and were subsequently excluded from analysis. We therefore opted to use Prolific (www.prolific.co) instead of Amazon Mechanical Turk to recruit participants in our subsequent experiments. In light of the results of the pilot experiment, we pre-registered four central predictions for Experiment 1: (1) Participants will most frequently select the cause relation to describe Alice's contribution to the outcome, the neither cause nor allow relation to describe Bob's contribution to the outcome, and the allow relation to describe Carol's contribution to the outcome. (2) Participants will be more likely to classify Alice's contribution to the outcome as causing and less likely to classify the Alice's contribution to the outcome as not causing or allowing when Alice behaved intentionally. Similarly, participants will be more likely to classify Carol's action as allowing the outcome and less likely to classify the Carol's action as not causing or allowing the outcome when Carol behaved intentionally. (3) The appropriateness ratings for Alice will be significantly higher when Alice acts intentionally (in the two A+ conditions) relative to when she acts unintentionally (in the two A- conditions). (4) The appropriateness ratings will be significantly higher for Carol than the Alice in the A-C+ condition.

3.1. Method

3.1.1. Participants

This experiment was pre-registered before data collection began. The pre-registration on Open Science Framework can be accessed at <https://osf.io/87dbh>. We used the software program G*Power to conduct power analyses for this experiment. We sought to have (a) .80 power to detect an effect size of .25 with a one-tailed independent means t-test at the standard .05 alpha error probability and (b) .80 power to detect an effect size of .25 with a one-tailed paired-samples t-test at the standard .05 alpha error probability. Our power analyses indicated that

a sample size of approximately 100 participants in each condition would be needed to reach these goals. We therefore sought to have a final sample size of 400 participants. We only included the first 100 participants from each condition who met the criteria, and any additional participants were excluded from the study even if they met our other inclusion criteria.

We initially recruited 440 participants (age range 18–65) from the Prolific online participant recruitment platform. Eighteen participants were excluded from Prolific in this first data collection wave because the participants provided the incorrect survey code or left the code field empty. We excluded participants who had not answered one or more of the questions. Additionally, participants who answered more than one of the comprehension check questions incorrectly were also excluded from subsequent analysis. After excluding participants, we did not have a sufficient number of includable participants for Condition 2 (A-C-) or Condition 4 (A-C+). We therefore recruited an additional 20 participants across those two conditions. After participants were excluded according to the aforementioned criteria, we had a final sample with data from 100 participants in each of the four experimental conditions. Participants received GBP 1.34 in compensation for their time.

3.1.2. Materials

The experiment consisted of an online survey that presented participants with the double prevention scenario from Lombrozo (2010) involving the characters Alice, Bob, and Carol. Participants in the current study were randomly allocated to one of four versions of the scenario (A+C+, A+C-, A-C+, and A-C-). Each of these scenarios is provided in the Supplementary Materials. After reading the scenario, participants were asked to respond to questions like the following for each character (example for Alice shown):

Which of the following statements do you agree with the most?

- 1) Alice caused the music to play.
- 2) Alice allowed the music to play.
- 3) Alice did not cause or allow the music to play.

Following these questions, participants had to evaluate the appropriateness of claims regarding Alice, Bob, and Carol on a 6-point scale (from “Completely Inappropriate” to “Completely Appropriate”). As noted above, the statements presented for the appropriateness ratings differed from Lombrozo’s (2010) statements, which featured the word “cause” (e.g., “Alice caused the music to play.”). Participants judged the appropriateness of a series of sentences stating that the outcome occurred “because of” each character in the scenario (e.g., “The music played because of Alice.”). Lastly, we ended the questionnaire by asking participants five true-or-false comprehension questions about the scenario. All the materials used in this experiment and the subsequent experiments are provided in the Supplementary Materials.

3.1.3. Procedure

The experiment consisted of a questionnaire created on an online platform called Qualtrics (www.qualtrics.com). Participants were recruited through an online platform called Prolific (www.prolific.co). Participants were presented with a description of the study and consented to participate by clicking a link from Prolific that redirected them to the Qualtrics questionnaire. Participants were randomly assigned to one of the four conditions described above. Additionally, the order of the forced-choice questions, the order of the statements about each character for the appropriateness ratings, and the order of the comprehension questions were randomized for each participant. The participants could read the scenario at the top of the page when completing the force-choice questions and the questions regarding the appropriateness ratings for each character. For the last section featuring the comprehension questions, participants could not read the scenario on their screen when answering the questions.

3.2. Results

We pre-registered four central predictions for this experiment. Our first hypothesis stated that participants will most frequently select the cause relation to describe Alice’s contribution to the outcome, the neither cause nor allow relation to describe the Bob character’s contribution to the outcome, and the allow relation to describe Carol’s contribution to the outcome. Responses to the agreement questions were analyzed with a series of chi-square tests, following Wolff and Barbey (2015). A one-sample chi-square test was performed to examine the causal ratings for each character across all four conditions (see Fig. 1). The results for Alice yielded a significant effect, $\chi^2(2, N = 400) = 382.66, p < .001$, with the cause relation selected significantly more often than the allow relation or the neither cause or allow relation. The results for Bob yielded a significant effect, $\chi^2(2, N = 400) = 679.42, p < .001$, with the neither relation selected significantly more often than the cause relation or the allow relation. The results for Carol yielded a significant effect, $\chi^2(2, N = 400) = 149.43, p < .001$, with the allow relation

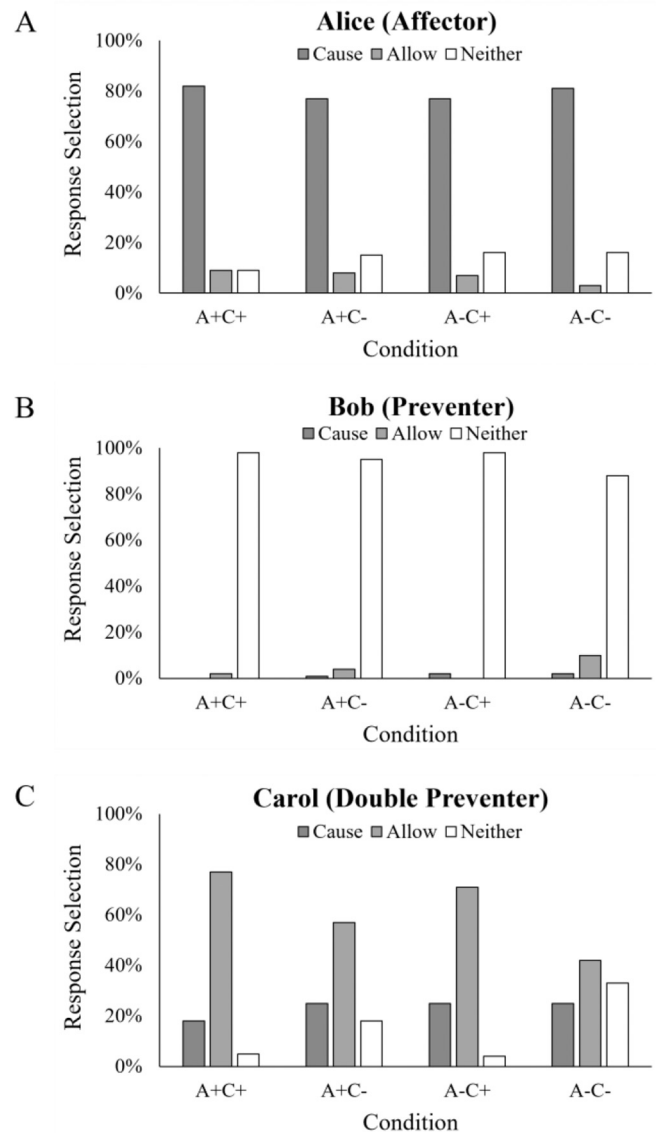


Fig. 1. Percent of Cause, Allow, and Neither Responses for Alice (Affector), Bob (Preventer), and Carol (Double Preventer) for Each Condition. Note. Percentage of responses when both were intentional (A+C+), only Alice was intentional (A+C-), only Carol intentional (A-C+), or both were unintentional (A-C-).

selected significantly more often than the cause or the neither relation.

Our second hypothesis stated that participants would be more likely to classify Alice’s action as causing the outcome and less likely to classify Alice’s action as not causing or allowing the outcome when Alice behaved intentionally (as opposed to unintentionally). We evaluated how the intentional status of the affector (i.e., Alice) impacted responses to the agreement question for the affector. A chi-square test of independence was performed to examine the relation between the intentionality of Alice and the proportion of each response type (see Fig. 2a). The relation between these variables was not significant, $\chi^2(2, N = 400) = 2.96, p = .23$. As illustrated in Fig. 2A, there was no significant difference in the frequency of cause, allow, or neither responses when Alice behaved intentionally versus unintentionally.

Our second hypothesis also stated that participants would be more likely to classify Carol’s contribution to the outcome as "allowing" the outcome and less likely to classify Carol’s contribution to the outcome as "not causing or allowing" when Carol behaved intentionally (as opposed to unintentionally). We evaluated how the intentional status of double preventer (i.e., Carol) impacted responses to the agreement question for the double preventer. A chi-square test of independence was performed to examine the relation between the intentionality of Carol and the proportion of each response type (see Fig. 2B). The effect of intentional status had a significant effect on the distribution of responses, $\chi^2(2, N = 400) = 39.65, p < .001$. The results showed that the allow relation was selected for Carol significantly more frequently when Carol behaved intentionally as compared to unintentionally and, correspondingly, the neither cause or allow relation was selected for Carol significantly less frequently when Carol behaved intentionally as compared to unintentionally.

