

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Causation and norms of proper functioning: Counterfactuals are (still) relevant

Permalink

<https://escholarship.org/uc/item/26s0x7qd>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors

Phillips, Jonathan

Kominksy, Jonathan

Publication Date

2017

Peer reviewed

Causation and norms of proper functioning: Counterfactuals are (still) relevant

Jonathan Phillips (phillips01@g.harvard.edu) & Jonathan Kominsky (jkominsky@g.harvard.edu)

Department of Psychology, Harvard University, 33 Kirkland St. Cambridge, MA, 02138

Abstract

Causal judgments are well-known to be sensitive to violations of both prescriptive moral and descriptive statistical norms. There is ongoing discussion as to whether both effects are best explained through changes in the relevance of counterfactual possibilities, or if moral norm violations should be independently explained through a potential polysemy whereby ‘cause’ may simply mean ‘is morally responsible for’. In support of the latter view, recent work has pointed out that moral norm violations affect judgments of agents, but not inanimate objects, and that these effects are moderated by agents’ knowledge states. We advance this debate by demonstrating that judgments of counterfactual relevance exhibit precisely the same patterns, and that judgments of inanimate objects are actually highly sensitive to whether the object violated a prescriptive norm by malfunctioning. The latter finding is difficult to account for through polysemy, but is predicted by changes in the relevance of counterfactual alternatives. Finally, we show that direct (non-moral) interventions on the the relevance of counterfactual alternatives affect causal judgments in precisely the same way as functional and moral norm violations.

Keywords: causation; norms; counterfactuals; morality; teleology

Introduction

A central question in research on causal cognition concerns the role of norms. It is well-known that both descriptive norms (e.g., the probability of an event occurring) and prescriptive norms (e.g., the morality of an event occurring) influence judgments of actual causation, that is, a judgment that some event, *e*, was the cause of some outcome, *o* (Alicke, 2000; Gerstenberg & Tenenbaum, in press; Hitchcock & Knobe, 2009; Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015). Specifically, people are more inclined to judge that *e* was the cause of *o* if *e* was either very unlikely to happen or morally prohibited. Despite widespread agreement on the existence of the phenomenon, there has been little corresponding agreement on how these effects should be explained.

Most researchers take the impact of descriptive *statistical* norms to reveal part of the basic underlying processes that support causal reasoning (e.g., Gerstenberg & Tenenbaum, in press; Icard, Kominsky, & Knobe, in press; Samland & Waldmann, 2016). They differ, however, in whether they treat the impact of prescriptive moral norms similarly, or argue that it arises from a fundamentally different set of processes.

On one side, researchers have argued that the impact of both descriptive and prescriptive norms is best explained by changes in the relevance of counterfactual possibilities. These accounts propose that when a norm violation occurs, it increases the relevance of counterfactual alternatives wherein the norm violations are replaced by norm-conforming events (e.g., Halpern & Hitchcock, 2015; Kominsky et al., 2015; Bello, 2016). In support of this account, recent work demonstrated that norm violations affect explicit assessments of

counterfactual relevance in precisely the same way that they affect causal judgments (Phillips, Luguri, & Knobe, 2015).

On the other side, other researchers have argued for separate explanations of the two effects. The most recent approach has suggested that the term ‘cause’ is polysemous: It can be used to talk about whether an agent is morally responsible for an outcome, or it can be used to talk about whether some event causally contributed to an outcome (Samland & Waldmann, 2016). On this approach, the impact of violations of moral norms can instead be accounted for by arguing that participants are more likely to interpret the word ‘cause’ as being about moral responsibility in cases where moral norms have been violated.

Advancing this debate, Samland and Waldmann (2016) (S&W hereafter) reported two important new data points: First, the violation of moral norms selectively influences causal judgments about whether agents caused an outcome, but not causal judgments of whether inanimate objects used by the same agents caused the outcome. Second, factors that affect the moral responsibility of the norm violator (such as their knowledge states) also affect causal judgments (see also Samland, Josephs, Waldmann, & Rakoczy, 2016). S&W suggest that these findings are best accounted for by assuming that participants were interpreting the causal question to be about moral responsibility when asked about an agent, but about simple causal contribution when asked about an object.

