Cognition 190 (2019) 157–164

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

A counterfactual explanation for the action effect in causal judgment

Paul Henne[a,d,*], Laura Niemi[b], Ángel Pinillos[c], Felipe De Brigard[a,d], Joshua Knobe[e]

[a] Department of Philosophy, Duke University, United States
[b] Munk School of Global Affairs and Public Policy, University of Toronto, Canada
[c] School of Historical, Philosophical and Religious Studies, Arizona State University, United States
[d] Department of Philosophy and Center for Cognitive Neuroscience, Duke University, United States
[e] Program in Cognitive Science and Department of Philosophy, Yale University, United States

ARTICLE INFO

Keywords:
Action effect
Omissions
Omission effect
Causal reasoning
Counterfactual thinking
Causation by omission

ABSTRACT

People's causal judgments are susceptible to the action effect, whereby they judge actions to be more causal than inactions. We offer a new explanation for this effect, the *counterfactual explanation*: people judge actions to be more causal than inactions because they are more inclined to consider the counterfactual alternatives to actions than to consider counterfactual alternatives to inactions. Experiment 1a conceptually replicates the original action effect for causal judgments. Experiment 1b confirms a novel prediction of the new explanation, the *reverse action effect*, in which people judge inactions to be *more* causal than actions in overdetermination cases. Experiment 2 directly compares the two effects in joint-causation and overdetermination scenarios and conceptually replicates them with new scenarios. Taken together, these studies provide support for the new counterfactual explanation for the action effect in causal judgment.

1. Introduction

Suppose a company will send you a free sample of coffee if you are on their email list. Seeing that you are not subscribed, you change your subscription status by clicking a link, and later you receive the free sample. It seems reasonable to say that you received the free sample because you changed your subscription status. Now, consider a slightly different case: you are already subscribed to the email list, you decide not to change your subscription status, and then you receive the free sample. In this case, it seems less reasonable to claim that you received the free sample because you did not change your subscription status. In what we will refer to as *the action effect for causal judgment*, people consistently judge that actions like these are more causal than inactions (Cushman & Young, 2011; Feldman & Yay, 2018; Spranca, Minsk, & Baron, 1991; Walsh & Sloman, 2011; Willemsen & Reuter, 2016).

One explanation for the action effect is that people judge that actions, unlike inactions, have a direct physical connection to their effects and transfer force to their effects (for discussion, see Walsh & Sloman, 2011). We call this explanation the *generative explanation*. An alternative explanation for this difference relies on the view that counterfactual thinking affects causal reasoning (Byrne, 2016; Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum, 2017; Lewis, 1974; Mackie, 1974; Pearl, 2000; Roese & Olson, 1997). In line with this work, our new explanation suggests that people are more inclined to consider the counterfactuals of actions, relative to those of inactions, and that this propensity engenders the action effect. We call this explanation the *counterfactual explanation*.

To the best of our knowledge, there has been no direct comparison between these two explanations, and there is no evidence that favors one over the other. In fact, the standard action effect is consistent with both explanations (Cushman & Young, 2011; Walsh & Sloman, 2011). Accordingly, we offer a way to disambiguate them. Given recent work on counterfactuals and causal judgment (Icard, Kominsky, & Knobe, 2017), the counterfactual explanation makes a distinctive prediction in cases involving overdetermination, where a number of different causal factors are each individually sufficient but not individually necessary for the outcome. Specifically, the counterfactual explanation predicts that people should actually judge inactions to be *more* causal than actions in cases of overdetermination. By contrast, the generative explanation predicts that people should judge inactions to be less causal than actions in such cases—just as they do in non-overdetermination cases. In this article, we test these predictions and find evidence in favor of the counterfactual explanation.

* Corresponding author at: 201 West Duke Building, Duke University, Durham, NC 27708-0743, United States.
E-mail addresses: paul.henne@duke.edu (P. Henne), laura.niemi@utoronto.ca (L. Niemi), pinillos@asu.edu (Á. Pinillos), felipe.debrigard@duke.edu (F. De Brigard), joshua.knobe@yale.edu (J. Knobe).

https://doi.org/10.1016/j.cognition.2019.05.006
Received 8 October 2018; Received in revised form 25 April 2019; Accepted 6 May 2019

2. Explaining the action effect

Generally, the action effect refers to a phenomenon whereby, across a range of scenarios, people judge that agents who perform an action played a greater role in causing the outcome than agents who did nothing at all (Cushman & Young, 2011; Walsh & Sloman, 2011; Willemsen & Reuter, 2016; for a review, see Feldman & Yay, 2018). Many researchers have found this difference in causal judgment (Cushman & Young, 2011; Walsh & Sloman, 2011), and some have explored and suggested potential moderators (Willemsen & Reuter, 2016).

Much of the attention given to the action effect for causal judgment is due to its relationship to moral cognition. People consistently judge inactions, or omissions, to be less bad or less morally wrong than actions (Cushman, Young, & Hauser, 2006; Spranca et al., 1991), and this difference has been extensively studied both in moral judgment (Cushman & Young, 2011; Ritov & Baron, 1999) and in decision making (Anderson, 2003). Notably, some explanations for the difference in moral judgment identify the difference in causal judgment as a mediating factor (Baron & Ritov, 2004; Cushman & Young, 2011; Greene et al., 2009; Iliev, Sachdeva, & Medin, 2012; but see DeScioli, Bruening, & Kurzban, 2011).

In short, existing research suggests that the difference in moral judgment between actions and inactions may arise because there is a difference in the associated causal judgment. But this finding immediately leads to a new question: what explains the difference in causal judgment between actions and inactions?

2.1. The generative explanation

A class of theories, often called *process theories*, describe causation as a physical transfer of energy or force along a causal pathway (Dowe, 2000; 2004; Salmon, 1984; 1994; see also, Hall, 2004). On these accounts, causation is a continuous process along a pathway, or "world line" (Dowe, 1992; 1995; 2004; Salmon, 1994; 1997; 1998), and this process transmits a conserved quantity, like energy, along such a pathway (Dowe, 2004; Salmon, 1994). Causation, in short, is a physical, generative process. Such theories have a lot of intuitive backing; there is, for instance, an obvious generative process that occurs when you click your mouse, change your subscription status, and then receive a free sample of coffee.