Our next hypothesis stated that the appropriateness ratings for Alice will be significantly higher when Alice acts intentionally (in the two A+ conditions) relative to when Alice acts unintentionally (in the two A-

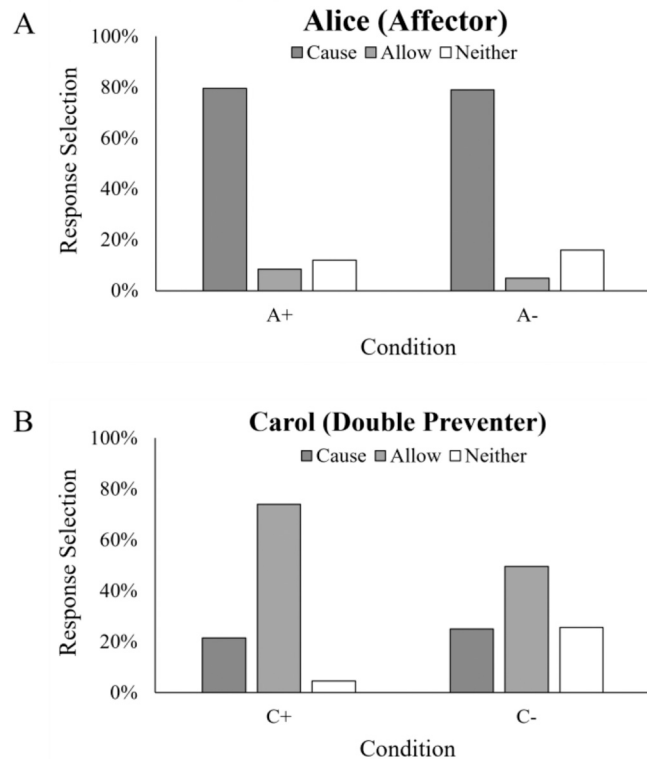


Fig. 2. Percent of Cause, Allow, and Neither Responses for Alice (Affector) and Carol (Double Preventer) as a Function of Intentional Status. Note. Percent of responses: when Carol and Alice are intentional (A+ and C+) and when Alice and Carol are unintentional (A- and C-).

conditions). The appropriateness judgments were analyzed using an independent samples t-test for Alice to see if there is a difference in the mean ratings between the conditions where Alice behaved intentionally (A+C+, A+C-) and unintentionally (A-C+, A-C-). A significant difference was observed between the mean appropriateness ratings for Alice based on intentional status, $t(398) = -2.27, p = .012$, with higher appropriateness ratings when Alice behaved intentionally ($M = 4.84, SD = 1.09$) than when Alice behaved unintentionally ($M = 4.57, SD = 1.23$). The mean appropriateness ratings for the affector (i.e., Alice) and the double preventer (i.e., Carol) are presented in Fig. 3 for each of the four conditions.

Our last hypothesis stated that the appropriateness ratings would be significantly higher for Carol than Alice in the A-C+ condition. The appropriateness judgments were compared for Alice and Carol using a paired samples t-test in the A-C+ condition only. There was not a significant difference in the appropriateness ratings for Alice ($M = 4.56, SD = 1.21$) and Carol ($M = 4.74, SD = 1.03$) conditions; $t(99) = -1.17, p = .12$.

3.2.1. Exploratory analyses

During the review process, we identified another approach to test whether appropriateness ratings reflected factors other than the differential weighting of separately maintained process and dependence relations. If appropriateness ratings reflected “boosts” from specific factors such as whether an action was intentional or was linked to the outcome via the direct transference of force, then it would stand to reason that the affector would receive higher causal ratings than the double preventer in the A+C+ condition, as the affector’s ratings would be receiving boosts for their action being both intentional and linked to the outcome via the direct transference of force, whereas the double preventer’s action would only receive a boost for their action being intentional. We therefore decided to test whether the affector received significantly higher causal ratings than the double preventer in the dataset featured in the current experiment and the dataset featured in our pilot experiment. The results of Experiment 1 revealed a significant effect of character, $t(99) = 2.19, p = .031$, with higher ratings for the affector ($M = 4.85, SD = 1.04$) than the double preventer ($M = 4.60, SD = 0.98$). Similarly, our pilot experiment also revealed a significant effect of character, $t(103) = 2.01, p = .047$, with higher ratings for the affector ($M = 5.12, SD = 1.05$) than the double preventer ($M = 4.83, SD = 0.88$).

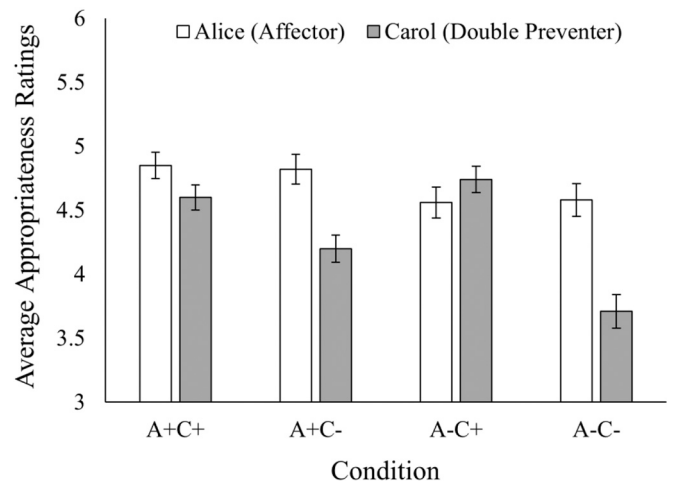


Fig. 3. Average “Because of” Appropriateness Ratings for Alice and Carol as a Function of Intentional Status.

Note. Appropriateness ratings are from a scale of 1 to 6. The conditions included scenarios in which both Alice and Carol behaved intentionally (A+C+), only Alice behaved intentionally (A+C-), only Carol behaved intentionally (A-C+), and both behaved unintentionally (A-C-). Error bars denote standard errors.

3.3. Discussion

Responses to our agreement questions were consistent with the results of Wolff and Barbey (2015), with the cause relation selected most often for the affector, the neither cause nor allow relation selected most often for preventer, and the allow relation selected most often for the double preventer. These results are consistent with the possibility that a subset of reasoners interpreted the double preventer's link to an outcome in terms of process relations despite the lack of a direct transmission of force between the double preventer and the outcome. This finding presents an important challenge to the causal pluralism account because it raises questions regarding (a) whether participants fail to adopt a mechanistic mode of explanation when reasoning about the double preventer and, more generally, (b) whether participants represent distinct process and dependence relations linking the characters to the outcome, as opposed to composite representations that reflect both process and dependence information. It is important to note, however, that these results could also be accommodated by dependency theories designed to account for allow or enable relations (Beller et al., 2020; Sloman et al., 2009).

As predicted, we found that intentionality had a significant effect on responses to the agreement questions pertaining to the double preventer's contribution to the outcome. Although most participants classified the double preventer as allowing the outcome, the degree to which this relation was attributed to the double preventer depended on the intentionality of their action. Participants were significantly more likely to classify the double preventer's contribution to the outcome as allowing the outcome and significantly less likely to classify the double preventer's contribution to the outcome as not causing or allowing the outcome when the double preventer behaved intentionally (as opposed to unintentionally). This provides an early demonstration that reasoning about allow relations is impacted by the double preventer's intentional status.

In contrast to our prediction and the results of our pilot study (see Supplementary Materials), we did not observe evidence that the affector's intentionality impacted the classification of their action, as participants overwhelmingly classified the affector as causing the outcome regardless of their intentional status. We did, however, observe significantly higher appropriateness ratings for the affector when the affector behaved intentionally (in the two A+ conditions) versus unintentionally (in the two A- conditions). On our view, the most parsimonious interpretation of this effect is that intentionality led to a boost in appropriateness ratings for the affector's action beyond the effect of being linked to the outcome via a direct transfer of force. This interpretation was supported by exploratory analyses comparing the appropriateness ratings awarded to the affector and the double preventer in the A+C+ condition of the current experiment and our pilot experiment. These analyses showed that the affector received significantly higher appropriateness ratings than the double preventer, suggesting that the affector's ratings reflected both an effect of intentionality and an effect of being linked to the outcome via a direct transfer of force, whereas the double preventer's ratings reflected an effect of intentionality alone.

From the perspective of the version of causal pluralism proposed and tested by Lombrozo (2010), it is unclear why the differential weighting of the process and dependence links would result in the affector receiving significantly higher appropriateness ratings in the A+ conditions relative to the A- conditions. Indeed, upon failing to observe such an effect, Lombrozo (2010, Experiment 1a) performed additional analyses to demonstrate that the lack of an effect was not due to a ceiling effect in which causal ratings for the affector were too high to show an effect of intentionality. It is also unclear how the causal pluralism account would explain the affector receiving significantly higher appropriateness ratings in the A+C+ condition without appealing to considerations beyond the differential weighting of separately maintained process and dependence relations.

Finally, we failed to observe a significant difference in the

appropriateness ratings for the affector and the double preventer in the A-C+ condition, though the appropriateness ratings were descriptively higher for double preventer than the affector. Given this descriptive difference, along with the descriptive difference reported by Lombrozo (2010, Experiment 1a) and the significant difference observed in our pilot experiment, we decided to conduct a second experiment with a larger sample size that focused on the A-C+ condition.

4. Experiment 2

In Experiment 1, we hypothesized that appropriateness ratings would be significantly higher for the double preventer (Carol) than the affector (Alice) in the A-C+ condition. However, a significant difference between the appropriateness ratings given to the double preventer and affector was not observed in the A-C+ condition. We therefore used G*Power to determine the sample size necessary to obtain .80 power to detect an effect size of .12 with a one-tailed paired-samples t-test at the standard alpha level of .05. This power analysis revealed that a sample size of approximately 430 participants was required. We selected a final target sample size of 450 participants.

4.1. Method

4.1.1. Participants

This experiment was pre-registered before data collection commenced. The pre-registration on Open Science Framework can be accessed at <https://osf.io/78sfj>. Five hundred participants (age range 18–65) were recruited from an online platform (Prolific) and participated in exchange for monetary compensation (GBP 0.75) for a 5-minute study. Based on the average participant time for completion for Experiment 1, we reduced the study time to 5 minutes. Fifteen participants were excluded from Prolific initially because they provided the incorrect survey code or left the code field empty, resulting in a total of 485 participants available for analyses. Participants were excluded from the study if they (a) had not answered one or more of the questions or (b) answered more than one of the comprehension-check questions incorrectly. This resulted in 451 participants who met our inclusion criteria. As stipulated in our pre-registration, our final sample was restricted to the first 450 participants who met the aforementioned inclusion criteria.

4.1.2. Materials

The experiment consisted of an online survey in which participants were presented with the A-C+ condition used in Experiment 1.

4.1.3. Procedure

The procedure was identical to Experiment 1 with the exception that participants were only presented with the A-C+ condition of the double prevention scenario wherein the affector behaved unintentionally and the double preventer behaved intentionally. The order of the force-choice questions, the order of the statements about each character for the appropriateness ratings, and the order of the comprehension questions were randomized for each participant.

