Causal Supersession

Jonathan F. Kominsky[1] (jonathan.kominsky@yale.edu), Jonathan Phillips[1] (jonathan.phillips@yale.edu), Tobias Gerstenberg[2] (tger@mit.edu), David Lagnado[3] (d.lagnado@ucl.ac.uk) Joshua Knobe[1] (joshua.knobe@yale.edu),

[1] Yale University, 2 Hillhouse Ave, New Haven, CT 06520, USA
[2] MIT, Building 46-4053, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[3] University College London, Gower Street, London, WC1E 6BT, UK

Abstract

When agents violate norms, they are typically judged to be more of a cause of resulting outcomes. In this study, we suggest that norm violations also reduce the causality of other agents, a novel phenomenon we refer to as "causal supersession." We propose and test a counterfactual reasoning model of this phenomenon in three experiments. Experiment 1 shows that causal judgments of one actor are reduced when another actor violates moral norms, even when the outcome in question is neutral. Experiment 2 shows that this causal supersession effect is dependent on a particular event structure, following a prediction of our counterfactual model. Experiment 3 demonstrates that causal supersession can occur with violation of non-moral norms.

Keywords: Causal reasoning; Counterfactuals; Morality; Supersession

Introduction

In the 1870 case of *Carter v. Towne*, the court faced an intriguing question. The defendant sold gunpowder to a child. The child's mother and aunt hid the gunpowder, but in a location that they knew the child could find and access. The child found the gunpowder and was injured. The court found that the defendant could not be considered to be the cause of the child's injuries, because of the negligence of the mother and aunt (Hart & Honoré, 1985, pp. 281-282).

This case leaves us with an interesting puzzle about causal reasoning. The question before the court was not whether the mother and aunt caused the outcome; it was whether the defendant caused the outcome. Yet the court determined that the fact that the actions of the mother and aunt were negligent had some effect on the causal relationship between the defendant's actions and the outcome. What we find here is somewhat counter-intuitive: the defendant's responsibility for the outcome is being affected not by anything that the defendant himself did, but by someone else's actions! This suggests a broader phenomenon: the causality of one agent may be affected not only by the moral status of their own actions, but by the moral status of others' actions. We refer to this as "causal supersession."

It is well-established that moral judgments can affect causal judgments. Previous research has shown that when an agent acts in a way that is judged to be morally wrong, that agent is seen as more causal (e.g., Alicke 1992). Recent work has suggested that, rather than being about morality specifically, these effects are rooted in the normality of an agent's actions, that is, how much they diverge from moral or statistical norms (Halpern & Hitchcock, in press; Hitchcock & Knobe, 2009). However, most of the experimental work to date has focused on how the normality of an agent's actions affects that agent's own causality, not anyone else's. The present experiments aim to demonstrate and explore causal supersession.

The novelty of causal supersession

We suggest that causal supersession provides novel insight into causal reasoning, but first we wish to acknowledge another explanation for the phenomenon from an existing well-supported motivational theory (Alicke, 2000). It is already known that people's causal judgments can be impacted by motivational factors. Notably, recent work has found that people's judgments are often distorted by "excuse validation" (Turri & Blouw, in press), the motivation to *not* assign causality for bad outcomes to individuals whom we do *not* feel are blameworthy. For example, if a driver is speeding because of an accelerator malfunction and gets into a lethal accident, we might be inclined to exculpate the driver as a cause of the accident because her actions are blameless. We do not wish to say that a blameless individual is the cause of a blameworthy outcome. This basic idea could then be used to explain causal supersession. If one agent does something morally wrong and is therefore seen as the one who is to blame for the outcome, people will be motivated to find some way to conclude that all other agents are not to blame, and they will therefore conclude that those other agents did not cause the outcome.

This explanation draws on claims that have already received extensive support in the existing empirical literature (Alicke, 2000), and we do not mean to call the empirical claims into question here. Instead, we will suggest that causal supersession provides evidence for a different process that goes beyond what has been demonstrated in existing work.