In arguing that these results provide evidence against a unified counterfactual account, S&W rely on the assumption that when a norm violation occurs, people consider counterfactual alternatives to the event in its entirety. That is, they consider a counterfactual alternative that involves both the agent who violated a norm and the inanimate object used by that agent. If this assumption is correct, then a polysemy account seems to better capture S&W’s results, since a unified counterfactual account would predict that causal judgments of the agent and the inanimate object would *both* be affected.

At the same time, though, it is possible that the counterfactual alternatives people represent are more granular. That is, when a moral norm violation occurs, people may consider a counterfactual alternative that involves the norm-violating agent, but not the inanimate object used by that agent. If this turns out to be correct, then S&W’s findings should be understood as perfectly compatible with a unified counterfactual explanation, as this accounts would then predict that causal judgments of the agent, but not the object, would be affected.

To distinguish these possibilities, we begin by asking whether the effects uncovered by S&W also arise in participants’ assessments of which counterfactuals are relevant. One possibility is that, because moral norms apply to agents but not inanimate objects, participants will regard counter-

factual alternatives to what the agent did as relevant, but not alternatives to what the inanimate object did. If so, it would suggest that they are represented somewhat independently of one another. Furthermore, changes to agents' mental states may affect both the agent's moral responsibility and similarly whether it is relevant to consider counterfactual alternatives to their actions, which may help explain why changes to agents' mental states affect causal judgments (see, e.g., Lombrozo, 2010 on how intentions affect causal judgments in double-prevention scenarios).

Experiment 1

Methods

Participants. 610 participants ($M_{age} = 37.28$, $SD_{age} = 12.14$; 338 females, 1 unreported) from Amazon Mechanical Turk participated for a \$0.25 compensation. Participant recruitment was automated through TurkPrime (www.turkprime.com) to prevent repeat participation and limit recruitment to participants with a high approval rating.

Stimuli and procedure. This experiment was nearly identical to S&W's Experiment 4, but with an additional DV. The overall design was 4 (norm condition) \times 3 (question) and administered fully between-subjects. Participants read one of four vignettes (see Supplementary Materials available at <https://github.com/phillipsjs/stillRelevant>). In all conditions, Tom owns a garden and has two gardeners, Alex and Benni, who each take care of 1/3 of the plants on their own, and jointly tend to the remaining 1/3. Additionally, Alex and Benni always use two fertilizers "A-X200®" and "B-Y33®". Tom reads that fertilizers are good for plants, but using more than one kind of fertilizer could damage his plants, so Tom decides he wants both gardeners to use only fertilizer A-X200. In all cases, however, Alex applies fertilizer A-X200 and Benni applies fertilizer B-Y33, and the plants cared for by both of them are damaged.

The four conditions varied the reason that Benni used B-Y33. In the Standard norm-violation condition, Benni simply decides to use B-Y33; in the Unintended norm-violation condition, Benni believed he was applying A-X200, but accidentally applied B-Y33; in the Ignorant norm-violation condition, Tom neglects to tell Benni to use only A-X200, and he uses B-Y33 instead; and in the Deceived norm-violation condition, Alex deliberately lies to Benni about which fertilizer he is supposed to use to get him in trouble. We additionally varied the focus of the questions. Participants were either asked questions that focused on the two agents ("Alex" and "Benni"), the two actions ("the application of fertilizer by Alex" and "the application of fertilizer by Benni"), or the two chemicals ("the application of chemical A-X200" and "the application of chemical B-Y33").