4.2. Results

We first analyzed responses to the agreement questions with a series of chi-square tests, following Wolff and Barbey (2015). This analysis was of secondary interest and was not pre-registered but is provided to be consistent with the results of Experiment 1. A one-sample chi-square test was performed to examine the causal ratings for each character in the A-C+ condition. The results for Alice yielded a significant effect, $\chi^2(2, N = 450) = 393.17, p < .001$, with the cause relation selected significantly more often than the allow or the neither relation (see Fig. 4). The results for Bob yielded a significant effect, $\chi^2(2, N = 450) = 756.81, p < .001$, with the neither relation selected significantly more often than the cause or the allow relation. The results for Carol yielded a significant effect, χ^2

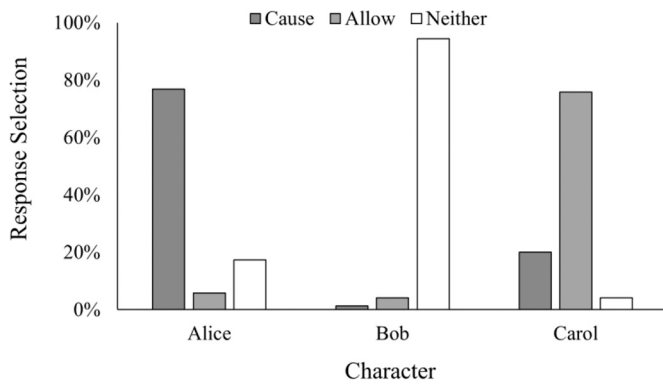


Fig. 4. Percent of Cause, Allow, and Neither Responses for Alice (Affector), Bob (Preventer), and Carol (Double Preventer) for the A-C+ Condition. *Note.* Percent of responses: when Alice was unintentional and Carol was intentional (A-C+).

(2, $N = 450$) = 381.61, $p < .001$, with the allow relation selected significantly more often than the cause or the neither relation.

Next, appropriateness judgments for the Alice and Carol characters were analyzed with a one-tailed paired-samples t-test. A significant effect was observed, $t(449) = -3.16, p < .001$, with higher appropriateness ratings for Carol ($M = 4.90, SD = 0.99$) than Alice ($M = 4.68, SD = 1.19$) (see Fig. 5).

4.3. Discussion

Experiment 2 revealed significantly lower appropriateness ratings for the affector than the double preventer in the A-C+ condition, consistent with the descriptive difference reported by Lombrozo (2010, Experiment 1a), the significant difference observed in our pilot experiment, and the descriptive difference reported in Experiment 1. This finding presents a challenge for the causal pluralism account as it lends credence to the claim that intentionality can directly influence causal ratings in double prevention scenarios without appealing to a weighting mechanism. On this view, the intentionality of the double preventer’s action led to increased causal ratings that extended beyond the boost in causal ratings provided to the affector for being linked to the outcome via the direct transference of energy.

From the perspective of the causal pluralism account proposed and tested by Lombrozo (2010), it is unclear why the affector, who is purported to share both a process and dependence link to the outcome (with the former heavily weighted in the A-C+ condition), would receive

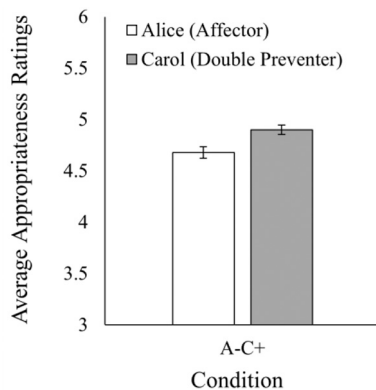


Fig. 5. Average “Because of” Appropriateness Ratings for Alice and Carol as a Function of Intentional Status. *Note.* Appropriateness ratings are from a scale of 1 to 6 for the A-C+ condition where Alice behaved unintentionally and Carol behaved intentionally. Error bars denote standard errors.

lower appropriateness ratings than the double preventer, who is purported to share only a dependence link to the outcome. Although the results of Experiment 2 do not constitute definitive counterevidence to the causal pluralism account, we would argue that our account of the results is more parsimonious, as it does not posit the differential weighting of separately maintained representations of process and dependence links between the characters and the outcome.

5. Experiment 3

Experiments 1 and 2 explored two challenges to the causal pluralism account of reasoning in double prevention scenarios by (1) questioning whether intentionality only affects causal ratings by increasing the weighting placed on dependency relations during reasoning and (2) lending credence to the claim that reasoners might interpret the double preventer’s action in terms of process. Experiments 3 and 4 explore a third challenge to causal pluralism by (3) investigating how the temporal sequence of the characters’ actions in double prevention scenarios impacts the causal ratings awarded to the characters. To explore this challenge, we developed an original double prevention scenario that allowed for the order with which each of the characters acted to be altered.

We propose that even an unintentional double preventer can receive high causal ratings depending on when the affector and preventer act. For example, imagine a scenario in which the preventer blocks the path linking the affector to an outcome, the affector attempts to bring about the outcome, and the double preventer removes the blockade put in place by the preventer. We will refer to this as a PA2 scenario because the preventer (P) acts first, the affector (A) acts second, and the double preventer (2) acts third (see Fig. 6A). In such scenarios, we predict that the affector is likely to receive lower ratings than the double preventer because the affector’s attempt is never represented as being capable of bringing about the outcome until the double preventer intervenes.

By contrast, imagine a scenario in which the preventer and double preventer act before the affector (P2A scenario) (see Fig. 6B). In this case, we predict that the affector will receive higher ratings than the double preventer because the affector’s action is not represented to be threatened when it is performed. Similarly, in scenarios in which the affector behaves first, the preventer behaves second, and the double preventer behaves third (AP2) (see Fig. 6C), the affector’s action is represented as being capable of bringing about the outcome, at least briefly. Given that the AP2 scenario used in Experiments 1 and 2 showed that the affector received high causal ratings, we expected the same results for this experiment as well. Thus, we predicted that the affector will receive high causal ratings in the P2A and AP2 scenarios relative to the PA2 scenario.

5.1. Method

5.1.1. Participants

This experiment was pre-registered before any data were collected. The pre-registration on Open Science Framework can be accessed at <https://osf.io/hf274>. We recruited participants through an online platform (Prolific), and they participated in exchange for monetary compensation (GBP 0.75) for a 5-minute study. We aimed to recruit 250 participants per condition (3 conditions in total) to get a final sample size of 750. A total of 901 participants completed the questionnaire. Participation was limited to individuals who reported a fluent comprehension of English and being between 18 and 65 years of age. Participants were excluded from the final sample if they (a) failed to answer one or more of the appropriateness ratings questions, (b) answered one or more of the comprehension questions incorrectly, or (c) were not within the first 250 participants in their condition to meet criteria (a) and (b). After participants were excluded according to these criteria, we had a total of 750 participants, with 250 in the AP2 condition, 250 in the PA2 condition, and 250 in the P2A condition.

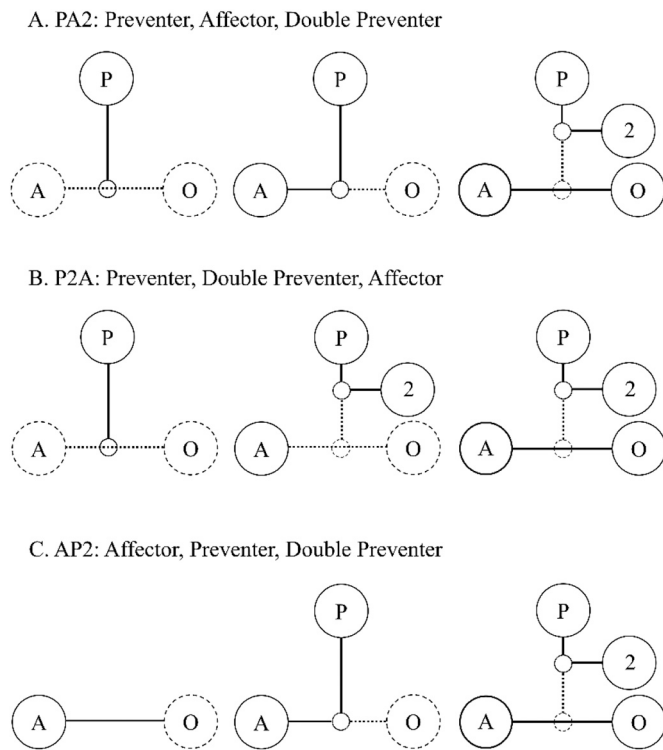


Fig. 6. Illustration of the PA2, P2A, and AP2 Double Prevention Scenarios. *Note.* Circles represent actions and events and lines represent the relations among them. Circles with a solid perimeter represent that the action or event has been realized, whereas circles with a dotted perimeter indicate an unrealized action or event. Solid lines represent that a relation is presently realizable, whereas dotted lines denote presently unrealizable relations. Successive frames depict the temporal sequence of events. In the PA2 condition, the preventer (P) attempts the prevention, the affector (A) carries out the action, and then the double preventer (2) removes the preventer’s prevention, resulting in the outcome (O) occurring. In the P2A condition, the preventer attempts the prevention, the double preventer removes the preventer’s prevention, and then the affector carries out the action. In the AP2 condition, the affector carries out the action, the preventer attempts the prevention, and the double preventer removes the preventer’s prevention.

5.1.2. Materials

The experiment consisted of an online survey that presented participants with a double prevention scenario involving the characters Anna, Brad, and Clara. Participants in the current study were randomly assigned to one of three versions of the scenario (AP2, PA2, P2A). In all versions of the scenario, the affector (Anna) is intentional, and the double preventer (Clara) is unintentional. An example of one version of the scenario (AP2) is below:

Anna, Brad, and Clara work at a shipping facility. Anna’s job is to place packages onto a conveyor belt that leads to a shipping dock on the other side of the building. The packages travel down the conveyor belt to a sorting machine that scans the packages and then sorts them into different collection bins. The packages in each collection bin are then loaded onto different delivery trucks at 6 PM. Anna finishes her shift at 5 PM but just before leaving she realizes that she forgot to place a package on the conveyor belt. She places the package on the conveyor belt and leaves work. Brad is finishing his work and believes that no more packages will be coming on the conveyor belt, so he flips a switch in the control room to turn off the sorting machine and then leaves work. When the scanner is turned off, packages on the conveyor belt are blocked from moving forward and do not reach the collection bins. Just as the package is approaching the scanner, Clara accidentally turns on the sorting machine by bumping into the switch in the control room as she is leaving for the

day. The sorting machine scans the package and directs it to the appropriate collection bin at the shipping dock before 6 PM.