A counterfactual account of causal supersession

We propose an account of the supersession effect based on counterfactual reasoning. Independent of any motivational factors, we will argue that counterfactual reasoning affects causal judgment in these cases, and that norm violations influence counterfactual reasoning. This account follows

from two key claims. First, counterfactual reasoning affects causal judgment. Second, moral valence affects counterfactual reasoning.

Counterfactual reasoning and causal judgment

Previous work on counterfactuals and causal judgment has suggested that people regard an event as a cause of the outcome when it satisfies two conditions, *necessity* and *sufficiency* (e.g., Woodward, 2006). Take the causal relationship "A caused B". Roughly speaking, this relationship would have the following necessity and sufficiency conditions:

Necessity: If A had not occurred, B would not have occurred.

Sufficiency: If A occurs, B occurs.

Our focus here will be on the second of these conditions – sufficiency – and on the role it plays in ordinary causal cognition.

Woodward (2006) defines a property he calls *sensitivity* to characterize the robustness with which a condition is satisfied. The sufficiency of a cause is 'sensitive' if it would cease to hold if the background conditions were even slightly different. By contrast, a condition is 'insensitive' if it would continue to hold even if the background conditions were substantially different. Woodward argues that when the sufficiency condition is highly sensitive, people will show some reluctance to attribute causation.

To give a more concrete example, Woodward describes a case in which Billy tells Suzy that if she scratches her nose, he will throw a rock at a bottle. She does, he throws the rock at the bottle, and it breaks. Looking at the causal relationship between the rock hitting the bottle and the bottle breaking, the sufficiency counterfactual ("If the rock were to hit the bottle, the bottle would break") is very insensitive, as the rock hitting the bottle will cause the bottle to break under many circumstances. By contrast, looking at the causal link between Suzy scratching her nose and the bottle breaking, the sufficiency counterfactual ("If Suzy were to scratch her nose, the bottle would break") is quite sensitive, because any number of small changes to the background conditions would render Suzy's action no longer sufficient. Thus, the account predicts that people are less likely to agree with the statement: "Suzy's action caused the bottle to break."

This claim about the importance of sufficiency is the first piece of our account of causal supersession. In the case of *Carter v. Towne,* for example, the defendant's action was only sufficient to bring about the outcome because the mother and aunt happened to act negligently. If the mother and aunt had not acted negligently, then even if the defendant had performed exactly the same action, the outcome would not have come about. It is for this reason, we claim, that people are somewhat disinclined to regard the action as fully causal. Certain facts about the child's guardians make the relationship between the defendant and the outcome *sensitive*.

Moral valence and counterfactuals

We noted above that a relationship could be considered "sensitive" to the extent that it would not have held if the background circumstances had been slightly different. Yet, there will always be *some* way that the background circumstances could have been different such that sufficiency would no longer hold. (For example, suppose that someone said: "The rock only happened to be sufficient to break the bottle because the bottle wasn't covered in a steel casing. If it had been, the rock would not have been sufficient.") But not all counterfactuals are treated equally. Even if this counterfactual claim is correct, there seems to be some important sense in which it is *irrelevant* – not even worth thinking about. If we want to understand the notion of sensitivity, we need to say a little bit more about this issue, providing a sense of how to determine whether a given counterfactual is relevant.

Drawing on the substantial body of research on counterfactual reasoning (for reviews, see Byrne 2005; Kahneman & Miller, 1986), we will focus here on two principal findings. First, studies show that *likelihood* judgments play a role in people's intuitions about which counterfactuals are relevant and which are not (Byrne, 2005; Kahneman & Tversky, 1982). When something unlikely occurs, people tend to regard as relevant counterfactuals that involve something more likely occurring. Second, studies show that *moral* judgments can influence people's intuitions about the relevance of counterfactuals (McCloy & Byrne, 2000; N'gbala & Branscombe, 1995). When an agent performs a morally bad action, people tend to regard as relevant counterfactuals that involve the agent doing something morally neutral.

To unify these two findings, we can say that people's intuitions about the relevance of counterfactuals are affected by violations of *norms* (Hitchcock & Knobe, 2009). In some cases, an event is seen as unlikely (and hence violates a statistical norm); in other cases, an event is seen as morally wrong (and hence violates a moral norm). Thus, we can formulate a more general principle, which should apply across both types of norm violation. The general principle is: when an event in the actual world is perceived as violating a norm, people tend to regard as relevant counterfactuals in which the norm-violating event is replaced by a norm-abiding event.