Some recent work in cognitive science suggests that people reason about causation in a similar way. Relying on theories of force dynamics (Talmy, 1988), Wolff and colleagues suggest that causal reasoning involves the mental simulation of interacting entities and their perceived vectors (Wolff, 2007; Wolff, Barbey, & Hausknecht, 2010; see also Lombrozo, 2010; but see Walsh & Sloman, 2011). Regardless of the nuances of these different kinds of process theories, they can generally agree on a potential explanation for the action effect: force. In the action version of the coffee case, you physically click your mouse and change your subscription status. But in the inaction version, you do not change anything at all—you just leave your subscription status as it is. These views suggest a very natural explanation for the action effect: actions generate the outcome through some kind of transfer of force, energy, or *oomph*, while inactions do not transfer force or energy to the outcome at all. We call this explanation for the action effect the *generative explanation*.

2.2. The counterfactual explanation

An alternative explanation draws on views according to which causal judgment relies on counterfactual reasoning (Byrne, 2016; Gerstenberg et al., 2017; Lewis, 1974; Mackie, 1974; Pearl, 2000; Roese & Olson, 1997). On such counterfactual accounts, people's causal judgments depend on the consideration of counterfactual alternatives. That is, when people reason about whether some event caused an outcome, they do so by considering a counterfactual alternative and asking whether the outcome would still occur. Suppose, for example, that you subscribe to the company's email list, and then you receive a free sample. When deciding whether subscribing to the email list caused you to receive the free sample, people consider the counterfactual alternative where you *do not* subscribe to the email list and judge whether you would still receive the free sample. In this case, not subscribing to the email list would make a difference—you would not receive the free sample in this counterfactual alternative—so your subscribing, in fact, caused you to receive the free sample.

To see how counterfactual theories might explain the action effect, it will be helpful to focus on one specific aspect of such theories: people do not think equally about all counterfactuals. Instead, they tend to focus on certain counterfactuals and ignore others (Byrne, 2016; Kahneman & Miller, 1986; Phillips, Luguri, & Knobe, 2015; Kahneman & Tversky, 1982). Research has identified a number of different, specific factors that influence which counterfactuals people consider. For example, people are more likely to consider counterfactuals in which statistically infrequent events in the actual world are replaced by frequent events (Byrne, 2016; Kahneman & Miller, 1986; Kahneman & Tversky, 1982), and they are more likely to consider counterfactuals in which morally bad actions are replaced by morally good actions (Byrne, 2016; McCloy & Byrne, 2000; Phillips et al., 2015).

Intriguingly, these very same patterns appear in people's causal judgments (see Halpern & Hitchcock, 2014; Henne, Pinillos, & De Brigard, 2017; Hitchcock & Knobe, 2009). That is, existing research consistently finds a tendency to pick out as causes events that are statistically infrequent (Hart & Honoré, 1985; Hilton & Slugoski, 1986) or morally bad (Knobe & Fraser, 2008; Phillips et al., 2015). Counterfactual theories of causal judgment can explain these effects in terms of people's tendencies to consider the corresponding counterfactuals. For example, if you strike a match and a fire starts, people judge that the fire was caused by the lighting of a match (infrequent) rather than the presence of oxygen in the atmosphere (frequent). On the counterfactual account, this is explained by the fact that people are more inclined to consider the counterfactual alternative to lighting a match—i.e., not lighting the match—than the alternative to the presence of oxygen in the atmosphere—i.e., the absence of oxygen in the atmosphere.

One intriguing hypothesis would be that the action effect can now be explained in precisely the same way. Just as existing research suggests that people are more inclined to consider counterfactuals in which statistically infrequent events in the actual world are replaced by frequent events, existing research suggests that people tend to be especially drawn to consider counterfactuals in which an action is replaced by an inaction rather than counterfactuals in which an inaction is replaced by an action (Byrne & McEleney, 2000; Kahneman & Tversky, 1982; see also Byrne, 2016). This fact about counterfactual thinking then might engender the action effect that we find in people's causal judgments. Consider again that it seems more reasonable to claim that you received the free sample because you changed your subscription status (the action case) than it does to claim that you received it because you did not change your subscription status when you were already subscribed (the inaction case). This difference could be explained by people's greater tendency to consider the counterfactual where you *do not* change your subscription status, in the action case, than the counterfactual where you *do* change your subscription status, in the inaction case. We will refer to this second explanation as the *counterfactual explanation*.

2.3. Testing the explanations

So far, we have two explanations for the action effect for causal judgment: the generative explanation and the counterfactual explanation. These explanations give us two distinct, testable hypotheses. The generative hypothesis is that people consider actions to be more causal than inactions because they believe that actions impart a force to the

outcome, whereas inactions do not involve a transfer of force. The counterfactual hypothesis is that people judge actions to be more causal than inactions because of a difference in their tendencies to consider the corresponding counterfactuals. The generative hypothesis and the counterfactual hypothesis are not easy to tease apart experimentally, and existing work does not favor either one.

Despite these limitations, recent empirical and theoretical work provides a viable way to test differential predictions that follow from these hypotheses. Specifically, the two explanations yield different predictions in cases of *overdetermination*—i.e., in cases where multiple causal factors are each individually sufficient but not individually necessary for the outcome. As an example, suppose that you will receive the free sample if you have previously purchased coffee beans from the company or if you have subscribed to the email list (inclusive 'or'). Now suppose that you have previously purchased coffee beans from the company and that you have subscribed to the email list. The outcome (receiving the free sample) would then be overdetermined: each action would be sufficient for the outcome just by itself, and each action is not individually necessary. So, if either action occurs but the other does not, the outcome would still occur.

