The Extent of Causal Superseding
Justin Sytsma

Abstract: Research indicates that norms matter for ordinary causal attributions. Across a range of cases in which two agents jointly bring about an outcome, with one violating a norm while the other does not, causal ratings are higher for the agent who violates the norm. Building off such findings, Kominsky et al. (2015) note a related phenomenon that they term “causal superseding”—whether or not one agent violates a norm also affects causal ratings for the other agent. Kominsky et al. offer an explanation of this phenomenon (the counterfactual sufficiency account) and describe the results of four experiments testing their account. In this paper, I explore the proposed phenomenon further. I present a sequence of new studies covering a range of cases, finding that the superseding effect is not as consistent as Kominsky et al. would predict, that the effect does not occur in all of the cases where they say it should, and that it sometimes occurs in cases where they say it shouldn’t. Finally, I offer an alternative deflationary explanation of causal superseding, presenting evidence suggesting that it is simply a context effect.

A growing body of research indicates that norms, especially injunctive norms, matter for ordinary causal attributions (e.g., Hilton and Slugoski 1986, Alicke 1992, Knobe and Fraser 2008, Hitchcock and Knobe 2009, Sytsma et al. 2012, Reuter et al. 2014, Kominsky et al. 2015, Livengood et al. 2017). Across a range of cases, when two agents jointly bring about an outcome, with one violating a norm while the other does not, causal ratings are higher for the agent who violates the norm. Further, causal ratings are higher when an agent violates a norm than when that agent does not violate a norm.

---

1 I want to heartily thank David Rose and Jonathan Livengood who were involved in the earlier stages of this project. I also want to thank Livengood for helpful feedback on an earlier draft of this paper. Finally, I want to thank the audiences at the AAP conference, Experimental Philosophy Group Germany Bi-annual Workshop, the Australasian Experimental Philosophy Conference, and the Empirical Philosophy Workshop at Victoria University of Wellington.

2 By “ordinary causal attributions” I specifically mean the use of language like “X caused Y” (see Sytsma et al. 2019 for further discussion). Some researchers arguably go further than this, asserting that norms matter not just for ordinary causal attributions, but for causal cognition more generally (Danks et al. 2014), but I will focus on just ordinary causal attributions here. Further, there is ongoing debate about which types of norms impact ordinary causal attributions, and specifically whether descriptive norms (often referred to as “statistical norms”) generally have an independent effect or whether they play a role in mediating injunctive norms (e.g., Knobe and Fraser 2008; Sytsma et al. 2012; Livengood et al. 2017). Injunctive norms include both prescriptive norms (what should be done) and proscriptive norms (what should not be done). In the existing literature, authors often use the expression “prescriptive norm” to refer to both prescriptive and proscriptive norms.

3 There is also evidence that this effect is not restricted to entities that philosophers would classify as agents (e.g., Hitchcock and Knobe 2009, Livengood and Sytsma under review), although I will focus on agents in this paper.
Building off such findings, Kominsky et al. (2015) note a related phenomenon that they term causal superseding: in situations in which two agents jointly bring about an outcome, whether or not one agent violates a norm also affects causal ratings for the other agent. More specifically, causal superseding occurs when causal ratings for the agent who does not violate the norm (the fixed agent) are lower when the other agent (the varied agent) violates a norm than when she does not violate a norm.\footnote{Using this terminology, the superseding effect might also be termed the fixed agent effect.} Kominsky et al. offer an explanation of causal superseding—the counterfactual sufficiency account—and note that it makes three clear predictions: (1) causal superseding should occur for good as well as bad outcomes; (2) causal superseding should occur when the actions of two agents are jointly necessary for bringing about the outcome, but not when either action is alone sufficient for bringing about the outcome; and, (3) causal superseding should occur for violations of descriptive norms as well as injunctive norms. They then report the results of four experiments demonstrating the effect and testing their three predictions.

Kominsky et al.’s findings on the causal superseding effect are important not simply because they show an interesting effect with regard to ordinary causal attributions, but because they speak to the larger debate concerning how to explain the effect of norms more generally. Specifically, Kominsky et al.’s results support one type of account in the literature, e.g. Hitchcock and Knobe’s intervention account, while raising difficulties for competing accounts, potentially including my own responsibility account. As such, understanding how the superseding effect works promises to illuminate the wider discussion.