After reading the vignette, participants were asked whether it was relevant to consider counterfactual alternatives to some aspect of the event, following Phillips et al., (2015). For example, in the Agent condition, participants indicated both whether they thought it was relevant or irrelevant to consider

what Alex could have done differently and also whether it was relevant or irrelevant to consider what Benni could have done differently. Subsequently, as in S&W, participants were asked to judge who or what caused the plants to dry up (again depending on the Question condition). In the Agent condition, participants indicated both whether they thought Alex was a cause and also whether they thought Benni was a cause. Because the causal question was simply a replication of S&W (who did not include a counterfactual question), the counterfactual question was always presented first and mirrored S&W's causal question as closely as possible.

Following these question, participants received two check questions that tested their understanding of which chemicals were applied by which gardener, and which chemicals Tom wanted each gardener to use. Following S&W, they were also asked to estimate the proportion of the flowers that dried when (1) only fertilizer A-X200 was applied, (2) only fertilizer B-Y33 was applied, and (3) both were applied.

Results

We excluded participants who did not answer both of the check questions correctly, and analyzed the remaining 439 participants' judgments. (Note that here and throughout the following experiments, all of the key results remain when these exclusion criteria are relaxed.) To examine the effects of our manipulation on both causation and relevance judgments, we categorized participants' responses as assigning causal responsibility (or counterfactual relevance) to (1) only the norm-violating agent, (2) both agents, or (3) only the norm-conforming agent, and then subjected both kinds of judgments to a proportional odds logistic regression using the probit function in the MASS package in R. For causal judgments, we observed an effect of the norm-condition ($LRT = 20.49$ [$df = 3$], $p < .001$), an effect of question ($LRT = 44.53$ [$df = 2$], $p < .001$), and critically, a norm-condition \times question interaction effect ($LRT = 19.94$ [$df = 6$], $p = .003$). This precisely replicates the pattern of data observed in S&W (the complete information on the replication of the key statistical tests reported in S&W is available in the Supplementary Materials). Importantly, this direct replication of S&W provides evidence that answering the counterfactual question first did not unduly influence participants' causal judgments.

We next analyzed participants' relevance judgments, and observed a highly similar pattern of results: an effect of norm-condition ($LRT = 13.93$ [$df = 3$], $p = .003$), an effect of question ($LRT = 73.34$ [$df = 2$], $p < .001$), and a norm-condition \times question interaction effect ($LRT = 14.15$ [$df = 6$], $p = .028$). Critically, because participants answered this question first, the observed pattern cannot have been influenced by participants' causal judgments. All the same, at the level of each participants' responses, judgments of the causal responsibility were highly correlated with judgments of whether it was relevant to consider alternatives to the agents' actions. This was true both for judgment of the norm-violating agent/action/object, (Pearson's $r = 0.553$, $p < .001$), and for the norm-conforming agent/action/object

(Pearson’s $r = 0.406$, $p < .001$), and moreover, held whether participants were making judgments about agents (Pearson’s $r = 0.651$, $p < .001$), actions (Pearson’s $r = 0.262$, $p < .001$), or simply inanimate objects (Pearson’s $r = 0.280$, $p < .001$). The similarity in the overall pattern of these judgments across all of the conditions can be seen in Figure 1.

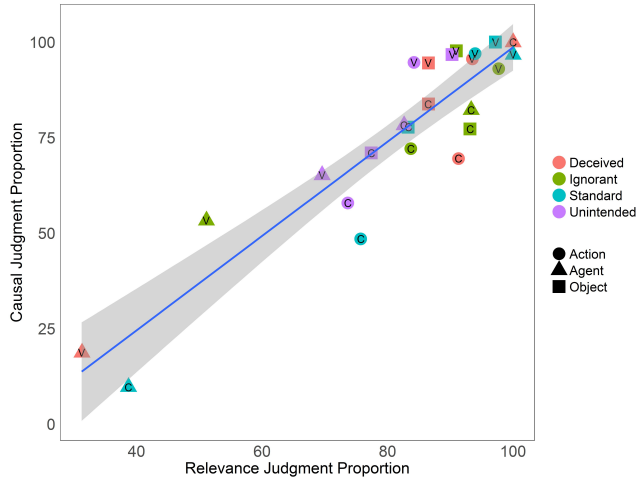


Figure 1: Depiction of the relationship between participants’ causal ratings and relevance ratings. Judgments related to the norm-conforming agent are marked with a ‘C’; Judgments related to the norm-violating agent are marked with a ‘V’.