Interestingly, existing research suggests that the impact of both statistical and moral considerations is *reversed* in these cases (Icard et al., 2017; Kirfel & Lagnado, 2018; Morris, Phillips, Gerstenberg, & Cushman, 2019). Although people tend to pick out infrequent and morally bad events as causes in most ordinary cases, their judgments in overdetermination cases show the opposite pattern. In other words, in overdetermination cases, people tend to select frequent and morally permissible events as causes.

This effect is easily explained by counterfactual theories of causal judgment, and it is specifically predicted by one recent computational theory of the impact of counterfactuals on causal judgment (Icard et al., 2017). This theory predicts that in overdetermination cases, people should be more inclined to regard an event as causal when they consider counterfactuals in which that event did occur than when they consider counterfactuals in which that event did not occur (see the General Discussion for a brief, non-technical summary).

These recent developments give us an opportunity to better assess the two explanations for the action effect. On the counterfactual explanation, it is the tendency to consider counterfactuals that engenders the difference in causal judgments for actions and inactions. Hence, the counterfactual explanation predicts a *reverse action effect*: in overdetermination cases, inactions should actually be regarded as more causal than actions. By contrast, the generative explanation does not predict that inactions should be regarded as more causal than actions in overdetermination cases. If people judge inactions to be less causal than actions because they do not transfer force, then they should always judge that inactions to be less causal, regardless of the causal structure. So, on this hypothesis, even in cases of overdetermination, people should judge inactions to be less causal than actions.

2.4. The present studies

Prior to conducting this research, we had no settled opinion as between the generative explanation and the counterfactual explanation; hence, we had no specific prediction as to whether the reverse action effect would emerge. To investigate these two explanations, we conducted three studies. Experiments 1a and 1b asks whether the action effect is indeed reversed in overdetermination cases. Experiment 2 conceptually replicates these effects with new materials and directly compares them in joint-causation and overdetermination scenarios.

3. Experiment 1a

To ensure that our vignettes showed the original action effect for causal judgments, we developed three vignettes involving ordinary causal scenarios that were not morally or emotionally salient and that could be transformed into overdetermination cases.

3.1. Methods

3.1.1. Participants

Because there is no standard method for calculating the sample size for mixed models, we calculated the sample size for a single vignette as it would be analyzed with an ANOVA. At 80% power with a small-medium effect size ($f = 0.17$), a sample size of 274 per vignette (822 total) was required. We aimed for the same sample size in all experiments that follow. A total of 826 subjects completed the survey on Amazon Mechanical Turk (AMT). 20 participants reported not paying attention, so they were excluded. Data were analyzed from the remaining 806 participants ($M_{age} = 35.60$, $SD = 11.20$, $\text{Range}_{age} = [18–80]$, 51.73% female). After completing the survey, participants were compensated $0.25.

3.1.2. Materials and procedure

Participants were randomly assigned to 1 of 6 conditions in a 3 (Vignette: Motorboat, Guitar, Dryer) × 2 (Event Type: Action or Inaction) between-subjects design (vignettes provided in Supplement A). Each participant read a vignette and answered the causal question (see sample in Table 1). They were asked for their level of agreement with the causal statement on a 1–7 scale [1 = strongly disagree, 4 = neutral, 7 = strongly agree]. Participants were then asked for basic demographic information. One explicit attention check was used in all experiments (Supplement C). Data collection was completed in all experiments in this manuscript prior to any analysis by the authors. All materials, data, and analyses scripts for all experiments are available at https://osf.io/gk9dj/.

3.2. Results

Data were analyzed using R with the lme4 software package (Bates, Maechler, Bolker, & Walker, 2015) and the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Data were fit to linear

Table 1
Sample vignette from Experiment 1a (Motorboat) in both the action and inaction conditions and the dependent variable, a measure of causal judgment.

Action	Inaction
Ned has a new motorboat. When Ned turns the key, the motorboat starts if the motor is in the lock position. Today, the motor is not in the lock position. Ned checks the motor to see if it is in the lock position. He sees that it is not in the lock position. So, he changes its position, and he puts it in the lock position. Because the motorboat would start if the motor is in the lock position, the motorboat starts when Ned turns the key. To what extent do you agree with the following statement about the passage you just read? The motorboat started because Ned changed the position of the motor.	Ned has a new motorboat. When Ned turns the key, the motorboat starts if the motor is in the lock position. Today, the motor is in the lock position. Ned checks the motor to see if it is in the lock position. He sees that it is in the lock position. So, he does not change its position, and he leaves it in the lock position. Because the motorboat would start if the motor is in the lock position, the motorboat starts when Ned turns the key. The motorboat started because Ned did not change the position of the motor.