In the PA2 version of the scenario, Brad attempts the prevention first, then Anna acts, and then Clara removes Brad’s prevention attempt. In the P2A version of the scenario, Brad attempts the prevention first, then Clara removes Brad’s prevention attempt, and then Anna acts at the end. After reading the scenario, participants had to evaluate the appropriateness of claims regarding Anna, Brad, and Clara on a 6-point scale (from “Completely Inappropriate” to “Completely Appropriate”). We ended the questionnaire by asking participants five yes-or-no comprehension questions about the scenario.

5.1.3. Procedure

The experiment consisted of a questionnaire created on an online platform called Qualtrics (www.qualtrics.com). Participants were presented with a description of the study and consented to participate by clicking a link from Prolific that redirected them to the Qualtrics questionnaire. At the beginning of the questionnaire on Qualtrics, participants were provided with a Participant Information Sheet. Participants were randomly assigned to one of the three conditions described above. Additionally, the order of the statements about each character for the appropriateness ratings and the order of the comprehension questions were randomized for each participant. The participants could read the scenario at the top of the page when completing the questions regarding the appropriateness ratings for each character and the comprehension questions.

5.2. Results

We preregistered three central predictions for this experiment: (1) appropriateness ratings will be significantly lower for the affector in the PA2 condition as compared to the P2A and AP2 conditions; (2) appropriateness ratings will be significantly lower for the double preventer in the P2A condition as compared to the PA2 and AP2 conditions; and (3) appropriateness ratings will be higher for the affector than the double preventer in the P2A condition and the appropriateness ratings will be higher for the double preventer than the affector in the PA2 condition.

We first tested the prediction that appropriateness ratings will be significantly lower for the affector in the PA2 condition as compared to the P2A and AP2 conditions. A significant effect of condition was observed in the affector’s appropriateness ratings, $F(2, 747) = 18.49, p < .001, \eta^2p = .047$. Post-hoc analyses using Bonferroni corrections revealed significantly higher appropriateness ratings for the affector in the P2A condition ($M = 4.53, SD = 1.24$) than the PA2 ($M = 4.05, SD = 1.23$) and the AP2 conditions ($M = 3.84, SD = 1.41$), p -values $< .001$ (see Fig. 7). The difference in the affector’s appropriateness ratings between the PA2 and AP2 conditions was not significant, $p = .22$.

Next, we tested the prediction that appropriateness ratings will be significantly lower for the double preventer in the P2A condition as compared to the AP2 and PA2 conditions. A significant effect of condition was observed in the double preventer’s appropriateness ratings, $F(2, 747) = 16.36, p < .001, \eta^2p = .042$. Post-hoc analyses using Bonferroni corrections revealed significantly lower appropriateness ratings for the double preventer in the P2A condition ($M = 4.35, SD = 1.20$) than the AP2 condition ($M = 4.84, SD = 1.04$) and PA2 condition ($M = 4.84, SD = 1.04$), p -values $< .001$ (see Fig. 7). The difference in the double preventer’s appropriateness ratings between the AP2 and PA2 conditions was not significant, $p > .99$.

Finally, we tested the prediction that appropriateness ratings will be higher for the affector than the double preventer in the P2A condition and the appropriateness ratings will be higher for the double preventer than the affector in the PA2 condition. The 2 (Condition: P2A, PA2) x 2 (Character: Affector, Double Preventer) ANOVA revealed a significant main effect of character, $F(1, 498) = 15.38, p < .001, \eta^2p = .030$, with the double preventer receiving higher appropriateness ratings ($M =$

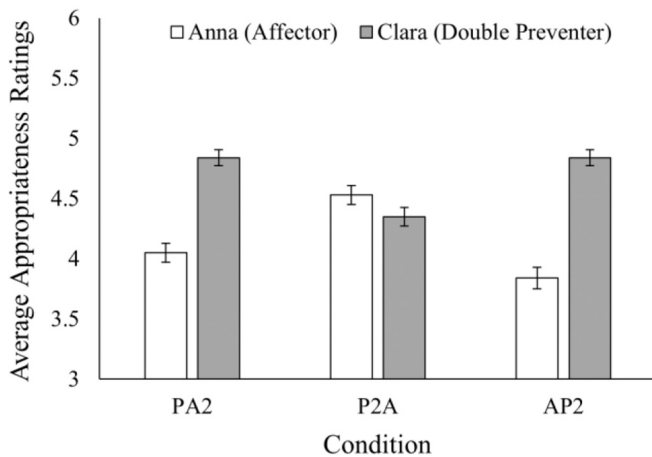


Fig. 7. Average “Because of” Appropriateness Ratings for Anna and Clara as a Function of Condition.

Note. Appropriateness ratings are on a scale of 1 to 6. The conditions included scenarios in which the preventer attempted the prevention first, then the affector carried out the action, and then the double preventer removed the preventer’s prevention (PA2); the preventer attempted the prevention first, then the double preventer removed the preventer’s prevention, and then the affector carried out the action (P2A); the affector acted first, then the preventer attempted the prevention, and then the double preventer removed the preventer’s prevention (AP2).

4.59, $SD = 1.15$) than the affector ($M = 4.29$, $SD = 1.26$). However, a significant interaction between condition and character was also observed, $F(1, 498) = 38.16$, $p < .001$, $\eta^2p = .071$. Follow-up t-tests demonstrated that the double preventer ($M = 4.84$, $SD = 1.04$) received significantly higher appropriateness ratings than the affector ($M = 4.05$, $SD = 1.23$) in the PA2 condition, $t(249) = -7.85$, $p < .001$, whereas the appropriateness ratings in the P2A condition were descriptively – but not significantly, $p = .14$ – higher for the affector ($M = 4.53$, $SD = 1.24$) than the double preventer ($M = 4.35$, $SD = 1.20$) (see Fig. 7).

5.3. Discussion

Process theories emphasize that reasoners’ causal ratings will be sensitive not only to the occurrence or non-occurrence of an outcome, but also how the events leading up to an outcome unfolded. We systematically altered double prevention scenarios so that the characters’ actions followed different temporal sequences to understand how the causal ratings attributed to the affector and double preventer would be impacted. Our first prediction for Experiment 3 was based on the notion that the affector would receive high causal ratings in conditions in which they were represented – however briefly – as being able to bring about the outcome on their own. Consequently, we predicted that the affector would receive significantly lower appropriateness ratings in the PA2 condition as compared to the P2A and AP2 conditions. This prediction was partially supported. The affector did receive significantly lower appropriateness ratings in the PA2 condition than the P2A condition. However, the affector also received significantly lower appropriateness ratings in the AP2 condition relative to the P2A condition.

Originally, we expected that the affector would receive high appropriateness ratings in the AP2 condition because we considered the scenario presented in the condition to be comparable to the scenario featured in Experiments 1 and 2. However, the scenario presented in the AP2 condition of the current experiment differed from the scenario featured in Experiments 1 and 2 in a crucial manner that we failed to consider. Specifically, the prevention attempt in the AP2 condition was represented to be *in place* in Experiment 3, with the route connecting the affector to the outcome fully blocked at one point in time. In contrast, the prevention attempt presented in the scenario used in Experiments 1

and 2 never resulted with the route connecting the affector and the outcome being blocked. Instead, it was implied that the preventer would have been able to prevent the outcome if they had been able to continue pulling on the electrical cord without intervention by the double preventer.

This suggests that if the preventer’s prevention is ever represented as being in place (posing an *actual* as opposed to a *counterfactual* threat), the affector will receive lower ratings (i.e., AP2 and PA2 conditions). Hence, although the preventer’s prevention attempt was not in place at the time of the affector’s action in the AP2 condition, the prevention attempt was in place later and presented an actual threat. This was sufficient for participants to provide a lower rating to the affector in the AP2 condition as compared to the P2A condition. This suggests that causal ascriptions to the affector are sensitive not only to the conditions at the time of action, but also to subsequent threatened preventions, even when the preventions themselves are ultimately thwarted.

For our next hypothesis, we predicted that the appropriateness ratings would be significantly lower for the double preventer in the P2A condition as compared to the PA2 and AP2 conditions. The results were in line with our hypothesis. We found that an unintentional double preventer can receive high appropriateness ratings when their action removes a prevention attempt that posed an actual threat to an outcome’s occurrence. In the P2A condition, although the double preventer acts upon an actual (i.e., in place) prevention attempt, the actions of the preventer and double preventer are not linked to the outcome because the affector has yet to act. This indicates that causal ascriptions to the double preventer depend upon a link extending not solely to the prevention attempt but also to the outcome. In other words, the unintentional double preventer does not receive high causal ratings unless they are linked to the outcome via a prevention attempt that poses an actual (as opposed to counterfactual) threat.

6. Experiment 4

In Experiment 3, we examined how the temporal sequence of the characters’ actions in double prevention scenarios influences causal ascriptions. The results of Experiment 3 suggest that a key determinant of the causal ratings awarded to the affector and double preventer concerns whether the prevention attempt was *actual* (represented as in position to prevent an action that had taken place) or *counterfactual* (represented as nearly in a position to prevent an action that had taken place or would later take place). To further test this account, Experiment 4 compared a modified version of the AP2 scenario from Experiment 3, in which the prevention attempt was actual, with a new condition in which the prevention attempt was counterfactual: the A2P condition (see Fig. 8). In this new condition, the prevention attempt is never

A2P: Affector, Double Preventer, Preventer

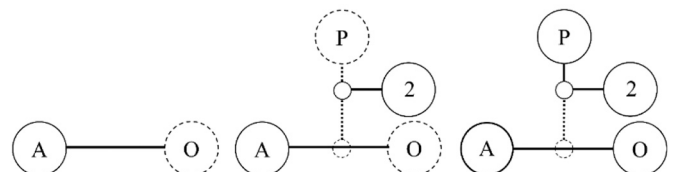


Fig. 8. Illustration of the A2P Double Prevention Scenario.

Note. Circles represent actions and events and lines represent the relations among them. Circles with a solid perimeter represent that the action or event has been realized, whereas circles with a dotted perimeter indicate an unrealized action or event. Solid lines represent that a relation is presently realizable, whereas dotted lines denote presently unrealizable relations. Successive frames depict the temporal sequence of events. In the A2P condition, the affector (A) attempts to bring about an outcome (O), the double preventer (2) blocks a hypothetical prevention route, and then the preventer (P) attempts the prevention.

represented as being in place because the double preventer has acted before the preventer. Consequently, the A2P condition is most comparable to the scenario used in Experiments 1 and 2. We therefore predicted that appropriateness ratings will reveal a significant interaction between condition and character such that the double preventer will receive higher appropriateness ratings relative to the affector in the AP2 condition but not in the A2P condition.