Discussion

In sum, we replicated S&W’s causal judgments, and found a corresponding pattern for which counterfactual alternatives participants regarded as relevant. These findings mirror those in observed in Phillips et al. (2015), which used a continuous rather than dichotomous measure of counterfactual relevance, and more importantly are predicted by a unified counterfactual account of the impact of norms on causal judgments.

Experiment 2

We next investigated whether causal judgments of inanimate objects are sensitive to violations of prescriptive norms of *proper functioning* (e.g., a machine malfunctioning). According to a unified counterfactual account, when a machine malfunctions, it should become more relevant to consider the counterfactual possibility that the machine could have instead functioned as intended, and thus the machine should be judged as more causal. Polysemy accounts do not predict such an effect, as participants should not interpret the word ‘cause’ to mean ‘morally responsible’ when discussing an inanimate object. We test these two predictions.

Methods

Participants. 403 participants ($M_{age} = 34.96$, $SD_{age} = 11.90$; 205 females, 1 unreported) from Amazon Mechanical Turk participated for a \$0.25 compensation. Participant recruitment was again automated through TurkPrime.

Stimuli and procedure. This experiment used a 3 (Norm violation; norm-conforming vs. moral violation vs. malfunction) x 2 (Question: agent vs. object) design, administered fully between-subjects.

Participants read one of three vignettes involving a vending machine in an academic department. In every condition, the machine has three levers (red, black, and white): two produce pencils and one produces an eraser but frequently malfunctions and also produces a broken pencil. There were also two agents: an administrative assistant, and Professor Smith (a recent hire who did not know about the malfunctioning lever). Prof. Smith always pulls the red lever, and the assistant always pulls the black lever. This later results in a problem for a student who needs a pencil to take a test but cannot get one.

In the norm-conforming condition, the red lever and black lever both produce pencils, and the white lever produces erasers (but also consistently malfunctions). Additionally, both administrators and faculty were allowed to take pencils from the machine. Both agents request pencils using the black and red levers, which both function appropriately. The moral violation condition was identical to the norm-conforming condition, except that the faculty are not allowed to get pencils from the machine (but administrative assistants are allowed), and this rule was known by Prof. Smith. Lastly, the malfunction condition was identical to the norm-conforming condition except that it was the red lever that produced erasers (and malfunctioned), and Prof. Smith wanted an eraser, so Prof. Smith pulled the red lever and got an eraser and a broken pencil.

Participants were then asked a question about the relevance of counterfactual alternatives and a causal question in random order on separate pages. The relevance of alternatives question was worded and presented the same way as Experiment 1, and either focused on the agents (Prof. Smith, administrative assistant) or the objects (red lever, black lever). The causal question similarly asked either who caused the problem (agent condition) or what caused the problem (object condition), and participants could select one or both.

These were followed by three comprehension check questions and two additional manipulation-check questions. The comprehension questions ensured that participants understood the key facts about the levers, agents, and outcome of the scenario. Additionally, participants rated, on a 0-100 scale, how likely the malfunction was to occur, in order to verify that participants did not think the malfunction also violated a descriptive (statistical) norm. Finally, participants rated their agreement with the statement “It was morally wrong for Prof. Smith to pull the red lever” on a 7-point Likert scale, with the expectation that ratings should be higher in the moral violation condition than the other two conditions, which should not differ from each other. The predictions for both manipulation-check questions were overwhelmingly confirmed (see Supplementary Materials). Thus, any effect of the functional norm violation cannot be explained by appealing to statistical or moral norms.