The complete counterfactual account

Combining these two ideas yields a counterfactual account of causal supersession. Take the causal claim "The defendant selling gunpowder to the child caused the child's injuries." The sufficiency condition for this claim reads as follows: "If the defendant sells gunpowder to the child, then the child is injured." Now suppose that sufficiency only holds because the mother and aunt negligently hid the gunpowder where the child could find it. Since this act violates a norm, people will tend to regard as highly relevant the possibility in which the gunpowder is put somewhere that the child could not get it. In that possibility, the defendant's action would not have been

sufficient, so the negligent actions of the mother and aunt make the defendant's sufficiency more sensitive. Thus, the defendant is regarded as less of a cause of the outcome, or in other words, is superseded.

Putting this point more abstractly: When an actor does something that violates a norm, it makes the possibility that they did *not* do that thing very relevant. If the sufficiency condition is not met in those possibilities, the sufficiency of the causal link between that actor and the outcome becomes sensitive. Because the sufficiency of that causal link is sensitive, the second actor is seen as less of a cause of the outcome.

Predictions of the counterfactual account

The first novel prediction of the counterfactual account is that causal supersession should occur even for completely neutral outcomes. This goes beyond, but does not contradict, motivational accounts. While you may be motivated to blame someone even in the absence of a bad outcome, such blame cannot be justified by insisting that the agent caused some *neutral* outcome. Similarly, saying that an agent did not cause a neutral outcome would not help to exonerate her from blame. Rather, one might well be motivated to find some way of justifying the claim that the agent is not blameworthy. Existing work on motivational biases in causal cognition has used precisely this logic to show that certain effects are indeed the product of motivation (Alicke, Rose & Bloom, 2011). In contrast, the counterfactual account does not require that the outcome be bad in order for supersession to occur. From the standpoint of the counterfactual account, the only thing that matters is the (ab)normality of what the superseding actor did.

The second prediction is not about when supersession should occur, but rather when it should not. The counterfactual account does not treat the assignment of causality to different actors as a zero-sum game. Our account predicts that supersession should only occur when the sufficiency of the superseded actor is threatened. More concretely, if one agent violates a norm, we suggest that people consider the possibility in which that agent did not act. If the other agent is still sufficient even if the first does not act, then the wrongness of the first agent's actions should not affect the judged causality of the second (see the introduction to Experiment 2 for further details).

Finally, the third prediction is that supersession should arise for any norm violation, not just for violations of moral norms. The key thing that moral valence is doing in the counterfactual account is making certain counterfactual possibilities more relevant, and those possibilities make the sufficiency of the superseded actor sensitive. Since violations of statistical norms also make people regard counterfactual possibilities as more relevant (Kahneman & Tversky, 1982), violations of these purely statistical norms should yield similar supersession effects.

We test each of these predictions in turn in Experiments 1 through 3.

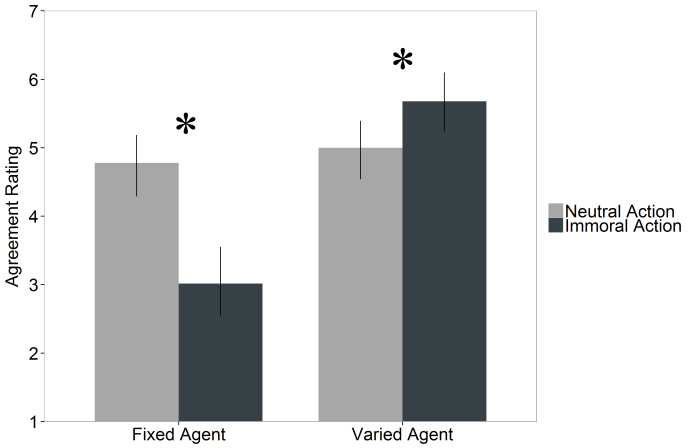

Figure 1: Mean agreement ratings as a function of agent and valence of the varied agent's actions. * = $p<.05$ Error bars indicate bootstrapped 95% CI.