6.1. Method

6.1.1. Participants

This experiment was pre-registered before data collection commenced. The pre-registration on Open Science Framework can be accessed at <https://osf.io/hm94y>. We recruited participants through an online platform (Prolific), and they participated in exchange for monetary compensation (GBP 0.75) for a 5-minute study. We aimed to recruit 250 participants per condition (2 conditions in total) to get a final sample size of 500. We based this number on previous research with similar scenarios (Experiments 1-3). In accordance with our pre-registration, we only included the first 250 participants from each condition who met the criteria, and any additional participants were excluded. A total of 675 participants completed the questionnaire. Participation was limited to individuals who reported a fluent comprehension of English and being between 18 and 65 years of age. Participants were excluded from the final sample if they (a) failed to answer one or more of the appropriateness ratings questions or (b) answered one or more of the comprehension questions incorrectly. After participants were excluded according to the aforementioned criteria, we had a total of 500 participants, with 250 in the AP2 condition and 250 in the A2P condition.

6.1.2. Materials

The experiment consisted of an online survey that presented participants with a double prevention scenario involving the characters Anna, Brad, and Clara. In all versions of the scenario, the affector (Anna) is intentional, and the double preventer (Clara) is unintentional. Participants in the current study were randomly allocated to one of two versions of the scenario (AP2, A2P). For this experiment, we altered the scenario from Experiment 3 to accommodate the A2P double prevention sequence. This alteration was necessary because the scenario from Experiment 3 did not allow for a sequence where the double preventer acted before the preventer. The AP2 version of the new scenario is presented below:

Anna, Brad, and Clara work at a shipping facility. Anna's job is to place packages onto a conveyer belt that leads to a shipping dock on the other side of the building. The packages travel down the conveyer belt and are then loaded onto delivery trucks at 6 PM. It is 5 PM and Anna has finished her shift, but just as she is about to leave, she notices a package on the floor. She places the package on the conveyer belt and leaves. Earlier that day, Brad installed a new scanner on the conveyer belt. The scanner is currently turned off, but when it is turned on, it is designed to scan the packages for specific labels. If it doesn't detect one of these labels on a package, it will push the package into a separate collection bin to be sorted out manually the following day. Since it is after 5 PM, Brad believes that no more packages will be on the conveyer belt for the rest of the day, so he flips a switch in the control room to turn the scanner on. If this scanner is turned on, the package Anna sent through will not reach the shipping dock because the package does not have the specific label that is compatible with the scanner. As the package is approaching the scanner, Clara accidentally bumps into the scanner and disconnects it from the power outlet which turns the scanner off. The package continues down the conveyer belt past the scanner and reaches the shipping dock before 6 PM.

In the A2P version of the scenario, the affector carries out the action first, then the double preventer blocks the preventer's prevention attempt, and then the preventer attempts to prevent the outcome from

occurring. After reading the scenario, participants had to evaluate the appropriateness of claims regarding Anna, Brad, and Clara on a 6-point scale (from "Completely Inappropriate" to "Completely Appropriate"). We ended the questionnaire by asking participants five yes-or-no comprehension questions about the scenario.

6.1.3. Procedure

The experiment consisted of a questionnaire created on an online platform called Qualtrics (www.qualtrics.com). Participants were presented with a description of the study and consented to participate by clicking a link from Prolific that redirected them to the Qualtrics questionnaire. At the beginning of the questionnaire on Qualtrics, participants were provided with a Participant Information Sheet. Participants were randomly assigned to one of the two conditions described above. Additionally, the order of the statements about each character for the appropriateness ratings and the order of the comprehension questions were randomized for each participant. The participants could read the scenario at the top of the page when completing the questions regarding the appropriateness ratings for each character and the comprehension questions.

6.2. Results

We preregistered one central prediction for this experiment; namely, that an interaction effect will be observed when evaluating the appropriateness ratings given to the affector and double preventer across the AP2 and A2P conditions such that the double preventer will receive higher appropriateness ratings relative to the affector in the AP2 condition but not in the A2P condition.

A 2×2 mixed ANOVA featuring condition (AP2, A2P) as a between-subject factor and character (Affector, Double Preventer) as a within-subjects factor revealed a significant interaction between condition and character, $F(1, 498) = 10.01, p = .002, \eta^2 p = .020$. Follow-up *t*-tests demonstrated that the double preventer received significantly higher appropriateness ratings in the AP2 condition ($M = 4.45, SD = 1.18$) than the affector ($M = 4.11, SD = 1.35$), $t(249) = -3.21, p = .001$. In the A2P condition, the ratings awarded to the double preventer ($M = 3.84, SD = 1.52$) and the affector ($M = 4.03, SD = 1.54$) did not differ significantly, $t(249) = 1.46, p = .14$ (see Fig. 9).

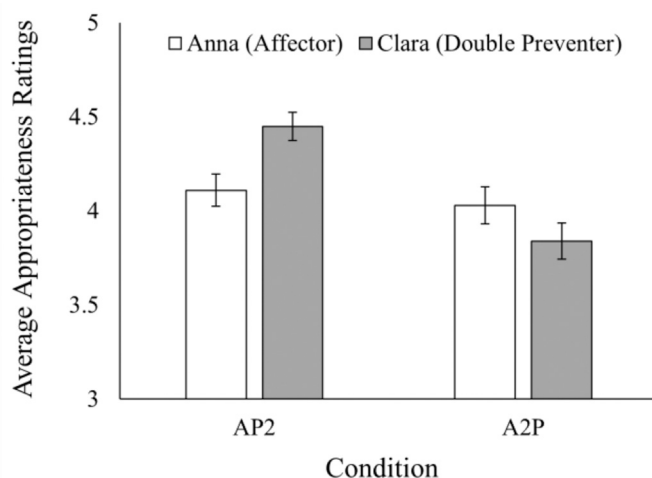


Fig. 9. Average "Because of" Appropriateness Ratings for Anna and Clara as a Function of Condition.

Note. Appropriateness ratings are on a scale of 1 to 6. The conditions included scenarios in which the affector acted first, then the preventer attempted the prevention, and then the double preventer removed the preventer's prevention (AP2); the affector acted first, then the double preventer removed the preventer's future prevention attempt, and then the preventer attempted the prevention (A2P). Error bars denote standard errors.

6.3. Discussion

As predicted, we observed a significant interaction between the ratings provided to the affector and double preventer across the AP2 and A2P conditions. In the AP2 condition, the double preventer received significantly higher appropriateness ratings than the affector, whereas the double preventer received descriptively *lower* appropriateness ratings than the affector in the A2P condition. These results confirm that the distinction between scenarios featuring *actual* prevention threats (the AP2 condition in the current experiment, and the AP2 and PA2 conditions from Experiment 3) and *counterfactual* prevention threats (the A2P condition in the current experiment and the P2A condition in Experiment 3) is consequential for the causal ratings provided to the double preventer.

Although process theories emphasize the importance of going beyond the occurrence or non-occurrence of events to consider how events unfold over time, research within the dependency tradition has also highlighted how temporal sequences can impact causal reasoning. For example, Spellman (1997) suggested that reasoners attribute causality to different events within a sequence by computing how each event altered the probability of an outcome. In the context of the current experiment, this account offers an explanation for why the double preventer received low causal ratings in the A2P condition; namely, because the double preventer's action did not increase the probability of the outcome when the action took place. However, without further elaboration, this account does not offer an explanation as to why the intentional – but not the unintentional – double preventer received high causal ratings in the A2P condition presented in Experiment 1. Additionally, Spellman's (1997) theory appears to predict that the affector should receive high causal ratings in the AP2 condition because the affector's action increased the probability of the outcome when the action was performed. However, we found that the affector received low causal ratings in the AP2 condition.

Although Spellman's theory was not tested in the context of double prevention scenarios that manipulated intentional status, it is possible that the theory could be elaborated to account for our findings. For instance, Gerstenberg and Lagnado (2012) presented an extension of Spellman's account in which judgments of responsibility reflect not only whether an action made a difference to the outcome as it occurred, but also whether the action would have made a difference if the situation had been somewhat different. Thus, an intentional double preventer in an A2P scenario may be judged as more causal of an outcome than the unintentional double preventer because reasoners expect that the actions of an intentional double preventer would be more likely to make a difference in the scenario, even if the events unfolded in a slightly different manner. We further explore why intentionality may have increased causal ratings for the double preventer in the A2P condition in the following General Discussion.

7. General discussion

Understanding how we reason about causal relations is essential to understanding a wide range of human behavior. The current study built on foundational work by Lombrozo (2010) to evaluate three challenges to the account of causal pluralism she developed to capture reasoning in double prevention scenarios featuring intentional or accidental actions. In the following, we discuss each of these challenges in turn.

Our first challenge evaluated whether intentionality might influence causal ratings through a route other than the mechanism outlined by the causal pluralism account proposed and tested by Lombrozo (2010) in which intentional actions lead reasoners to adopt a teleological mode of explanation and increase the weight placed on dependence relations. Results from Experiment 1 showed that the affector received significantly higher ratings when acting intentionally (A+ conditions) versus unintentionally (A- conditions). Results from our pilot experiment and Experiment 2 showed that the double preventer received significantly

higher appropriateness ratings than the affector in the A-C+ condition in which the double preventer acted intentionally and the affector did not. This effect did not reach significance in Experiment 1, though a descriptive difference in the predicted direction was observed.

We suggest that the most parsimonious interpretation of the effects of intentionality observed in these experiments is that intentionality led to a boost in appropriateness ratings given to each of the characters. On this view, the affector's ratings in the A+ conditions received boosts from (a) being linked to the outcome via a direct transfer of force and (b) acting intentionally. In the A- conditions, the affector did not receive a boost in ratings from acting intentionally, leading to significantly lower ratings in the A- condition relative to the A+ condition. In the A-C+ condition, the double preventer received significantly (pilot experiment, Experiment 2) or descriptively (Experiment 1) higher appropriateness ratings than the affector because the effect of intentionality was greater than the effect of being linked to the outcome via a direct transfer of force.