Results

We excluded participants who did not answer all three of the check questions correctly, and analyzed the remaining 258 participants' judgments. To facilitate comparison of participants' judgments, we computed a measure of participants' preference for selecting the norm-violating event as a cause. Participants who selected only the norm-violating event as a cause were assigned a score of 1; participants who selected both or neither events as causes were assigned score of 0; and participants who selected only the norm-conforming event were assigned a score of -1. We then analyzed participants' causal preference scores with a 2 (Causal Question: Agent vs. Object) \times 3 (Norm condition: Immoral vs. Malfunction vs. Normal) proportional odds logistic regression, as in Study 1. This analysis revealed a main effect of Norm condition, ($LRT = 71.49$ [$df = 2$], $p < .001$), no main effect of Causal question ($LRT = 0.045$ [$df = 1$], $p = .832$), and critically a Norm condition \times Causal question interaction effect ($LRT = 31.42$ [$df = 2$], $p < .001$).

We decomposed this interaction effect by separately analyzing participants' causal preference scores for each of the different conditions. When the relevant norm was moral and thus applied to the agent but not the object, participants tended to prefer the norm-violating agent as a cause, but did not similarly prefer the norm-violating object as a cause ($LRT = 15.33$ [$df = 1$], $p < .001$). When the relevant norm was functional, and thus the norm applied to the object but not the agent, this pattern was reversed: participants tended to prefer the norm-violating object as a cause, but did not similarly prefer the norm-violating agent as a cause ($LRT = 12.36$ [$df = 1$], $p < .001$). When there was no norm that applied to either the agent or the object, there was small and non-significant preference for the norm-conforming agent but not the object ($LRT = 1.13$ [$df = 1$], $p = .288$).

We next analyzed participants' judgments of the relevance of counterfactual alternatives in exactly the same way. Just as with participants' causal judgments, we observed a main effect of Norm condition, ($LRT = 40.53$ [$df = 2$], $p < .001$), no main effect of Relevance question ($LRT = 0.10$ [$df = 1$], $p = .747$), and critically a Norm condition \times Relevance question interaction effect ($LRT = 33.70$ [$df = 2$], $p < .001$). We decomposed this interaction effect by separately analyzing participants' counterfactual preference scores for each of the different conditions. When a moral norm was salient, participants tended to prefer counterfactuals for the agent, but not the object ($LRT = 16.63$ [$df = 1$], $p < .001$). When the relevant norm was functional, this pattern was reversed: participants preferred counterfactuals for the object, but not the agent ($LRT = 11.20$ [$df = 1$], $p < .001$). When there was no norm violation that applied to either the agent or the object, there was a small and significant preference for the norm-conforming agent, but not the object ($LRT = 4.48$ [$df = 1$], $p = .034$). A similar pattern is found when only participants' first responses are analyzed, allowing for a between-subjects analysis (see Supplementary Materials).