Experiment 1

In the first experiment, we wanted to demonstrate causal supersession for events with neutral outcomes. We constructed a scenario with two agents. One agent, whom we will call the "fixed" agent, always acted in the same way. Her actions were always morally neutral. The second agent, whom we will call the "varied" agent, acted differently depending on condition. We manipulated the varied agent's actions on two dimensions: intent (intending to bring about the specific outcome vs. not intending to bring about the specific outcome) and moral valence (morally neutral vs. morally wrong).

The counterfactual account predicts that the fixed agent should be seen as less causal when the varied agent's actions are morally wrong, compared to when the varied agent's actions are morally neutral, since both are necessary to bring about the outcome. Motivational accounts do not make this prediction, since assigning greater causal responsibility for a neutral outcome does not help to justify blaming or exonerating from blame. We made no specific predictions about the impact of intentionality with regard to the supersession effect, but since existing research has found an impact of intentionality on causal cognition (Lombrozo, 2010) it was included in this initial investigation as a potentially informative factor.

Methods

Participants 120 participants were recruited via Amazon Mechanical Turk and paid $0.20 each for completing the two-question survey.

Materials and procedure. Four vignettes[1] were created featuring a varied agent (Bill) and fixed agent (Bill's Wife). All vignettes started with the fixed agent finding the left side of a specific paired set of bookends at an antique store.

[1] All vignettes are available in full online: http://goo.gl/9tiM4i

The varied agent then acquired the right-side bookend for the paired set. The varied agent's actions were manipulated along two dimensions, the moral wrongness of his actions (buying the right-side bookend from a friend vs. stealing it), and whether he intended to bring about the outcome (intended to get a matching bookend vs. intended to get any right-side bookend). This created a 2x2 design, which was run between-subjects. In all conditions, participants were asked to rate on a 1-7 scale how much they agreed with each of the two sentences: "Bill's wife caused them to possess the paired set of bookends" (the fixed agent) and "Bill caused them to possess the paired set of bookends" (the varied agent), 1 being "Disagree" and 7 being "Agree". The two questions were presented in random order.

Results and discussion

The results of Experiment 1 can be found in Fig. 1, collapsed across the levels of intention. We conducted separate 2 (intent) X 2 (moral valence) ANOVAs for the fixed and varied agents. For the varied agent, there was only a main effect of moral valence, with agreement ratings being higher when the varied agent's actions were morally wrong (M = 5.68, SD = 1.81) than when they were not (M = 5.00, SD = 1.75), $F(1, 116) = 4.18$, $p = .04$, $\eta_p^2 = .04$.

For the fixed agent, there was also a main effect of moral valence, with lower agreement ratings when the *varied* agent's actions were morally wrong (M = 3.02, SD = 1.98) than when they were not (M = 4.78, SD = 1.74), $F(1, 113) = 26.74$, $p < .001$, $\eta_p^2 = .19$.

There was also a main effect of intent, with higher agreement ratings when the varied agent intended the specific outcome (M = 4.26, SD = 1.98) than when he did not (M = 3.52, SD = 2.08), $F(1, 113) = 4.60$, $p = .03$, $\eta_p^2 = .04$. While we did not specifically predict this, Lombrozo (2010) suggested that when an action is intentional rather than accidental, the counterfactual possibility that it does not occur seems less likely. Following from this, if people are less willing to consider the counterfactual possibility that the varied agent did not act, it becomes more important to the outcome whether the fixed agent acted or not, making them seem more causal. Regardless, our primary interest was in the causal supersession effect, and in that light this effect is largely irrelevant as there was no interaction between intent and moral valence, $F(1, 113) = .15$, $p > .5$.

In short, we demonstrated the causal supersession effect, with the causality of one agent affected by the normative status of another's actions. In addition to being the first experimental demonstration of causal supersession (to our knowledge), it confirmed one of the key predictions of the counterfactual account: Even when the outcome was something as trivial and morally neutral as possessing a paired set of bookends, causal supersession occurred. This confirms the first prediction of our counterfactual model, and goes beyond the predictions of motivational accounts.