In this paper, I explore the phenomenon of causal superseding further, beginning in Section 1 with a discussion of Kominsky et al.’s account and the evidence they present. In Section 2, I note that this phenomenon can be found in earlier work in the literature, although it does not arise in all cases (raising concerns about the extent of the effect) and for all types of
descriptive norms (raising concerns about Kominsky et al.’s third prediction). In Section 3, I then present the results of a sequence of five new studies investigating the proposed effect for a range of cases. I argue that the results cast doubt on both the generality of the effect and the counterfactual sufficiency account. In Section 4, I offer a deflationary explanation of causal superseding, arguing that what we are seeing (when the effect occurs at all) is a context effect, with the perceived causal strength of the agents relative to one another affecting how participants interpret the scale used in eliciting their judgments. This is illustrated by a pair of studies showing that a comparable “superseding effect” can be found for height judgments. I then show that the strength of the causal superseding effect varies as a function of the scale used. Finally, in Section 5, I consider Kominsky et al.’s second prediction, presenting the results of two cases in which we find a superseding effect even though sufficiency is not threatened.

1. The Causal Superseding Effect

Causal superseding occurs when the relative causal ratings for one agent are affected in a specific way by changes in the normative status of another agent’s actions. To better understand the proposed effect, let’s begin by considering the cross-agent effect of normative status on ordinary causal attributions that has been shown most often in the literature. In the typical case, two agents perform symmetric actions, jointly bringing about an outcome, with each of the two actions being described as individually necessary for the outcome and the two actions together being described as jointly sufficient for the outcome. What we find in such conjunctive cases is that when one agent’s action violates an injunctive norm and the other agent’s action does not, the agent violating the norm receives higher causal ratings than the agent who did not violate the norm.

For instance, Knobe and Fraser (2008) presented participants with a story in which two agents perform symmetric actions, leading to a problem. In their Pen Case both the
administrative assistants and faculty members in a philosophy department help themselves to pens from the secretary’s desk, but while administrative assistants are allowed to take the pens, faculty members are not supposed to take the pens. One day, both Professor Smith and an administrative assistant take a pen. Later that day, the secretary needs to take an important message but there are no pens left. Participants were then asked whether Professor Smith caused the problem and whether the administrative assistant caused the problem. Despite the two agents performing symmetric actions, participants were significantly more likely to affirm that Professor Smith caused the problem than that the administrative assistant did, and the difference between the mean ratings was striking (3.4 points on a 7-point scale). This difference shows the cross-agent effect.

Now imagine that we repeat the experiment, but this time we specify that neither agent violates a norm. Looking across the two cases, we see that the normative status of one agent’s action is varied (in one scenario the agent violates a norm, in the other the agent does not violate a norm) while the normative status of the other agent’s action is fixed (the agent does not violate a norm in either scenario). Comparing causal ratings across the two scenarios, there are now two further effects that we might see. First, we might find that causal ratings for the varied agent are higher in the scenario in which that agent violates the norm than in the scenario in which that agent does not violate the norm. Call this the varied agent effect. And, in fact, such an effect has also been reported in the literature. For instance, in Sytsma et al. (2012) we found a varied agent effect for the Pen Case. We ran a second version of the case in which there was no departmental policy with regard to taking pens. Comparing the causal ratings for Professor Smith between the two scenarios, we found that participants were significantly more likely to affirm that he caused the problem when he violated a norm than when he did not (the difference between the mean ratings was 1.05 on a 7-point scale).
Second, we could compare the ratings for the fixed agent between the two scenarios. Given that the fixed agent’s action is described the same way in each scenario—given that she doesn’t violate a norm in either case—we might expect that the causal ratings will be roughly the same. Kominsky et al., however, present results in which the ratings for the fixed agent are significantly different between the two scenarios, with the ratings being higher in the scenario where the varied agent does not violate a norm than in the scenario where the varied agent does violate a norm. And as described in Section 2, such an effect is also sometimes found in the previous literature. Kominsky et al. are the first to highlight this effect, however, terming it the *superseding effect.*

1.1 Two Explanations

Assuming that the superseding effect reliably occurs, how should we explain it? A number of explanations of the cross-agent effect and the varied agent effect have been offered in the literature, including Hitchcock and Knobe’s (2009) *intervention account* and my *responsibility account* (Sytsma et al. 2012). While other accounts could be discussed here, including Alicke’s (1992) bias view and the pragmatic account put forward by Samland and Waldmann (2016), I will focus on just these two views—the intervention account for purposes of laying the foundation for Kominsky et al.’s counterfactual sufficiency account and the responsibility account as a foil.