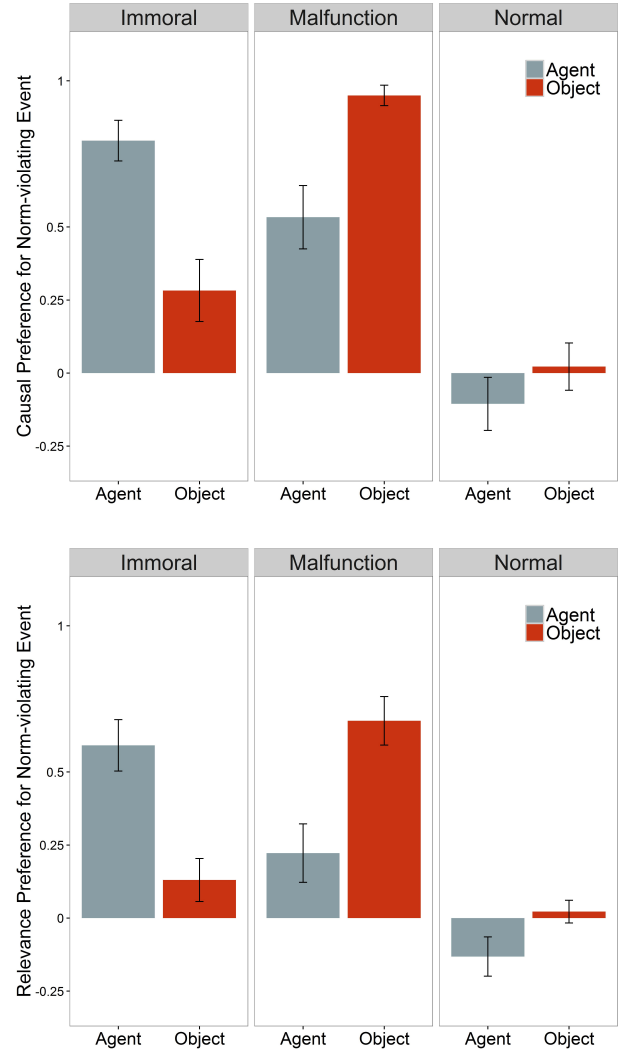


Figure 2: Average preference score for the norm-violating event in causal judgments (top) and counterfactual relevance judgments (bottom), as a function of which norms were relevant (split into panels). Grey bars depict responses to questions about agents; Red bars depict responses to questions about inanimate objects. Error bars depict ± 1 SEM.

Discussion

Experiment 2 found that judgments of inanimate objects are sensitive to violations of prescriptive norms of proper functioning, even though they are not sensitive to violations of moral norms. Specifically, we found that when an inanimate object violated a functional norm, participants' thought it was relevant to consider counterfactual alternatives to that malfunction, and that this effect was mirrored by a corresponding change in participants' causal judgments. This pattern is uniquely predicted by a unified counterfactual accounts of the impact of norms on causal judgments, and is not predicted by an account on which the term 'cause' is polysemous.

Experiment 3

Previous research (Phillips et al., 2015) has demonstrated that causal judgments are also sensitive to more direct counterfactual manipulations: participants tend to judge an event to be more causal after they generate alternative ways that the event could have occurred. In Experiment 3, we extend this method by asking participants to generate alternatives to one particular aspect of the causal structure that contributes to the outcome (i.e., to the agent or to the inanimate object). We then measure how their causal judgments are affected by this manipulation. This allows us to test a precise prediction of a unified counterfactual account: participants causal judgments should be affected by the generation of counterfactual alternatives primarily for the part of the causal structure that the counterfactual alternative focused on.

Methods

Participants. 601 participants ($M_{age} = 35.96$, $SD_{age} = 15.58$; 304 females, 2 unreported) from Amazon Mechanical Turk participated for \$0.35 in compensation. Participant recruitment was again automated through TurkPrime.

Stimuli and procedure. This experiment used a 3 (Agent-Counterfactual vs Object-Counterfactual vs No Counterfactual) \times 2 (Agent Question vs Object Question) design. Counterfactual condition was manipulated between-subjects and Question was manipulated within-subjects.

All participants read the norm-conforming condition from Experiment 2, where the red lever and black lever both produce pencils, and the white lever produces erasers (but also consistently malfunctions). In this scenario, both the administrators and the faculty are allowed to take pencils from the machine, both the administrative assistant and Professor Smith request pencils using the black and red levers respectively, and both levers function appropriately to produce pencils. A problem then arises from a lack of pencils.

After reading the vignette, participants underwent the counterfactual manipulation. In the Agent-Counterfactual condition, for example, participants were asked to think about Professor Smith's decision to take a pencil from the vending machine, and then to consider and describe one relevant way that things could have gone differently such that the professor would not have taken one of the pencils from the vending machine. In the Object-Counterfactual condition, by contrast, participants were instead asked to consider and describe a relevant way in which the red lever could have functioned differently such that it didn't produce a pencil from the vending machine. In the No Counterfactual condition, participants were simply asked to describe the story they read.