Readers may wonder if these causal judgments are a zero-sum game, which would account for supersession without appealing to counterfactual reasoning. Under this view, any increase due to a norm violation must decrease the judged causality of other agents. There is evidence against the notion of zero-sum causal judgments in previous work (Lagnado, Gerstenberg, & Zultan, 2013), but also in our results. We examined the effect of norm violations on the summed causal ratings of the fixed and varied actors in this experiment. A zero-sum account predicts that there should be no difference between conditions, but in fact the sum of the two causal ratings was higher in the non-violation condition (M = 9.75, SD = 2.57) than the violation condition (M = 8.67, SD = 2.65), $F(1, 113) = 4.93$, $p = .026$.

Experiment 2

The counterfactual account predicts that A will only supersede B if A's action makes B's sufficiency more sensitive. However, in situations where B's sufficiency is independent of A, there should be no causal supersession.

Take a concrete example: Suppose that Billy and Suzy work together in the same office. Suzy is supposed to come in at 9 AM, whereas Billy has specifically been told not to come in at that time. The office has a motion detector, and the motion detector will be set off if it detects *two or more people* entering the room. Both Suzy and Billy arrive at 9am the next day, and the motion detector goes off. This case has the same basic structure as those examined in Experiment 1, and the counterfactual theory predicts that the scenario should produce the same supersession effect. Since Billy's action is bad, the possibility in which he doesn't act will be seen as highly relevant. Then, since Suzy's act would not be sufficient for the outcome in that possibility, her causality becomes more sensitive.

But now consider a slightly modified version of the case. What if, instead, the motion detector will be set off if it detects *one or more people* entering the room? In this case, either Suzy or Billy would be sufficient to bring about the outcome. Since Billy's action is bad, the possibility in which he doesn't act is seen as highly relevant. However, even in that possibility, Suzy's action is sufficient to bring about the outcome. Thus, we predict that Suzy's causality should be unaffected by Billy's actions if either individual action is sufficient to bring about the outcome.

The difference between these two scenarios comes down to a difference in their causal structures. In the first case, the causal structure is *conjunctive*, as the outcome requires the actions of both agents. In the second case, where we do not predict causal supersession, the causal structure is *disjunctive*, that is, the outcome can be generated by either agent (or both).

More abstractly, if the varied agent's actions are morally wrong, the possibility that the varied agent does not act becomes more relevant. However, if in that possibility the fixed agent can still bring about the outcome on her own, then her sufficiency is unaffected, and according to the counterfactual account, she should not be superseded.

We tested this prediction directly in Experiment 2 by manipulating the event structure such that the scenario was either disjunctive or conjunctive. We predicted that there

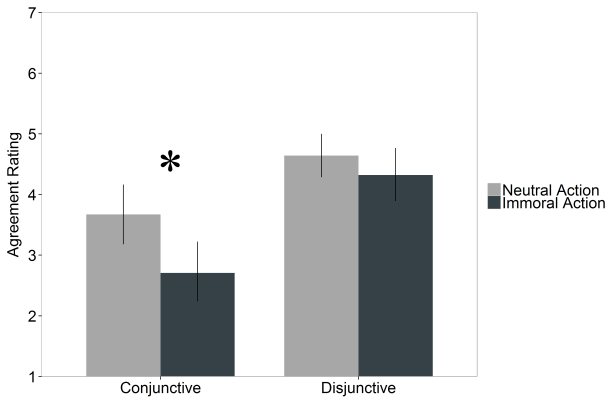

Figure 2: Mean agreement ratings as a function of causal structure and valence of the superseding actor's action. * = $p < .05$ Error bars indicate bootstrapped 95% CI.

would be a causal supersession effect in the conjunctive scenario but not in the disjunctive scenario.

Methods

Participants 240 participants were recruited via Amazon Mechanical Turk and paid $0.20 each for completing the two-question survey.