To bolster this interpretation, we performed exploratory analyses comparing the appropriateness ratings awarded to the affector and the double preventer in the A+C+ condition in Experiment 1 and our pilot experiment. These analyses revealed that the affector received significantly higher appropriateness ratings than the double preventer. This finding indicates that the affector's ratings reflected both an effect of being linked to the outcome via the direct transfer of force and an effect of acting intentionally, whereas the double preventer's ratings reflected the latter but not the former. Notably, this difference was not observed in Lombrozo's original experiment, leading the author to note that the lack of this difference suggested that the affector's process relation, "played a more minor role in evaluating relationships supporting teleological explanations than in those supporting only mechanistic explanations" (Lombrozo, 2010, pg. 314). On our view, the presence of a significant difference in the ratings for the affector and double preventer in the A+C+ condition casts doubt on the weighting process proposed by the causal pluralism account. This is because being linked to the outcome via a direct transference of energy appears to have led to a boost in appropriateness ratings for the affector. Although the effects of intentionality outlined above do not provide conclusive counterevidence to the causal pluralism account, we believe the effects are most parsimoniously explained by a reasoning process that takes into consideration relevant factors (e.g., such as the intentionality of the character or whether the character is linked to the outcome via a direct transfer of energy) without the need to appeal to separately maintained and differentially weighted process and dependence relations.

The second challenge presented in the current study concerned the causal pluralism account's claim that double preventers are not linked to the outcome via a process link. Lombrozo (2010) proposed that the lack of a direct transference of energy between the double preventer and outcome led participants to award low causal ratings to unintentional double preventers. However, research from Wolff and Barbey (2015) indicates that double preventers are often judged to bring about outcomes through allow (as opposed to cause) relations, as predicted by force theory. This finding indicates that a direct transference of force from the double preventer to the outcome is not necessary for reasoners to form a process link between the double preventer and the outcome. We therefore asked participants to judge whether characters in double prevention scenarios "caused", "allowed", or "did not cause or allow" the outcome to occur (as in Wolff & Barbey, 2015). In keeping with the predictions of force theory, our results lend credence to the claim that reasoners are capable of thinking about the double preventer's relation to the outcome in terms of process even though the double preventer is not connected to the outcome via a direct transference of energy.

Again, this finding does not provide conclusive counterevidence to the causal pluralism account, but it does suggest that the explanatory power of process theories may have been underestimated by the causal pluralism account. Further, this result raises important questions regarding the extent to which reasoners maintain multiple, distinct

representations when reasoning about the link(s) connecting a character and an outcome. Although it is theoretically possible that reasoners maintain separate dependence and process representations linking characters to outcomes, it is arguably more parsimonious to propose that reasoners instead maintain composite representations that reflect a range of factors, including information about dependency, process, and intentionality.

Related to the second challenge, we also explored how the presence or absence of intentional actions affects reasoners' judgments of cause and allow relations for the double preventer. As noted above, participants in Experiment 1 gave a higher proportion of "allow" responses for the double preventer as compared to "cause" and "did not cause or allow" responses, across all four conditions. To investigate the extent to which the intentional status of the double preventer impacted these judgments, we directly compared the responses to the agreement questions for the double preventer when she behaved intentionally versus unintentionally. Participants were significantly more likely to classify the double preventer's action as allowing the outcome and significantly less likely to classify the double preventer's action as not causing or allowing the outcome when the double preventer behaved intentionally (as opposed to unintentionally). The finding demonstrates that intentional status does have an effect on the allow relations that participants attribute to the double preventer. Although further research is needed to explore how, when, and why the intentional status of the double preventer impacts the classification of allow and neither cause nor allow relations, in the following discussion of our third challenge, we speculate as to how the intentional status of the double preventer may have impacted these classifications.

Our third and final challenge to the causal pluralism account also took inspiration from process theories by considering how the temporal sequence in which the characters featured in double prevention scenarios act impacts causal ratings. In direct contrast to the predictions of the causal pluralism account, the results of Experiments 3 and 4 demonstrate that an unintentional double preventer can receive significantly higher causal ratings than an intentional affector when the prevention attempt described in the scenario is represented as actual (in position to prevent an action that had taken place) as opposed to counterfactual (nearly in a position to prevent an action that had taken place or temporarily in position to prevent an action that would later take place). Crucially, the prevention attempt described in the scenarios used in Experiments 1 and 2 never presented an actual threat, as the preventer's action never blocked the affector's link to the outcome (i.e., the power cord was never removed from the socket). Consequently, our distinction between prevention attempts that were represented to present an actual versus a counterfactual threat provides an explanation as to why the unintentional double preventer received low causal ratings and was more likely to be characterized as not having cause or allowed the outcome in Experiment 1. However, can this distinction account for the relatively high causal ratings received by the intentional double preventer in Experiments 1 and 2?

On our account of the judgments given to the unintentional double preventer described in Experiment 1, reasoners awarded the double preventer relatively low causal ratings because the threat posed by the prevention attempt is never represented as actual (meaning in position to prevent an action that had taken place). In the case of the intentional double preventer described in Experiments 1 and 2, however, we contend that reasoners *do* represent the prevention attempt as actual, but reasoners do so by representing the double preventer's representation of the scenario. That is, reasoners represent the intentional double preventer as having the option to select between two possibilities: (1) fail to intervene, in which case the prevention attempt will be in place (actual) and the outcome will not occur; or (2) intervene, in which case the prevention attempt will never have been in place (counterfactual) and the outcome will occur. Although the prevention attempt itself is only counterfactual in the scenario (the path connecting the affector to the outcome is never actually blocked), we contend that most reasoners

nevertheless represent a world in which the prevention attempt is successful when they consider the intentional double preventer's decision to intervene. That is, many reasoners do, however briefly, represent the threat of the prevention attempt as actual (or in place) when they consider the possible worlds represented by the intentional double preventer.

If this line of argument were correct, should we not expect that reasoners also represent a world in which the prevention attempt is successful when the double preventer is described as unintentional in scenarios like that of Experiment 1? Reasoners are surely aware that the outcome would have been prevented if the double preventer's unintentional action had not occurred. Would this not suggest that reasoners are at least implicitly representing a possible world in which the prevention attempt was successful? Although we agree that reasoners are aware that the outcome would have been prevented if the double preventer's unintentional action had not occurred, the scenario itself makes it clear that the unintentional double preventer did not entertain a possible world in which the prevention attempt was successful. In contrast, when the double preventer behaves intentionally, reasoners understand that the double preventer must choose between two possible worlds: one in which the double preventer fails to act and the outcome is prevented, and one in which the double preventer acts and the outcome occurs.

In other words, reasoning about a possible world in which the outcome is prevented goes directly against what reasoners were told in the scenarios featuring an unintentional double preventer, whereas doing so is arguably essential to effective reasoning about the scenarios featuring an intentional double preventer. Our account can therefore be viewed as a sort of process-grounded dependency theory (or dependency-enabling process theory) in which process-based considerations (e.g., the order in which each character acts) serve to generate richer models of possible worlds (e.g., a world in which the double preventer chooses to not intervene). In this way, we view our account as consistent with the overarching insight of the causal pluralism account (namely, that reasoners might think about causal scenarios in terms of both process and dependency), without being tied to the specific mechanistic claims featured in the account developed and tested by [Lombrozo \(2010\)](#) regarding the differential weighting of separately maintained process and dependence relations.

7.1. Connections to counterfactual models

The current study took inspiration from recent developments in the literature on process theories of causation (e.g., [Wolff & Barbey, 2015](#)) in developing challenges to the causal pluralism account of reasoning about cases of double prevention. However, the causal pluralism account developed and tested by [Lombrozo \(2010\)](#) has also encountered challenges from recent developments within the literature on counterfactual theories (a type of dependence theory) of double prevention ([Henne & O'Neill, 2022](#); [O'Neill, Quillien, & Henne, 2022](#)). For instance, [Henne and O'Neill \(2022\)](#) investigated the extent to which causal judgments in double prevention scenarios were linked to reasoner's beliefs that a particular outcome would not have occurred if either the affector or the double preventer had not acted. In their first experiment, they presented reasoners with a vignette in which an affector accidentally knocked against a bottle of beer. A preventer then reached out to catch the bottle, but a double preventer accidentally bumped into the preventer, causing the preventer to miss the catch, resulting in the bottle spilling on the ground. [Henne and O'Neill \(2022\)](#) found that reasoners judged the affector to be highly causal of the outcome and agreed strongly with the claim that the bottle would not have spilled if the affector had not knocked against it. By contrast, reasoners judged the double preventer to be less causal of the outcome and were more accepting of the possibility that the bottle would have spilled even if the double preventer had not bumped into the preventer. These results suggest that reasoners were using counterfactual reasoning to generate their causal ratings for the affector and the double preventer,

leading the authors to question whether reasoning in double prevention scenarios should be interpreted as evidence in favor of the causal pluralism account.

Henne and O'Neill (2022, Experiment 3) further explored how manipulating counterfactual reasoning would impact causal reasoning by presenting participants with a scenario in which reasoner's strongly agreed that the outcome would not have occurred if not for the double preventer. The authors accomplished this by presenting participants with a video in which a stationary ball (the preventer) blocked another ball's path from reaching a goal. Yet another ball (the double preventer) rolled into the stationary (preventer) ball, knocking it out of the path of the ball traveling toward the goal. In the video, it was clear that the ball traveling toward the goal would not have entered the goal if the double preventer had not cleared the path. Correspondingly, participants rated the double preventer as highly causal of the outcome.

Although we were unaware of the work of Henne and O'Neill (2022) when we designed Experiments 3 and 4, we see very clear parallels between their results and our own. For example, their experiment featuring a stationary ball acting as the preventer is comparable to our PA2 condition, wherein the preventer's prevention attempt is in place to block the outcome from occurring (presenting an actual threat). Our PA2 condition also revealed high causal ratings for the double preventer, consistent with their results and, more generally, their counterfactual reasoning account. However, Henne and O'Neill (2022) found that reasoners judged the affector in their scenario (another ball that set the ball travelling toward the goal in motion) to be highly causal of the outcome and agreed strongly with the claim that the ball would not have entered the goal if the affector had not set it in motion. In contrast, we found that the affector in our PA2 condition received significantly lower causal ratings than the double preventer.

Although we did not ask counterfactual reasoning questions in the same manner as Henne and O'Neill (2022), we did ask participants the following comprehension question: "Would the package have reached the appropriate collection bin before 6 PM if Anna had not placed the package on the conveyer belt?" Crucially, more than 99% of participants in the PA2 condition in Experiment 3 indicated that the outcome would not have occurred if Anna (the affector) had not acted, suggesting that our results might be difficult for Henne and O'Neill's (2022) counterfactual account to explain. However, additional research is necessary to evaluate the account directly. That said, we see our account as generally consistent with Henne and O'Neill's (2022) counterfactual account in the sense that their account at least implicitly underscores the importance of taking process information (e.g., details regarding how the double prevention scenario unfolded over time) into account.