The intervention account explains the cross-agent effect in terms of the cognitive processes that generate ordinary causal attributions taking normality judgments into account. Hitchcock and Knobe argue that causal attributions serve to identify suitable intervention points. Information about norms then plays a role because norms are relevant to identifying suitable intervention points in some situations. The basic idea is that when people assess intervention
points, they consider counterfactuals on which the outcome does not occur, and norms play a role in which counterfactuals they consider—people being more likely to consider a counterfactual in which an abnormal event is replaced with a more normal one. Based on this account, Hitchcock and Knobe argue that the effect of norms on ordinary causal attributions does not depend on either the outcome valence or the type of norm that is violated.

Kominsky et al.’s counterfactual sufficiency account extends the intervention account to offer an explanation of the superseding effect. As with the intervention account, the counterfactual sufficiency account starts with the claims that norm violations make counterfactuals in which the norm was not violated salient and that people are more likely to consider salient counterfactuals. They then focus on the sufficiency condition for a causal relation. The idea is that when this condition holds, the occurrence of an event is sufficient for the occurrence of the outcome (if the event occurs, then the outcome occurs). To this Kominsky et al. add the notion of sensitivity (Woodward 2006): the more likely it is that a causal condition would cease to hold if the background conditions were slightly different, the more sensitive it is. Putting these together, the idea is that in a scenario like the Pen Case, people recognize that Professor Smith did something abnormal (he broke the injunctive norm against faculty members taking pens). This makes the counterfactual in which he does something more normal instead (doesn’t take a pen) salient, such that people are likely to consider this counterfactual. And if Professor Smith didn’t take a pen, then the outcome would not have occurred. As such, considering this counterfactual highlights the sensitivity of the sufficiency condition for the administrative assistant: if Professor Smith did not take a pen, then the administrative assistant taking a pen would not lead to the problem. Finally, following Woodward, Kominsky et al. argue that when a sufficiency condition is judged to be highly sensitive, people are reluctant to attribute causation.
Taking these parts together, the counterfactual sufficiency account is able to explain the occurrence of the causal superseding effect.

As Kominsky et al. note, the counterfactual sufficiency account makes three clear predictions. First, it predicts that the superseding effect should occur regardless of the valence of the outcome. The reason is that what matters for this account is the normality of the actions, not whether the outcome is good or bad. Second, the superseding effect should only occur when the sufficiency condition is threatened. This means that it should occur in *conjunctive cases* like the Pen Case where either action alone is insufficient to bring about the outcome but should not occur in *disjunctive cases* such as a modified version of the Pen Case, for instance, in which *either* Professor Smith *or* the Administrative Assistant taking a pen would lead to the outcome. Third, the superseding effect should occur for any norm violation, not just violations of injunctive norms. This follows from the intervention account, which holds that what matters for the saliency of a counterfactual is the overall abnormality of the action, not specifically that it is abnormal with regard to an injunctive norm or a descriptive norm.

The responsibility account put forward by Sytsma et al. (2012) explains the cross-agent effect in a rather different way than the intervention account.\(^5\) The responsibility account holds that the concept of causation that people typically use when they make causal attributions is itself a normative concept, akin to a thick ethical concept.\(^6\) The idea is that causal attributions typically serve to indicate something more than simply that an agent brought about an outcome; they also express a normative evaluation of the action similar to saying that the agent is responsible or

---

\(^5\) See Livengood et al. (2017), Sytsma et al. (forthcoming), Livengood and Sytsma (under review), and Sytsma and Livengood (under review); see Sytsma and Livengood (2015) for a brief discussion under the label of the descriptive program in experimental philosophy.

\(^6\) More cautiously, “people” should be restricted throughout to native English-speaking adult Americans—the population that most (but not all—e.g., Samland and Waldmann 2016) of the participants in the relevant studies have been drawn from. The extent to which these findings generalize to other English-speakers and, especially, to attributions in other languages is an interesting open question.
accountable for the outcome. According to the responsibility account, the cognitive processes involved in making most causal attributions start with a normatively-laden concept. This contrasts with the intervention account, which holds that the concept involved is descriptive, but that norms come into play via the counterfactuals that we consider.