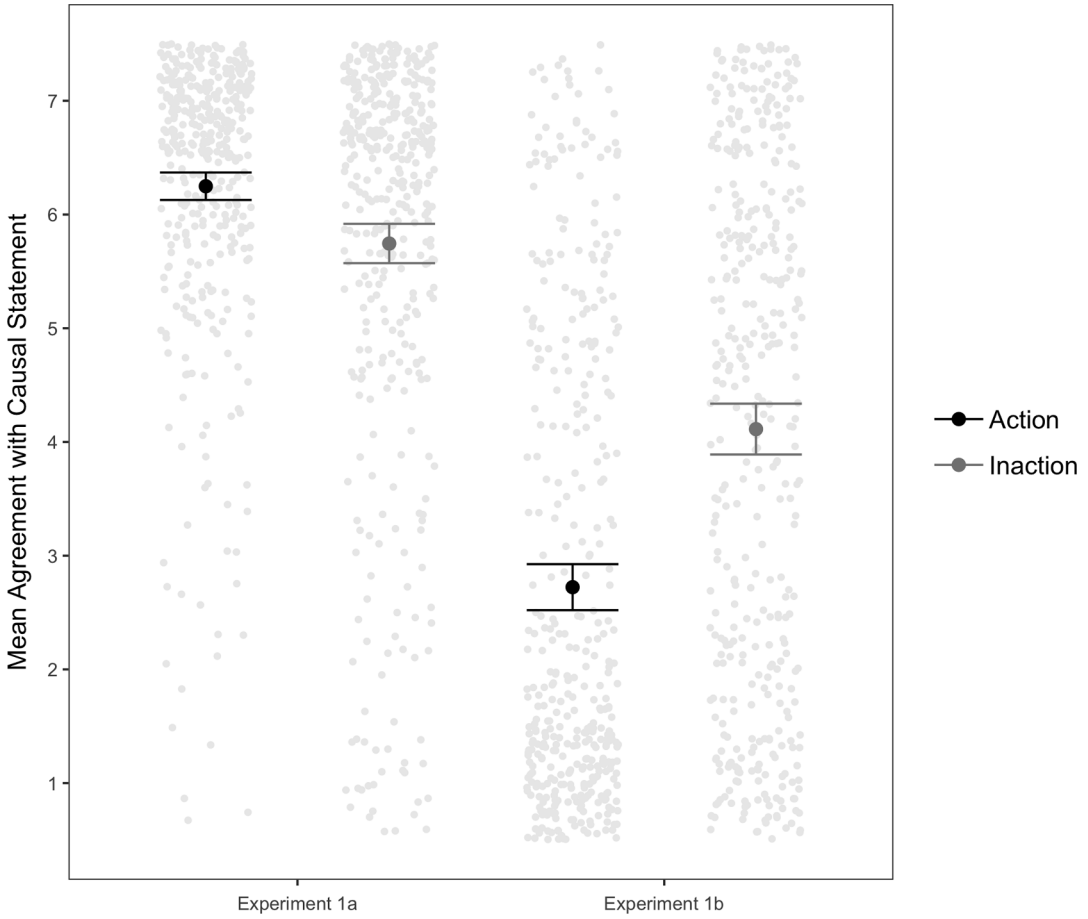

Fig. 1. Mean agreement with the causal statement in Experiment 1a and 1b collapsed across vignettes. Error bars indicate 95% confidence intervals. Lighter grey points represent individual data points evenly jittered.

mixed-effects models, and vignette was included as a random effect (random intercepts only) in all models. Significance for fixed effects was assessed via Satterthwaite's degrees of freedom method.

To test whether inactions were less likely to be judged as causes, we examined the difference in agreement with the casual statement as a function of event type across vignettes. In line with the standard action effect, participants judged actions (M = 6.25, SD = 1.23, n = 398) to be more causal than inactions (M = 5.75, SD = 1.78, n = 408) (b = -0.50, SE = 0.10, t = -4.70, $p < .001$, CI [-0.71, -0.29]) (Fig. 1). Mean agreement with the causal statement for each vignette is reported in Table 3 (Supplement D).

3.3. Discussion

Using three new vignettes, we conceptually replicated the original action effect—i.e. people are less likely to judge that inactions, relative to actions, are the cause of the outcome—found in previous research (Walsh & Sloman, 2011; Cushman & Young, 2011). This result was predicted by both the generative explanation and the counterfactual explanation.

4. Experiment 1b

In order to distinguish these explanations, we modified the vignettes from Experiment 1a to make them into overdetermination cases. If the generative explanation for the action effect is correct, then people, regardless of the shift in causal structure, should continue to be less likely to judge that inactions, relative to actions, are causes. Crucially, the counterfactual explanation makes the opposite prediction. Given the findings from Icard et al. (2017) and the predictions of the counterfactual explanation, inactions should be judged to be more causal than actions in overdetermination cases.

4.1. Methods

4.1.1. Participants

A total of 834 subjects completed the survey on AMT. 28 participants reported not paying attention, so they were excluded. Data were analyzed from the remaining 806 participants (M_{age} = 35.30, SD = 11.50, Range_{age} = [18–80], 49.25% female). After completing the survey, participants were compensated $0.25.

4.1.2. Materials and procedure

Participants were randomly assigned to 1 of 6 conditions in a 3 (Vignette: Motorboat, Guitar, Dryer) × 2 (Event Type: Action or Inaction) between-subjects design (Vignettes provided in Supplement B). Vignettes from Experiment 1 were modified such that alternative actions that are also sufficient for the outcome were added with disjunctives "either" and "or." Dependent variables were identical to those used in Experiment 1a. Each participant read a vignette like the sample in Table 2.

4.2. Results

We conducted the same analysis that was used in Experiment 1a. To test whether inactions were more likely to be judged as causes in the overdetermination cases, we examined the difference in agreement with the causal statement as a function of event type across vignettes.

Table 2
Sample vignette from Experiment 1b (Motorboat) in both the action and inaction conditions and the dependent variable, a measure of causal judgment.

Action	Inaction
Ned has a new motorboat. When Ned turns the key, the motorboat starts if *either* the gear is in neutral *or* the motor is in the lock position. Today, the gear is in neutral, and the motor is not in the lock position. Ned checks the motor to see if it is in the lock position. He changes its position, and he puts it in the lock position. Because the motorboat would start if *either* the gear is in neutral *or* the motor is in the lock position, the motorboat starts when Ned turns the key	Ned has a new motorboat. When Ned turns the key, the motorboat starts if *either* the gear is in neutral *or* the motor is in the lock position. Today, the gear is in neutral, and the motor is in the lock position. Ned checks the motor to see if it is in the lock position. He sees that it is in the lock position, and he does not change its position at all. Because the motorboat would start if *either* the gear is in neutral *or* the motor is in the lock position, the motorboat starts when Ned turns the key
To what extent do you agree with the following statement about the passage you just read?	
The motorboat started because Ned changed the position of the motor	The motorboat started because Ned did not change the position of the motor

Participants actually judged inactions (M = 4.11, SD = 2.27, n = 398) to be *more* causal than the actions (M = 2.72, SD = 2.08, n = 408) (b = 1.39, SE = 0.15, t = 9.16, p < .001, CI [1.09, 1.68]) (Fig. 1). Mean agreement with the causal statement for each vignette is reported in Table 3 (Supplement D).