Materials and procedure We used the vignettes described above with Suzy as the fixed and Billy as the varied actor. These vignettes were manipulated along two dimensions. First, as in previous experiments, we manipulated the moral valence of the varied agent's actions, such that they were either neutral or wrong. Second, we manipulated the causal structure such that both the fixed and varied agent's actions were required to bring about the outcome (conjunctive), or either agent alone could bring about the outcome (disjunctive). In all conditions, participants were asked how much they agreed with the statement "Suzy caused the motion detector to go off", using the same 1-7 scale as in previous experiments. Following this, they were asked to complete a comprehension check: "Who was supposed to show up at 9am?" They could choose "Billy", "Suzy", or "Both of them."

Results and discussion

We excluded nine participants who failed the comprehension check, leaving 231 for analysis. Fig. 2 shows participants' mean agreement ratings as a function of the moral valence of the varied agent and the causal structure of the situation.

A 2 (moral valence) x 2 (causal structure) ANOVA revealed main effects of moral valence $F(1, 230) = 14.67$, $p < .001$, $\eta_p^2 = .06$, and causal structure, $F(1, 230) = 31.768$, $p < .001$, $\eta_p^2 = .12$, as well as a significant interaction between the two, $F(1, 230) = 11.58$, $p = .001$, $\eta_p^2 = .05$. Further analyses looked at the conjunctive and disjunctive structures separately. As predicted, there was a significant supersession effect in the conjunctive condition, with lower agreement ratings for the fixed agent when the varied agent's actions were morally wrong ($M = 2.46$, $SD = 1.87$) than when they were not ($M = 4.11$, $SD = 1.80$), $t(112) = 4.79$, $p < .001$. However, in the disjunctive condition, there was no such supersession effect. Agreement ratings did not differ when the varied agent's actions were immoral ($M = 4.53$, $SD = 1.76$) or neutral ($M = 4.62$, $SD = 1.54$), $t(118) = .32$, $p = .7$.

These results support the predictions of our counterfactual account of causal supersession: Causal supersession occurs only when the actions of one agent affect the sufficiency of the other agent's action.

Experiment 3

In Experiment 3, we aim to replicate the influence of causal structure on supersession and test another prediction of the counterfactual account. As discussed in the introduction, moral valence is just one example of a violation of norms. Any violation of norms, even non-moral ones, by the varied agent should make the specific possibility that those actions do not occur more relevant. Thus, according to the counterfactual account, we should also see causal supersession even when an event is seen as violating a purely statistical norm.

Methods

Participants 120 participants were recruited via Amazon Mechanical Turk and paid $0.20 each for completing the two-question survey.

Materials and procedure Experiment 3 followed the structure of Experiment 2 very closely, but with different content. The relevant outcome concerned a person named Alex winning a game that required a coin flip and a die roll of certain values. The fixed and varied factors were not agents' actions, but events that had different probabilities. The fixed event was a coin-flip, while the varied event was rolling two six-sided dice. The coin-flip was always a 50/50 chance, and therefore either result could be seen as normative, and the coin always came up heads. We manipulated the likelihood of the varied event by changing the minimum value of the dice roll required for Alex to win the game – higher than 2 (very likely) or higher than 11 (very unlikely). We also manipulated the event structure, such that both the coin flip and the die roll were necessary for Alex to win (conjunctive), or either one alone was sufficient (disjunctive). Participants were then asked how much they agreed with the statement, "Alex won because of the coin flip", on a 1-7 scale. They were additionally asked two comprehension check questions: "What did Alex need to roll higher than in order to win?" and "Which was more likely, that he would get heads on the coin flip or roll high enough on the dice roll?"

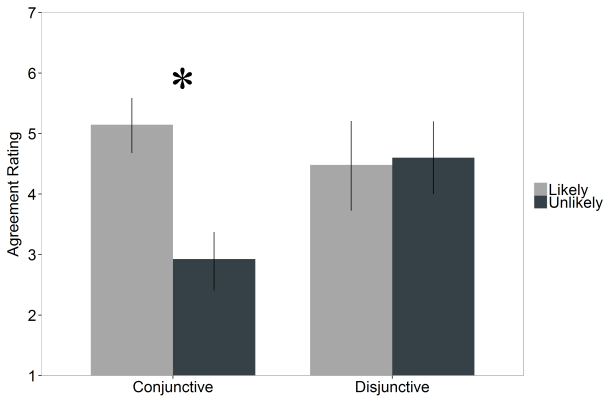

Figure 3: Mean agreement ratings as a function of the causal structure and the probability of the superseding event.
* = $p < .05$ Error bars indicate bootstrapped 95% CI.