Like the intervention account, the responsibility account does not directly predict the occurrence of the superseding effect. That said, the responsibility account is compatible with the occurrence of superseding effects. And it predicts the occurrence of superseding effects when coupled with the hypothesis, discussed in greater detail in Section 4, that the superseding effect is a type of context effect, arising from relative differences in the perceived causal strength of the agents affecting how participants interpret the scale used in eliciting causal attributions. If this hypothesis is correct, then insofar as the responsibility account explains the cross-agent effect, it can also explain the superseding effect.

The responsibility account, coupled with this context effect hypothesis, makes different predictions from the causal sufficiency account with regard to when we are likely to see the superseding effect occur. In contrast to Kominsky et al.’s first prediction, I would expect superseding effects to be more pronounced for bad outcomes than good outcomes. While the responsibility account does not make a direct prediction about the role of outcome valence on ordinary causal attributions, it does make a prediction when coupled with the observation that people are more likely to judge that an agent is responsible for a bad outcome than a good outcome (Sytsma et al. 2019). In contrast to Kominsky et al.’s second prediction, the responsibility account does not specifically predict that superseding effects will be absent in disjunctive cases. This would depend on whether people tend to judge that an agent who violates a relevant injunctive norm in a disjunctive case is more accountable for the outcome than an agent who does not violate the norm. While my intuition is that this will often be the case,
empirical testing is needed. Nonetheless, conditional on my intuition being correct, we would expect injunctive norms to continue to affect causal attributions in at least some disjunctive cases, and hence would expect that it is possible to get a superseding effect despite sufficiency not being threatened. And, in contrast to Kominsky et al.’s third prediction, the responsibility account does not predict that each type of norm will produce the same effects. Our account focuses on injunctive norms, although it allows that descriptive norms will sometimes also have an indirect effect on ordinary causal attributions by influencing people’s judgments about the relevant injunctive norms (Sytsma et al. 2012).

Finally, my account makes at least two new predictions. First, since I hold that the superseding effect is a byproduct of the cross-agent effect, I predict that it will only reliably occur when the cross-agent effect occurs. Second, in keeping with the literature on context effects, I would not be surprised to find that the superseding effect is somewhat fickle. As discussed in Section 4, the occurrence of context effects is often hard to predict. I would expect the same for the superseding effect across a range of scenarios and manipulations.

1.2 Four Experiments

Kominsky et al. report the results of four experiments indicating that the superseding effect occurs and providing initial support for each of the three predictions made by the counterfactual sufficiency account. Their first experiment has two conditions involving a fixed agent and a varied agent jointly procuring a matched pair of bookends. In each condition participants rated agreement with a causal attribution for each agent on a 7-point scale. As expected, Kominsky et al. found a cross-agent effect: in the condition where the varied agent violated an injunctive norm, the mean rating for the varied agent (M=5.97) was notably higher than the mean rating for the fixed agent (M=4.40). They also found a varied agent effect, with the mean rating for the varied
agent being higher when he violated the norm than when he did not (M=4.80). Finally, they found a superseding effect: the mean rating for the fixed agent when the varied agent violated the norm (M=3.37) was 1.03 points lower than when the varied agent did not violate the norm (M=4.40). Further, this experiment was successfully replicated as part of the experimental philosophy replication project (Cova et al. 2018), with the replication team finding a significant superseding effect of the same magnitude.

In their second experiment, Kominsky et al. again gave people scenarios with a fixed agent and a varied agent. This time they also varied the outcome valence, with the agents either jointly bringing about a good outcome or a bad outcome. Again, in each of the four conditions participants rated agreement with a causal attribution for each agent on a 7-point scale. While Kominsky et al. do not separately report the means for each condition, it is apparent from their Figure 3 that they find a cross-agent effect, varied agent effect, and superseding effect for both the good and the bad outcome. With regard to the superseding effect for the good outcome conditions, the mean rating for the fixed agent when the varied agent violated the norm (M=2.90) was 1.37 points lower than when the varied agent did not violate the norm (M=4.27). Combining the good and the bad outcome, the mean rating for the fixed agent when the varied agent violated the norm was 1.99 points lower (M=2.13) than when the varied agent did not violate the norm (M=4.12). Thus, we see that the superseding effect was notably larger when the outcome was bad than when the outcome was good. In fact, the effect looks to be roughly twice as large.