4.3. Discussion

In Experiment 1b, we found the reverse action effect: in cases of overdetermination, people were more likely to agree that inactions, relative to actions, are causes. These results are consistent with the counterfactual explanation for the action effect. The generative explanation, on the other hand, is inconsistent with the reverse action effect found in Experiment 1b; the generative explanation predicts that people should always judge inactions to be less causal than actions.

Despite our finding these two effects, there is a potential worry about our ability to compare them across experiments; in Experiment 1a, the vignettes describe only a single potential causal factor, while in Experiment 1b the vignettes describe two potential causal factors. In order to directly compare the effects—the action effect and the reverse action effect—we must match the number of potential causal factors in each contrasting scenario. Such a direct comparison will assuage potential worries about the number of potential causal factors—rather than frequency at which people consider the counterfactuals—producing the two effects. Hence, in Experiment 2, we directly compare these two effects by developing three new scenarios with the same number of potential causal factors. We also conceptually replicate the two effects with new vignettes.

5. Experiment 2

This experiment had two aims. First, we wanted to compare the action effect and the reverse action effect by investigating an interaction between causal structure and event type. Second, we wanted to conceptually replicate our findings in Experiment 1a and 1b with additional materials.

5.1. Methods

5.1.1. Participants

Our expected sample size was the same as calculated in Experiment 1, so we aimed to recruit 274 participants for each vignette per structure (1644 total). Expecting participant exclusions at a rate of 4%, we aimed for a sample size of 1710 participants. AMT workers were required to have a hit rate approval of greater than 90%, to have completed more than 50 hits on AMT, and be located in the United States. 1735 such participants completed more than 90% of the survey. 40 participants reported not paying attention or did not respond to the attention check, so they were excluded. Data were analyzed from the remaining 1695 participants (M_{age} = 38.50, SD = 12.80, $Range_{age}$ = [18–82], 51.50% female). After completing the survey, participants were compensated $0.35.

5.1.2. Materials and procedure

Participants were randomly assigned to 1 of 12 conditions in a 3 (Vignette: Implosion, Watch, Coffee) × 2 (Event Type: Action or Inaction) × 2 (Structure: Joint Causation or Overdetermination) between-subjects design (Vignettes provided in Supplement E). In joint-causation conditions, two events were described as jointly sufficient and individually necessary for the outcome to occur. This relationship was specified with the conjunctives “both” and “and.” The structure of the vignettes in the overdetermination conditions were similar to those used in Experiment 1b. In both structures, the two potential causal factors always occurred, as did the outcome. Each participant read a single vignette and answered the causal question just as they did in Experiment 1a and 1b. In order to ensure that participants understood what was necessary for the outcome to occur, participants also responded to a comprehension check, which was displayed on a its own page after the causal question. Each comprehension check was unique to each of the vignettes and causal structures (Supplement F). We planned to analyze the data including all participants and then re-analyze it excluding the participants who failed the comprehension check. Participants were then asked for basic demographic information, and they were asked to respond to the same explicit attention check that was used in Experiments 1a and 1b. A time-stamped document containing the methods, materials, and a plan for data collection, participant exclusion, and analysis was uploaded to OSF before we collected data, and it is available at https://osf.io/tkzgb/.

5.2. Results

First, we tested whether there was a significant interaction between structure and event type. To do so, we used a LMER to examine the difference in agreement with causal statement as a function of structure, event type, and their interaction across vignettes with vignette included as a random intercept in the model. The LMER was a singular fit because of an estimate of 0 variance for the intercept, suggesting that the model does not warrant a random effect of vignette. Hence, we simplified the model, using a linear model with no random effects. There was a significant interaction between causal structure and event type (b = 1.97, SE = 0.19, t = 10.20, p < .001, CI [1.59, 2.35]) (Fig. 2).

Second, we tested whether there was a main effect for event type in each causal structure. So, we ran post-hoc pairwise comparisons between event types for each causal structure (Tukey corrected). In line with the traditional action effect, participants in the joint-causation conditions agreed that actions (M = 6.25, SD = 1.31, n = 429) are more causal than the inactions (M = 5.51, SD = 1.80, n = 422) (b = 0.74, SE = 0.13, t = 5.43, p < .001, CI [0.39, 1.09]). In line with the reverse action effect, participants in the overdetermination conditions agreed that inactions (M = 4.50, SD = 2.31, n = 418) are more causal than the actions (M = 3.26, SD = 2.35, n = 426) (b = -1.23,