Results and discussion

13 participants were excluded for having failed to correctly answer the comprehension questions, leaving 107 for analysis. Results for the remaining participants are displayed in Fig. 3. We conducted a 2 (likelihood) x 2 (causal structure) ANOVA. There was a main effect of likelihood, $F(1, 106) = 11.29$, $p = .001$, $\eta_p^2 = .096$, no main effect of causal structure, $F(1, 106) = 1.10$, $p = .30$, but critically there was once again an interaction between the two, $F(1, 106) = 15.79$, $p < .001$, $\eta_p^2 = .13$. As in Experiment 3, further analyses revealed that there was a supersession effect only in the conjunctive scenario. In the conjunctive condition, the coin flip was seen as less causal when the dice roll was unlikely ($M = 2.88$, $SD = 1.31$) than when it was likely ($M = 5.19$, $SD = 1.40$), $t(56) = 6.42$, $p < .001$. However, in the disjunctive condition, the coin flip was equally causal when the die roll was unlikely ($M = 4.46$, $SD = 1.79$) and likely ($M = 4.27$, $SD = 2.01$), $t(50) = -.36$, $p = .7$.

General Discussion

In three experiments, we demonstrated the novel phenomenon of causal supersession. Experiment 1 showed that supersession occurs regardless of the valence of the outcome, and that abnormal actions can be superseded. Experiment 2 provided evidence for a counterfactual account by manipulating causal structure, and demonstrated that causal supersession does not occur when the event is *disjunctive*. Experiment 3 replicated these results and further demonstrated that any violation of normality, even non-moral violations, can lead to a causal supersession effect.

Taken together, these studies provide strong evidence for the claim that the normality of one cause's action can influence the perceived causality of another cause. They also suggest that this phenomenon is best understood in terms of the impact of normality on the availability of different counterfactuals. Norm violations lead people to consider counterfactual possibilities in which those violations do not occur, and whether the sufficiency of other causal relationships hold in those possibilities can affect causal judgments for the actual world.

On the theory proposed here, the phenomenon of causal supersession should not be regarded as specific to norm violations. Anything that leads people to focus on particular counterfactual possibilities can, in principle, bring about the supersession effect. Indeed, one could use the supersession effect to test when particular counterfactuals are being considered. Causal supersession is a phenomenon of causal reasoning more generally, and a rich topic for further inquiry.

Acknowledgments

TG was supported by the Center for Minds, Brains and Machines (CBMM), funded by NSF STC award CCF-1231216. DL supported by ESRC grant RES-062330004. The authors would like to thank Scott Shapiro for his contributions to the early stages of the project.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*, 368-378.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*, 556-74.

Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *Journal of Philosophy*, *108*, 670-696.

Byrne, R. M. (2005). *The rational imagination: How people create alternatives to reality*. The MIT Press.

Halpern, J., & Hitchcock, C. (in press). Graded causation and defaults. *British Journal for the Philosophy of Science*.

Hart, H. L. A., & Honoré, T. (1985). *Causation in the law*. Oxford: Oxford University Press.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, 11, 587-612.

Kahneman, D. and D. Miller (1986), Norm theory: Comparing reality to its alternatives, *Psychological Review* 80: 136-153.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, *37*(6), 1036-73.

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*, 303-32.

McCloy, R., & Byrne, R. M. (2000). Counterfactual thinking about controllable events. *Memory & Cognition*, 28, 1071-1078.

N'gbala, A., & Branscombe, N. R. (1995). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology*, 31, 139-162.

Turri, J., & Blouw, P. (in press). Excuse validation: A study in rule-breaking. *Philosophical Studies*.

Woodward, J. (2006). Sensitive and insensitive causation. The Philosophical Review, 115, 1-50.