Gerstenberg et al. (2021) also recently introduced a model of causal reasoning rooted in dependence theory called the *counterfactual simulation model* (CSM). The CSM has been rigorously tested using video clips of different scenarios featuring billiard ball collisions, including scenarios featuring double prevention. Like Henne and O'Neill's (2022) counterfactual theory outlined above, the CSM can be understood to underscore the importance of taking process information into account given that the precise sequence of events presented in the video clips is consequential for causal judgments. However, Gerstenberg et al. (2021) view their account as distinct from process theories such as force theory in that the CSM proposes that "causal judgments are intimately linked to counterfactuals, and that even understanding simple causal judgments requires considering counterfactual contrasts" (pg. 938). Although Sosa, Ullman, Tenenbaum, Gershman, and Gerstenberg (2021) applied the CSM to morally laden scenarios and Wu, Sridhar, and Gerstenberg (2022) extended the model to account for responsibility attributions by integrating causal attribution and mental state inferences, we are unaware of work applying the CSM to reasoning about intentional versus unintentional actions in double prevention scenarios. Future research should therefore explore the extent to which the model can account the results of intentionality manipulations (Experiments 1 and 2) and our temporal sequence manipulations (Experiments 3 and 4).

7.2. Voluntary actions in causal chains

Although we are unaware of previous research that systematically investigated sequence effects in double prevention scenarios in the manner presented in Experiments 3 and 4, previous work has examined how individuals attribute causality to voluntary actions and physical events differently at more distal or proximal points in a causal chain (e.g., Hilton, McClure, & Moir, 2016; Hilton, McClure, & Slugoski, 2005; Hilton, McClure, & Sutton, 2010; McClure, Hilton, & Sutton, 2007). For instance, to compare the effects of distal and proximal causes, researchers have used scenarios featuring *opportunity chains* wherein an initial event provides an opportunity for a second event to produce an outcome. An example of such a scenario comes from Hart and Honoré (1985): "[Person] A throws a lighted cigarette into [a fern at the edge of a forest] which catches fire. Just as the flames are about to flicker out [Person] B who is not working in concert with A, deliberately pours petrol [on the fern]. The fire spreads and burns down the forest." In this example, the distal cause (Person A throwing the cigarette) was voluntary, as was the proximal cause (Person B pouring petrol on the flame). Such an example can be contrasted with a case in which the proximal cause was physical instead of voluntary (e.g., a strong breeze sprang up just as the flame was about to flicker out, causing the fire to spread and the forest to burn down).

Hart and Honoré (1985) proposed that reasoners (a) prefer explanations featuring two voluntary causes as opposed to physical causes and (b) are inclined to work backwards from the outcome and settle upon the proximal voluntary action as being the most important cause of an event when the chain contains two voluntary causes. McClure et al. (2007) tested these predictions and found that reasoners did indeed judge chains containing two voluntary causes to be better explanations than chains featuring physical causes. However, McClure et al. did not observe evidence that reasoners preferred a proximal voluntary cause over a distal voluntary cause in chains featuring two voluntary causes.

Hilton et al. (2010) explored another type of causal chain called an *unfolding causal chain* in which prior events enable subsequent events to have a causal effect. In contrast to opportunity chains, unfolding causal chains imply transitivity of causation such that A should be judged a cause of C if A causes B and B causes C. In unfolding causal chains, prior events cause the subsequent events to happen, whereas in opportunity chains, prior events enable subsequent events to have a causal effect (Hilton et al., 2010). Hilton et al. provided stronger evidence for Hart and Honoré's (1985) notion that reasoners are inclined to backwards from the outcome and settle upon the first voluntary action in a chain as being the most important cause of the event. Although select results from the current study appear to be inconsistent with this notion (e.g., the results of Experiment 4 indicate that an unintentional double preventer can receive higher causal ratings than an intentional affector in the AP2 condition in which the double preventer behaves last), it is important to emphasize that the scenarios used in the current study differed substantially from those featured in Hilton et al. (2010).

Hilton et al. (2016) went on to examine how an agent's awareness of an eventual outcome impacts causal attributions for voluntary actions in double prevention scenarios. In one such scenario, rocks from a landslide bent a rail on a train track. Repairmen put up a warning light to warn train drivers of the bent rail. A group of saboteurs then cut the wires to the warning light, thus causing it not to work. A train then came along and went off the tracks at the bent portion of the rail, leading to a number of injuries. Crucially, the saboteurs were described as knowing that the rail was bent in one condition and unaware that the rail was bent in another condition. Thus, the saboteurs either knowingly or unknowingly took advantage of an opportunity that they were presented with. Hilton et al. (2016) found that reasoners judged the actions of the saboteurs to be more causally important and more morally blameworthy when they were aware of the bent rail relative to the condition in which they were unaware.

The results of Hilton et al. (2016) are consistent with the view that intentional double preventers receive a boost in causal ratings because

the threat of the prevention attempt is represented by the double preventer to be actual as opposed to counterfactual. In the scenario in which the saboteurs are unaware of the bent rail, the act of cutting the wires to the warning light is not directly linked to a specific possibility of a derailment in the minds of the saboteurs and, consequently, the threat of the prevention attempt (i.e., the activation of the warning light) is relatively minor. However, when the saboteurs are aware of the bent rail, the threat of the prevention attempt stands out in full relief. Consequently, the double prevention is judged to be more causally important and more morally blameworthy. Future research should explore this interpretation of [Hilton et al.'s \(2016\)](#) results by systematically manipulating the sequence of events and the intentional status and awareness of the characters.

7.3. Limitations and future directions

Causal reasoning and the role of intentionality have been investigated across a range of different topics including moral reasoning ([Alicke et al., 2015](#); [Knobe & Fraser, 2008](#); [Martin & Cushman, 2016](#); [Murray & Lombrozo, 2017](#)), economics ([Celli, 2022](#); [Little, 2010](#)), and philosophy ([Dinh & Danks, 2021](#)), across different scenarios ([Gerstenberg et al., 2021](#); [Lagnado & Channon, 2008](#); [Walsh & Sloman, 2011](#)) and age groups ([Buchanan & Sobel, 2011](#); [Erb, Buchanan, & Sobel, 2013](#); [Gopnik et al., 2004](#)). Experiments 1 and 2 focused on one scenario involving Alice, Bob, and Carol as the affector, preventer, and double preventer, respectively, which was adopted from [Lombrozo's \(2010\)](#) first experiment. This double prevention scenario focused on observable physical human interactions. However, [Lombrozo](#) analyzed a range of other scenarios involving non-human biological traits and artifacts. Consequently, it is important to note that the first two challenges outlined above are specific to a small subset of the experiments that [Lombrozo \(2010\)](#) put forward as evidence in favor of causal pluralism. Given the highly contrived nature of the double prevention scenarios used in Experiments 1-4 and their focus on observable human interactions, future research should focus on scenarios that feature alternative causal systems (e.g., non-human biological systems as in [Lombrozo, 2010](#), Experiment 2).

Similarly, although it was important to begin this line of research by focusing on scenarios designed to be as morally neutral as possible, future research should explore how morally relevant factors might impact causal reasoning in double prevention scenarios that manipulate intentionality and temporal sequence. Such factors could include foreseeability, responsibility, and accountability ([Lagnado & Channon, 2008](#); [Lagnado, Gerstenberg, & Zultan, 2013](#); [Martin, Buon, & Cushman, 2021](#)). For instance, future work could develop scenarios in light of [Shaver's \(1985\)](#) theory of blame, which discusses the considerations that reasoners make when attributing responsibility to a character. These include the affector's causal contribution to the outcome, the affector's awareness of the consequences of their actions, the affector's intentionality, and the understanding of the moral wrongfulness of the action. Similarly, drawing on the work of [Lagnado and Channon \(2008\)](#), future work could explore how different types of foreseeability impact causal reasoning, including subjective foreseeability (the likelihood of an event occurring from the affector's point of view), objective foreseeability (the likelihood of the outcome occurring regardless of what the affector believes will happen), and reasonable foreseeability (what is reasonable for the affector to expect based on other factors and information they have).

Finally, future work should also explore how subtle differences in the wording of key questions might impact causal attributions. For instance, some reasoners might interpret the word "allow" in terms of permission. Given that intention is generally considered to be required to grant permission for an event to occur, some reasoners in the current study may have judged that the double preventer could not have permitted the outcome to occur without consciously intending to do so. As noted by [Wolff and Barbey \(2015\)](#), the concept of allow is closely related to and

can be interchanged with the concepts of enable, let, and help. Additionally, a study conducted by [Rose, Sievers, and Nichols \(2021\)](#) used the same scenario involving Alice, Bob, and Carol as [Lombrozo \(2010, Experiment 1a\)](#) but changed the word "cause" in the causal claims to another special causal word, "start" (e.g., "Carol started the music."). In contrast to the results of [Lombrozo \(2010, Experiment 1a\)](#), [Rose et al. \(2021\)](#) found that reasoners considered Alice but not Carol to be a cause when both characters behaved intentionally.

Similarly, [Vasil & Lombrozo \(2022\)](#) recently demonstrated that reasoners do not always interpret statements of the form "X caused Y" to be equivalent to statements of the form "Y occurred because of X". For example, judgments of explanatory goodness for statements of the form "Y occurs because of X" were found to be more sensitive to information regarding mechanism than causal strength judgments assessing whether a causal relationship between X and Y existed. In the current study, we phrased our appropriateness rating prompts using a "because of" relation (e.g., "The music played because of Alice.") in light of previous research indicating that "because of" relations encompass "cause" and "allow" relations ([Wolff, Klettke, Ventura, & Song, 2005](#)). Given our desire to avoid interpretations of the word "cause" as meaning "cause as distinct from allow", we felt using a "because of" relation was appropriate. However, it is also possible that our usage of the "because of" relation prompted participants to think more in terms of explanatory goodness rather than causal strength, resulting in a greater focus on particular types of information (e.g., information regarding mechanism). Consequently, the results of [Rose et al. \(2021\)](#) and [Vasil & Lombrozo \(2022\)](#) underscore the importance of directly comparing how subtle differences in wording might impact responding.

8. Conclusion

Experiments 1 and 2 explored two challenges to the causal pluralism account of reasoning in double prevention scenarios developed and tested by [Lombrozo \(2010\)](#). The first challenge concerned the various ways in which the intentional status of the characters featured in the scenarios could hypothetically impact reasoning. Our results lend credence to the claim that intentionality can directly impact the causal ratings attributed to the affector and double preventer, calling into question whether the differential weighting process proposed by the causal pluralism account is required to explain reasoning performance. The second challenge was inspired by [Wolff and Barbey's \(2015\)](#) research on force theory indicating that reasoners are quite comfortable describing the link between a double preventer and an outcome in terms of process relations. Our results showed that when reasoners were given the opportunity to attribute an allow relation to the double preventer, they frequently did so, and this attribution was also influenced by the intentional status of the double preventer.