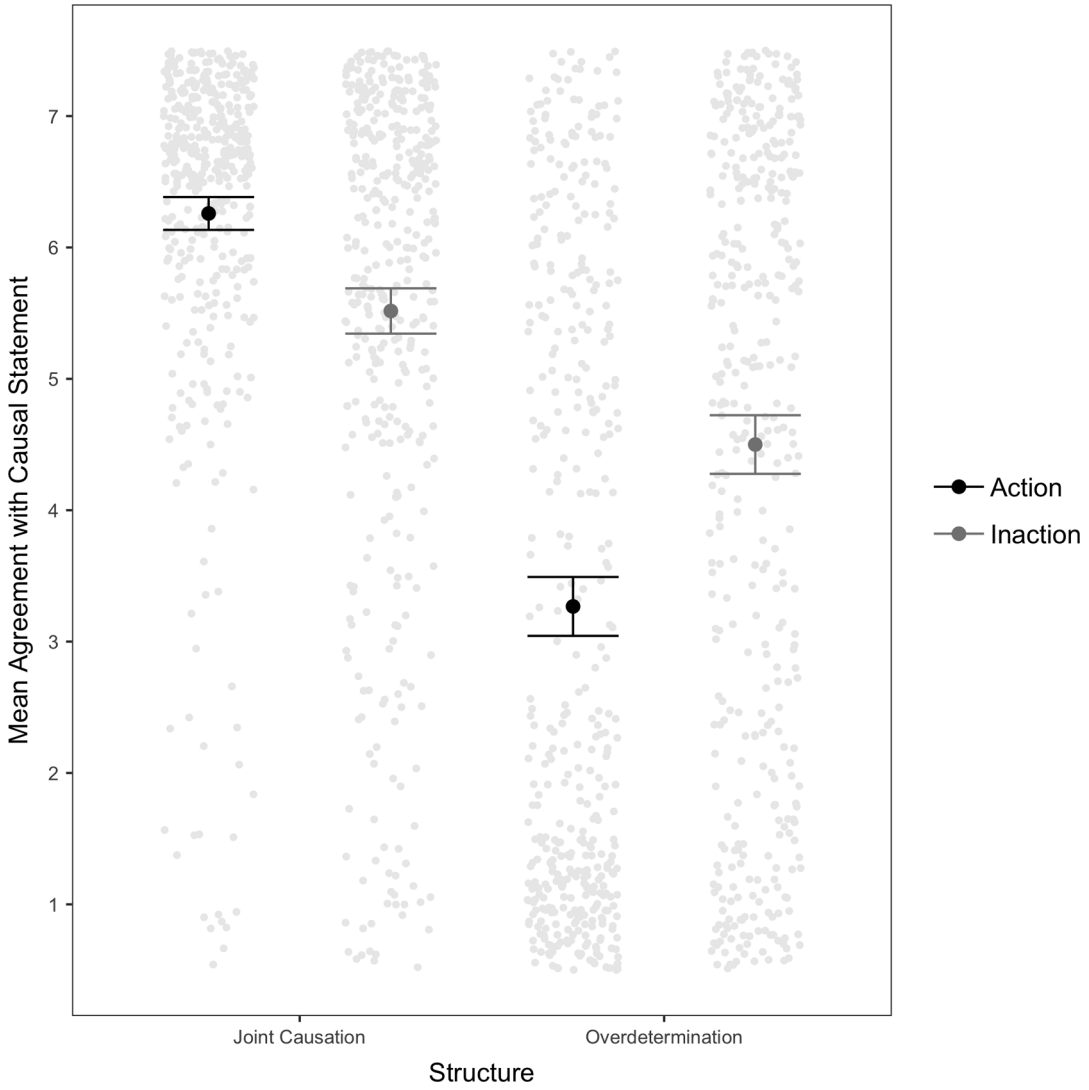

Fig. 2. Mean agreement with the causal statement in Experiment 2 collapsed across vignettes. Error bars indicate 95% confidence intervals. Lighter grey points represent individual data points evenly jittered.

$SE = 0.13$, $t = -8.99$, $p < .001$, CI $[-1.58, -0.87]$). Mean agreement with the causal statement for each vignette is reported in Table 4 (Supplement D). Of those participants included in this analysis, 92.09% appropriately responded "Yes" to the comprehension check. Excluding those participants who failed the comprehension checks made no difference to the significance of the results (Supplement G).

6. General discussion

Existing work has found an action effect for causal judgment, where people judge inactions to be less causal than actions when matched for outcome (Cushman & Young, 2011; Walsh & Sloman, 2011). We set out to contrast two explanations for this effect: the generative explanation and the counterfactual explanation. Experiments 1a and 2 conceptually replicate the action effect, showing a similar effect to that found in previous work. This finding is predicted by both explanations. Experiments 1b and 2 demonstrate a new effect for causal judgment, the *reverse action effect*, where people are more likely to agree that inactions, relative to actions, are the cause of the outcome in cases of overdetermination. This effect is predicted by the counterfactual explanation but not by the generative explanation. Experiment 2 directly compares these effects in causal judgment. Taken together, these studies bolster the counterfactual explanation for the action effect in causal judgment.

The generative explanation for the action effect does not predict these new findings. On this explanation, people should (*ceteris paribus*) consistently judge that inactions are *less* causal than actions when they are matched for outcome. Our studies do not confirm this prediction. Nonetheless, it may still be the case that perception of force and force relations accounts for some of important patterns in people's causal judgments. As Lombrozo (2010) has emphasized, it is possible that some causal judgments are driven by counterfactual reasoning while others are driven by reasoning about generative connections.

6.1. Inaction and counterfactual thinking

The present findings extend a pattern that has been observed in the existing literature on causal cognition. In cases of joint causation, people are more inclined to regard an event as causal to the extent that it is infrequent rather than frequent (e.g., Hilton & Slugoski, 1986), morally bad rather than morally good (e.g., Knobe & Fraser, 2008), and an action rather than an inaction (e.g., Walsh & Sloman, 2011). Previous studies found that the impact of the frequent-infrequent distinction and of the good-bad distinction are reversed in cases of overdetermination (Icard et al., 2017). The present studies find the same reversal for the inaction-action distinction.

Importantly, all three of these effects are mirrored in the existing literature on counterfactuals. That is, people are more inclined to consider counterfactuals that involve replacing infrequent events with frequent events, morally bad events with morally good events, and actions with inactions (see Byrne, 2016; Roese, 1997). Thus, the entire pattern would be straightforwardly explained on any counterfactual theory according to which (a) in the cases of joint causation, people are more inclined to regard a potential causal factor as causal to the extent that they consider counterfactuals in which the factor *does not* take place but also (b) in the cases of overdetermination, people are more inclined to regard a potential causal factor as causal to the extent that they consider counterfactuals in which the factor *does* take place.

One such theory is proposed in a recent paper by Icard et al. (2017). Stripped of its mathematical details, the core idea is that in trying to determine whether something is a cause, people tend to ask different questions depending on which counterfactual they consider. Suppose a person is wondering whether some event caused an outcome. If she considers counterfactuals in which the event did not occur, she will tend to ask whether the event was *necessary* for the outcome. By contrast, if she considers counterfactuals in which the event did occur (and various other background conditions were altered), she will ask whether the event was *sufficient* for the outcome. The impact of counterfactual thinking on causal judgments about a specific event, therefore, depends on the degree to which this event was necessary or sufficient. If the event is necessary but not sufficient, it will be regarded as *more* causal when people are *more* inclined to consider counterfactuals in which it did not occur. In overdetermination cases, however, an event is seen as sufficient but not necessary for the outcome. The model predicts that, in such cases, an event will actually be seen as *more* causal when people are *less* inclined to consider counterfactuals in which it did not occur.