After completing this task, they rated their agreement (on a scale from 0 ('Completely disagree') to 100 ('Completely agree')) with a statement that the Professor caused the problem, and separately with a statement that the red lever caused the problem. The statements were presented in counterbalanced order and on separate pages. Participants then completed a series of control questions that asked them about

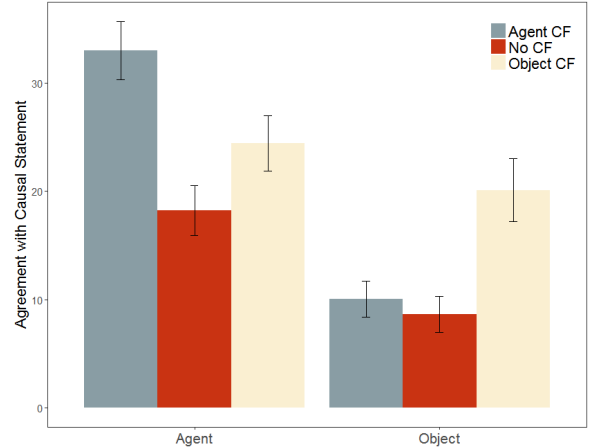


Figure 3: Agreement with the causal statement concerning the agent (left bars) and the object (right bars) as a function of Counterfactual condition). Error bars depict ± 1 SEM.

which levers were actually pulled and about who actually received a pencil in the original story.

Results

We excluded participants who did not answer both of the check questions correctly, and analyzed the remaining 423 participants' judgments. First, we analyzed the agreement with the two causal statements by comparing a series of linear mixed-effects models using the lme4 package in R (Bates, Maechler, Bolker, Walker, et al., 2014). This analysis revealed a main effect of Question ($\chi^2(1) = 53.135$, $p < .001$) and a main effect of Condition ($\chi^2(2) = 13.492$, $p = .001$). Critically, however, these were qualified by a significant Question \times Condition interaction ($\chi^2(2) = 23.04$, $p < .001$). We decomposed this interaction using a series of planned comparisons. These analyses revealed that participants strongly agreed that Professor Smith was a cause of the problem when they considered alternatives to Professor Smith's action ($M = 32.99$, $SD = 33.33$), but in comparison, agreed significantly less both when they did not consider counterfactual alternatives ($M = 18.22$, $SD = 27.28$), $t(282.48) = 4.12$, $p < .001$, $d = 0.482$, and when they only considered alternatives to the way the lever functioned ($M = 24.43$, $SD = 29.12$), $t(279) = 2.27$, $p = .024$, $d = 0.272$.

We also observed a corresponding pattern in participants' agreement with the statement that the red lever caused the problem: participants agreed that the lever was more of a cause when they considered alternatives to the way the lever functioned ($M = 20.11$, $SD = 33.34$), than when they did not generate any relevant counterfactual alternatives, ($M = 8.62$, $SD = 19.64$), $t(211.21) = 3.42$, $p < .001$, $d = 0.421$, or when they considered alternatives to what Professor Smith did ($M = 10.05$, $SD = 20.59$), $t(213.65) = 2.99$, $p = .003$, $d = 0.367$.

Discussion

In short, we found that directly manipulating the relevance of counterfactual alternatives affected participants' causal judgments. Moreover, in line with the predictions of a unified counterfactual account, we found this effect was specific to the factor that was altered in the counterfactual alternative.