Experiments 3 and 4 investigated how the temporal sequence of the characters' actions in double prevention scenarios influences causal ascriptions. Our results showed that the causal ratings attributed to the affector and double preventer were dependent on whether the preventer's prevention attempt was an actual (in place) versus a counterfactual (not in place) threat to the outcome. These findings indicate that process theories such as force theory have broader explanatory power than proposed by the causal pluralism account. Further, our results highlight fundamental questions concerning whether individuals form, maintain, and reason over (a) separable representations of dependence and process relations or (b) composite representations that reflect information about dependency, process, and intentionality. We believe that revisiting these questions will have important implications not only for our understanding of everyday causal reasoning in humans but also for efforts to design and test artificial systems tasked with learning and interacting with complex causal structures ([Glymour, Scheines, & Spirtes, 2014](#); [Pearl, 2018](#); [Pearl & Mackenzie, 2018](#)).

Author note

The data and analysis files associated with this study are available through the Open Science Framework at https://osf.io/nreqf/?view_only=815a37ead733448d983dd55152032300. The pre-registration for Experiments 1-4 are available at <https://osf.io/87dbh>, <https://osf.io/78sfj>, <https://osf.io/hf274>, and <https://osf.io/hm94y>, respectively.

CRedit authorship contribution statement

Huseina Thanawala: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Christopher D. Erb:** Writing – review & editing, Writing – original draft, Supervision, Formal analysis, Conceptualization.

Data availability

We have made all of our data files and analysis files available via the OSF website.

Acknowledgments

Special thanks to Fiery Cushman for formative input during the conceptualization of Experiments 3 and 4, and to Tania Lombrozo for insightful, constructive feedback throughout the review process.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2024.105786>.

References

- Alicke, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T., & Lagnado, D. A. (2015). Causal conceptions in social explanation and moral evaluation: A historical tour. *Perspectives on Psychological Science*, 10(6), 790–812. <https://doi.org/10.1177/1745691615601888>
- Aronson, J. L. (1971). On the grammar of 'cause'. *Synthese*, 22(3/4), 414–430.
- Beller, A., Bennett, E., & Gerstenberg, T. (2020). The language of causation. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd annual conference of the cognitive science society* (pp. 3133–3139). Cognitive Science Society.
- Buchanan, D. W., & Sobel, D. M. (2011). Mechanism-based causal reasoning in young children. *Child Development*, 82(6), 2053–2066. <https://doi.org/10.1111/j.1467-8624.2011.01646.x>
- Celli, V. (2022). Causal mediation analysis in economics: Objectives, assumptions, models. *Journal of Economic Surveys*, 36(1), 214–234. <https://doi.org/10.1111/joes.12452>
- Chang, W. (2009). Connecting counterfactual and physical causation. In , 31, No. 31. *Proceedings of the Annual Meeting of the Cognitive Science Society*. <https://doi.org/10.1037/0022-3514.58.4.545>
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, 58(4), 545–567. <https://doi.org/10.1037/0022-3514.58.4.545>
- Dinh, P. 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>
- Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science*, 59(2), 195–216. <https://doi.org/10.1086/289662>
- Erb, C. D., Buchanan, D. W., & Sobel, D. M. (2013). Children's developing understanding of the relation between variable causal efficacy and mechanistic complexity. *Cognition*, 129(3), 494–500. <https://doi.org/10.1016/j.cognition.2013.08.002>
- Fair, D. (1979). Causation and the flow of energy. *Erkenntnis*, 14(3), 219–250. <https://doi.org/10.1007/BF00174894>
- Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, 128(5), 936–975. <https://doi.org/10.1037/rev0000281>
- Gerstenberg, T., & Lagnado, D. A. (2012). When contributions make a difference: Explaining order effects in responsibility attribution. *Psychonomic Bulletin & Review*, 19, 729–736. <https://doi.org/10.3758/s13423-012-0256-4>
- Glymour, C., Scheines, R., & Spirtes, P. (2014). *Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling*. Academic Press.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review*, 111(1), 3–32. <https://doi.org/10.1037/0033-295X.111.1.3>
- Hart, H. L. A., & Honoré, A. M. (1985). *Causation in the law* (2nd ed.). Oxford: Oxford University Press.
- Henne, P., Kulesza, A., Perez, K., & Houcek, A. (2021). Counterfactual thinking and recency effects in causal judgment. *Cognition*, 212, Article 104708. <https://doi.org/10.1016/j.cognition.2021.104708>
- Henne, P., & O'Neill, K. (2022). Double prevention, causal judgments, and counterfactuals. *Cognitive Science*, 46(5), Article e13127. <https://doi.org/10.1111/cogs.13127>
- Hilton, D. J., McClure, J., & Moir, B. (2016). Acting knowingly: effects of the agent's awareness of an opportunity on causal attributions. *Thinking & Reasoning*, 22(4), 461–494. <https://doi.org/10.1080/13546783.2016.1191547>
- Hilton, D. J., McClure, J., & Slugoski, B. R. (2005). The course of events: Counterfactuals, causal sequences and explanation. In D. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The Psychology of Counterfactual Thinking* (pp. 44–60). London: Routledge.
- Hilton, D. J., McClure, J., & Sutton, R. M. (2010). Selecting explanations from causal chains: Do statistical principles explain preferences for voluntary causes. *European Journal of Social Psychology*, 40, 383–400. <https://doi.org/10.1002/ejsp.623>
- Khemlani, S. S., Barbey, A. K., & Johnson-Laird, P. N. (2014). Causal reasoning with mental models. *Frontiers in Human Neuroscience*, 8(849), 1–15. <https://doi.org/10.3389/fnhum.2014.00849>
- Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. *Moral Psychology*, 2, 441–447.
- Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, 108(3), 754–770. <https://doi.org/10.1016/j.cognition.2008.06.009>
- Lagnado, D. A., Gerstenberg, T., & Zultan, R. I. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, 37(6), 1036–1073. <https://doi.org/10.1111/cogs.12054>
- Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation. In A. Gopnik, & L. E. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154–172). Oxford University Press.
- Lewis, D. (1973). *Counterfactuals*. Blackwell Publishers.
- Little, D. (2010). *Current Issues in the Philosophy of Economics*. Colgate University.
- 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>
- Lombrozo, T., & Wilkenfeld, D. A. (2019). Mechanistic versus functional understanding. In S. R. Grimm (Ed.), *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology* (pp. 209–229). New York, NY: Oxford University Press.
- Martin, J. W., Buon, M., & Cushman, F. (2021). The effect of cognitive load on intent-based moral judgment. *Cognitive Science*, 45(4), Article e12965. <https://doi.org/10.1111/cogs.12965>
- Martin, J. W., & Cushman, F. (2016). The adaptive logic of moral luck. In *In, A Companion to Experimental Philosophy* (pp. 190–202). John Wiley & Sons, Ltd.. <https://doi.org/10.1002/9781118661666.ch12>
- McClure, J. L., Hilton, D. J., & Sutton, R. M. (2007). Judgments of intentional and physical causes in causal chains. *European Journal of Social Psychology*, 37, 879–901. <https://doi.org/10.1002/ejsp.394>
- Murray, D., & Lombrozo, T. (2017). Effects of manipulation on attributions of causation, free will, and moral responsibility. *Cognitive Science*, 41(2), 447–481. <https://doi.org/10.1111/cogs.12338>
- Norenzayan, A., & Nisbett, R. E. (2000). Culture and causal cognition. *Current Directions in Psychological Science*, 9(4), 132–135. <https://doi.org/10.1111/1467-8721.00077>
- O'Neill, K., Quillien, T., & Henne, P. (2022). A counterfactual model of causal judgment in double prevention. In *Conference in Computational Cognitive Neuroscience*.
- Pearl, J. (2018). Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv preprint*. arXiv:1801.04016.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.
- Rose, D., Sievers, E., & Nichols, S. (2021). Cause and burn. *Cognition*, 207(104517). <https://doi.org/10.1016/j.cognition.2020.104517>
- Schölkopf, B. (2022). Causality for machine learning. In H. Geffner, R. Dechter, & J. Halpern (Eds.), *Probabilistic and Causal Inference: The Works of Judea Pearl* (pp. 765–804). ACM Books.
- Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer Verlag.
- Sloman, S., Barbey, A. K., & Hotaling, J. M. (2009). A causal model theory of the meaning of cause, enable, and prevent. *Cognitive Science*, 33(1), 21–50. <https://doi.org/10.1111/j.1551-6709.2008.01002.x>
- Sosa, F. A., Ullman, T., Tenenbaum, J. B., Gershman, S. J., & Gerstenberg, T. (2021). Moral dynamics: Grounding moral judgment in intuitive physics and intuitive psychology. *Cognition*, 217, Article 104890. <https://doi.org/10.1016/j.cognition.2021.104890>
- Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General*, 126(4), 323–348. <https://doi.org/10.1037/0096-3445.126.4.323>
- Vasil, N., & Lombrozo, T. (2022). Explanations and causal judgments are differentially sensitive to covariation and mechanism information. *Frontiers in Psychology*, 13, 911177.
- 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>
- Wolff, P. (2014). Causal pluralism and force dynamics. In B. Copley, & F. Martin (Eds.), *Causation in grammatical structures* (pp. 100–119). <https://doi.org/10.1093/acprof:oso/9780199672073.001.0001>
- Wolff, P., & Barbey, A. K. (2015). Causal reasoning with forces. *Frontiers in Human Neuroscience*, 9, 1. <https://doi.org/10.3389/fnhum.2015.00322>

- Wolff, P., Barbey, A. K., & Hausknecht, M. (2010). For want of a nail: How absences cause events. *Journal of Experimental Psychology: General*, 139(2), 191–221. <https://doi.org/10.1037/a0018129>
- Wolff, P., Klettke, B., Ventura, T., & Song, G. (2005). Expressing causation in English and other languages. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside of the laboratory: Essays in honor of Douglas L. Medin* (pp. 29–48). Washington, DC: APA.
- Wu, S., Sridhar, S., & Gerstenberg, T. (2022). That was close! A counterfactual simulation model of causal judgments about decisions. In , 44, No. 44. *Proceedings of the Annual Meeting of the Cognitive Science Society*.