Although this theory correctly predicts the effects observed in the present studies, we hasten to add that it might be possible to develop or modify other counterfactual theories to accommodate these effects. In particular, on Icard et al. (2017) theory, people make use of a non-conscious process that samples counterfactuals for causal judgment, but there might also be other ways of spelling out the precise details within a broadly counterfactual approach. For example, another possible view would be that people explicitly represent certain counterfactuals as being simply more 'relevant' than others (Phillips et al., 2015). People might then regard counterfactuals as more relevant to the extent that they involve events that are frequent, morally good or inactions, and it might be these judgments of the relevance of counterfactuals that impact their causal judgments and yield these effects.

We initially found this counterfactual relevance hypothesis to be a plausible explanation for our effects and even obtained some preliminary evidence in favor of it (see Supplement H), but in a replication study, we found evidence that did not support it (see Supplement I). Specifically, using the Phillips and Kominsky (2018) method for measuring counterfactual relevance judgments and our materials from Experiment 1, we asked whether the impact of the action-inaction distinction on causal judgments was mediated by perceived counterfactual relevance. While we replicated the action effect and the reverse action effect, we found no significant mediation using the materials from Experiment 1a and, indeed, no impact of the action-inaction distinction on the counterfactual relevance measure. Thus, it may be that explicit judgments of counterfactual relevance explain certain aspects of people's causal judgments, but we do not have evidence that they explain the action effect for causal judgment.

A third possible way of spelling out the details would be in terms of people's representations of the probabilities of different counterfactuals. For example, on the *counterfactual potency* theory, people determine causal contribution from the product of the probability of the antecedent of the counterfactual and the probability of the consequent given the antecedent (Petrocelli, Percy, Sherman, & Tormala, 2011). As recent research has shown (Gilbert, Tenney, Holland, & Spellman, 2015), this ends up being strikingly similar to Spellman and colleagues' view according to which people compare the probability of the outcome before an event to the probability after that event (Spellman, Kincannon, & Stose, 2005). In their present versions, these theories do not predict a reversal in overdetermination cases, but perhaps it would be possible to develop modified versions of such theories that do predict this reversal.

Finally, it is possible, at least in principle, that the impacts on causal judgment of the frequent-infrequent distinction, the good-bad distinction, and the inaction-action distinction have nothing to do with counterfactuals and are instead best explained by some completely different approach to causal cognition. A defender of such an approach would have to explain why all three of these distinctions also impact counterfactual thinking, and we know of no attempt to provide such an explanation within the existing literature. Nonetheless, this is certainly a real possibility, which may be worthy of further exploration.

6.2. Moral cognition

As we mentioned in the introduction, much of the interest in the action effect for causal judgment stems from work on the omission effect for *moral* judgments (Cushman & Young, 2011). Numerous studies find that actions are regarded as more morally wrong than omissions or inactions, even when the consequences are exactly the same (Baron & Ritov, 1994; Baron & Ritov, 2009; Cushman et al., 2006; Feldman & Yay, 2018), and this "omission bias" affects many morally relevant behaviors (DeScioli et al., 2011; Ritov & Baron, 1994; Ritov & Baron, 1995; Royzman & Baron, 2002). So, why might people typically judge actions to be more morally wrong? Some evidence suggests that this pattern in people's moral judgments—or their judgments of blame in particular (Bostyn & Roets, 2016)—might be explained by the corresponding pattern in people's causal judgment (Cushman & Young, 2011; Siegel, Crockett, & Dolan, 2017). On this view, the reason people see the action as more morally wrong than the inaction is that they make different causal judgments in the two cases. When an agent performs an action, people are more inclined to regard the agent as the cause of the bad consequences that ensue. Hence, the agent is also more morally culpable.

If this view turns out to be correct, the present studies could shed light, at least indirectly, on questions in moral cognition. One common view is that the action effect for causal judgment is driven in some way by the fact that actions physically produce an outcome or transfer some force, while inactions do not transfer force to the outcome at all. If this assumption is correct, and if the action effect for moral judgment is driven by the action effect for causal judgment, then it seems that the action effect for moral judgment ultimately comes down to the fact that inactions lack the right sort of physical connection to outcomes (e.g., Greene et al., 2009; Iliev et al., 2012).

The current studies point to a different explanation. These studies suggest that the action effect for causal judgment is explained by facts about people's counterfactual reasoning. Thus, if the action effect for moral judgment is driven by the effect for causal judgment, these studies suggest that the effect for moral judgment is at least in part explained by which counterfactuals people consider. On this kind of view, the effect for moral judgment arises because when an agent performs a harmful action, people are naturally inclined to consider the counterfactual in which she refrains from performing this action, but when an agent omits to perform a helpful action, people are not nearly as inclined to consider the counterfactual in which she chooses to perform the helpful action. Future work could more directly explore the relationship between our new results for causal judgment and people's moral judgments. For example, we would expect that the action effect in moral cognition could be reversed in cases of overdetermination.

6.3. Conclusion

In this paper, we provide novel evidence for the counterfactual explanation for the action effect for causal judgment. We also show a surprising, novel effect on causal judgment, the reverse action effect. This explanation and the new effect deserve future investigation, for the results could be used to advance some areas of moral psychology, economic behavior, causal decision theory, and legal theory.

Acknowledgments

We thank the IMC lab and the MAD lab at Duke University. We also thank Jonathan Phillips, Matthew Stanley, Pascale Willemsen, Thomas Icard, Audrey Liu, Sara Bernstein, Paul Bello, Walter Sinnott-Armstrong, and Shenyang Huang for helpful feedback on different parts of the design, analysis, drafts, and the overall project. We are also thankful for audiences at Notre Dame's Department of Philosophy and at SPP 2018. This project was supported by an Office of Naval Research award (N00014-17-1-2603) to FDB.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.05.006.