General Discussion

The results of these three experiments favor a counterfactual relevance account of the impact of norms on causal judgments. Experiment 1 replicated S&W's finding that moral norm violations primarily affect causal judgments of intentional agents and not inanimate objects. Experiment 2 further found that violations of norms of proper functioning primarily affect judgments of inanimate objects but not intentional agents. In both experiments, judgments of counterfactual relevance tracked the impact of different norm violations on causal judgments for both intentional agents and inanimate objects. Finally, Experiment 3 demonstrated that non-normative manipulations of counterfactual relevance produce an analogous pattern in participants' causal judgments. These results support work on causal cognition that provides a central role for counterfactuals (Gerstenberg & Tenenbaum, in press; Kominsky et al., 2015; Icard et al., in press).

The extant literature on causal judgment now provides evidence for three distinct types of norms that all show similar effects: descriptive *statistical* norm violations (e.g., Kominsky et al., 2015), prescriptive *moral* norm violations (e.g., Hitchcock & Knobe, 2009), and prescriptive *functional* norm violations (demonstrated here). The demonstration of additional norms that have similar a impact on causal judgments makes a parsimonious explanation increasingly desirable. To extend the polysemy account, for example, one would now have to propose three independent explanations for three qualitatively similar effects. By contrast, an account that appeals to the relevance of counterfactual alternatives provides a unified explanation and predicts that these various norms should all have a qualitatively similar impact.

At the same time, many aspects of the relationship between counterfactual representation and causal cognition remain poorly understood. For example, a critical insight which arises in both S&W and in the current studies is that norms have a highly specific effect on causal judgments: they preferentially affect causal judgments of the entities to which the norm applies and typically do not extend to other aspects of the same event. Across three experiments, we find a similar pattern in participants' reasoning about counterfactual alternatives. Collectively, these findings suggest that, rather than representing a counterfactual alternative to an event in its entirety, participants' causal and counterfactual cognition represents events more granularly.

Not only does this shape our interpretation of S&W's original result, it opens an exciting new frontier in the study of causal cognition. How events are represented in causal and counterfactual cognition, and which aspects of an event are

represented as distinct variables, are almost completely unexplored topics (e.g. Halpern & Hitchcock, 2015 explicitly acknowledge this issue). However, as emerging research makes clear, it will be difficult to make precise predictions about the impact of norms without a more well worked-out theory of how events are represented in causal reasoning.

This opportunity cuts in both directions. These results are, to our knowledge, the first empirical investigation of which events are represented as distinct variables. Yet, as much as we need to build precise theories of how these events are represented in order to understand how norms will affect causal judgments, we can also use the effect of norms on causal judgments to determine which causes are distinct. We look forward to exploring these questions in future work.

Acknowledgments

We would like to thank Harvard's Moral Psychology Research Lab, D. Lagnado, T. Gerstenberg, and J. Knobe for helpful discussion. This work was supported by Grant N00014-14-1-0800 from the Office of Naval Research to JP.

References

- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological bulletin*, 126, 556.
- Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2014). lme4: Linear mixed-effects models using eigen and s4. *R package version*, 1.
- Bello, P. F. (2016). Machine ethics and modal psychology. In *Computing and philosophy* (pp. 245–258). Springer International Publishing.
- Gerstenberg, T., & Tenenbaum, J. B. (in press). Intuitive theories. In M. Waldmann (Ed.), *Oxford handbook of causal reasoning*. Oxford University Press.
- Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science*, 66, 413–457.
- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, 106, 587–612.
- Icard, T. F., Kominsky, J., & Knobe, J. (in press). *Causality, normality, and sampling propensity*. Elsevier.
- Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, 137, 196–209.
- Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, 61, 303–332.
- Phillips, J., Luguri, J., & Knobe, J. (2015). Unifying moral-ity's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, 145, 30–42.
- Samland, J., Josephs, M., Waldmann, M. R., & Rakoczy, H. (2016). The role of prescriptive norms and knowledge in childrens and adults causal selection. *Journal of Experimental Psychology: General*, 145, 125.
- Samland, J., & Waldmann, M. R. (2016). How prescriptive norms influence causal inferences. *Cognition*, 156, 164–176.