References

Anderson, C. J. (2003). The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin, 129*(1), 139.
Baron, J., & Ritov, I. (1994). Reference points and omission bias. *Organizational Behavior and Human Decision Processes, 59*(3), 475–498.
Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*(2), 74–85.
Baron, J., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. *Psychology of Learning and Motivation, 50*, 133–167.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effect models using lme4. *Journal of Statistical Software, 67*, 1–48.
Bostyn, D. H., & Roets, A. (2016). The morality of action: The asymmetry between judgments of praise and blame in the action–omission effect. *Journal of Experimental Social Psychology, 63*, 19–25.
Byrne, R. M. (2016). Counterfactual thought. *Annual Review of Psychology, 67*, 135–157.
Byrne, R. M., & McEleney, A. (2000). Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 1318–1331.
Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science, 35*(6), 1052–1075.
Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*(12), 1082–1089.
DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behavior, 32*(3), 204–215.
Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science, 59*, 195–216.
Dowe, P. (1995). Causality and conserved quantities: A reply to salmon. *Philosophy of Science, 62*, 321–333.
Dowe, P. (2000). *Physical Causation.* Cambridge University Press.
Dowe, P. (2004). Causation and misconnections. *Philosophy of Science, 71*(5), 926–931.
Feldman, G., & Yay, T. (2018). Action-inaction asymmetries in moral scenarios: Replication of the omission bias examining morality and blame with extensions linking to causality, intent, and regret. Preprint. http://doi.org/10.13140/RG.2.2.10240.74242.
Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological Science, 28*(12), 1731–1744.
Gilbert, E. A., Tenney, E. R., Holland, C. R., & Spellman, B. A. (2015). Counterfactuals, control, and causation: Why knowledgeable people get blamed more. *Personality and Social Psychology Bulletin, 41*(5), 643–658.
Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*(3), 364–371.
Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. A. Paul (Eds.). *Causation and Counterfactuals* (pp. 225–276). MIT Press.
Halpern, J. Y., & Hitchcock, C. (2014). Graded causation and defaults. *The British Journal for the Philosophy of Science, 66*(2), 413–457.
Hart, H. L. A., & Honoré, T. (1985). *Causation in the Law.* OUP Oxford.
Henne, P., Pinillos, Á., & De Brigard, F. (2017). Cause by omission and norm: Not watering plants. *Australasian Journal of Philosophy, 95*(2), 270–283.
Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review, 93*(1), 75.
Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy, 106*(11), 587–612.
Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition, 161*, 80–93.
Iliev, R. I., Sachdeva, S., & Medin, D. L. (2012). Moral kinematics: The role of physical factors in moral judgments. *Memory & Cognition, 40*(8), 1387–1401.
Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136.
Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, & A. Tversky (Eds.). *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.
Kirfel, L., & Lagnado, D. (2018). Statistical norm effects in causal cognition. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.). *Proceedings of the 40th Annual Conference of the Cognitive Science Society*. Madison, WI: Cognitive Science Society.
Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. *Moral Psychology, 2*, 441–448.
Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26.
Lewis, D. (1974). Causation. *The Journal of Philosophy, 70*(17), 556–567.
Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology, 61*, 303–332.
Mackie, J. L. (1974). *The cement of the universe.* Oxford: Oxford University Press.
McCloy, R., & Byrne, R. (2000). Counterfactual thinking about controllable events. *Memory and Cognition, 28*, 1071–1078.
Morris, A., Phillips, J. S., Gerstenberg, T., & Cushman, F. (2019). *Quantitative causal selection patterns in token causation. Unpublished manuscript.* Harvard University.
Pearl, J. (2000). *Causality: Models, reasoning and inference.* Cambridge, MA: Cambridge University Press.
Petrocelli, J. V., Percy, E. J., Sherman, S. J., & Tormala, Z. L. (2011). Counterfactual potency. *Journal of Personality and Social Psychology, 100*(1), 30.
Phillips, J. S., & Kominsky, J. F. (2018, July 24). Immoral professors and malfunctioning tools: Counterfactual relevance accounts explain the effect of norm violations on causal selection. https://doi.org/10.31219/osf.io/u83vk.
Phillips, J., Luguri, J. B., & Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition, 145*, 30–42.
Ritov, I., & Baron, J. (1994). Judgements of compensation for misfortune: The role of expectation. *European Journal of Social Psychology, 24*(5), 525–539.
Ritov, I., & Baron, J. (1995). Outcome knowledge, regret, and omission bias. *Organizational Behavior and Human Decision Processes, 64*(2), 119–127.
Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes, 79*(2), 79–94.
Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*(1), 133–148.
Roese, N. J., & Olson, J. M. (1997). Counterfactual thinking: The intersection of affect and function. *Advances in Experimental Social Psychology, 29*, 1–59.
Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research, 15*(2), 165–184.
Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton University Press.
Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science, 61*(2), 297–312.
Salmon, W. C. (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science, 64*, 461–477.
Salmon, W. (1998). *Causality and Explanation.* New York: Oxford University Press.
Siegel, J. Z., Crockett, M. J., & Dolan, R. J. (2017). Inferences about moral character moderate the impact of consequences on blame and praise. *Cognition, 167*, 201–211.
Spellman, B. A., Kincannon, A., & Stose, S. (2005). The relation between counterfactual and causal reasoning. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.). *The Psychology of Counterfactual Thinking* (pp. 28–43). London: Routledge Research.
Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27*(1), 76–105.
Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*(1), 49–100.
Walsh, C. R., & Sloman, S. A. (2011). The meaning of cause and prevent: The role of causal mechanism. *Mind & Language, 26*(1), 21–52.
Willemsen, P., & Reuter, K. (2016). Is there really an omission effect? *Philosophical Psychology, 29*(8), 1142–1159.
Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136*(1), 82.
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.