Looking at the data overall, an ANOVA showed no significant effect of outcome valence on ratings for the varied agent, while a second ANOVA showed a significant effect of outcome valence on ratings for the fixed agent and a marginal interaction effect between outcome valence and whether the varied agent violated a norm. Kominsky et al. also asked participants to rate the outcome valence. A regression looking at causal ratings for the fixed agent showed that both
ratings of outcome valence and the normative status of the varied agent’s action were significant predictors, but there was not a significant interaction. Kominsky interpret this to show that “judged outcome valence… did not alter the causal superseding effect” (202). And from this they conclude that “while outcome valence did have some impact on the causal ratings of the fixed agent, it did not impact the causal superseding effect, that is, the effect of the moral status of the varied agent’s actions on the fixed agent’s causality” (202).

This conclusion is tough to square with the fact that the superseding effect is much more pronounced in the conditions where the outcome is good than the conditions where the outcome is bad, however. Taking the superseding effect to be the difference in mean causal ratings for the fixed agent between the condition where the varied agent violates a norm and the condition where the varied agent does not violate a norm, the superseding effect was again roughly twice as large for the bad outcome conditions than for the good outcome conditions. Hitchcock and Knobe make a strong prediction with regard to outcome valence in laying out their intervention account. They write that their account “suggests that the impact of normative considerations should remain unchanged (because people still see that a norm has been violated)” (2009, 603). This prediction is not supported by the results of Kominsky et al.’s second experiment. Kominsky et al. make a weaker prediction for the causal sufficiency account, however, simply predicting that the superseding effect should occur regardless of the valence of the outcome. And it does in this experiment. Nonetheless it is unclear how either the intervention account or the causal sufficiency account are to explain the differences in causal ratings between the good outcome conditions and the bad outcome conditions. In contrast, as noted above, the responsibility account predicts that causal ratings will tend to be more pronounced for bad outcomes, and I make the same type of prediction for the superseding effect.
In Kominsky et al.’s third experiment, they test the second prediction of the causal sufficiency account—that the superseding effect should occur in conjunctive cases but not in disjunctive cases. As before they used vignettes in which there is a fixed agent and a varied agent, but this time they also varied whether the causal structure was conjunctive or disjunctive. Further, this time they only asked participants to rate agreement with a causal attribution for the fixed agent. For this experiment, the vignettes they used began by stating that “Suzy and Billy are working on a project that is very important for our nation’s security.” Participants are then told either that the boss tells both agents to be in a particular room at exactly 9am the next day or that the boss tells Suzy to be in a particular room at exactly 9am the next day and tells Billy not to come in at all that morning. The vignettes go on to state that there is a motion detector installed in the room and that it is either triggered to go off if more than one person is in the room (conjunctive) or if at least one person is in the room (disjunctive). In line with their second prediction, Kominsky et al. found a significant superseding effect for the conjunctive conditions (difference in means of 1.65 points), but not for the disjunctive conditions (difference in means of 0.09 points).

As discussed above, I tentatively expect that norms will sometimes have an effect on causal attributions in disjunctive cases. So, the results of Kominsky et al.’s third experiment potentially put some strain on the responsibility account, calling for an explanation of why the effect does not occur in this case. I believe that there is a ready explanation. To see this, note that in each version of the scenario used in Experiment 3 (the Motion Detector Case), Suzy is told that it is absolutely essential that she show up at 9am. In the versions in which Billy follows the injunctive norm, he too is told that it is absolutely essential that he show up at 9am. In the versions in which Billy violates the injunctive norm, however, he is told that it is absolutely essential that he not show up at 9am. Further, note that in the conjunctive versions, the motion
detector will only go off if *both* Suzy and Billy show up at 9am. So, in the conjunctive version where Billy follows the injunctive norm, the implication is that it is very important for national security that the motion detector goes off, from which we can infer that the motion detector going off is a *good outcome*. In the conjunctive version where Billy violates the injunctive norm, however, he was told not to show up at 9am, from which we can infer that it is very important for national security that the motion detector does not go off, and hence that the motion detector going off is a *bad outcome*.

Now consider the disjunctive versions. In these versions the motion detector will go off if *either* Suzy or Billy shows up. And in each condition, Suzy is told that it is absolutely essential that she show up at 9am. So, in each version, it is plausible that the boss wants the motion detector to go off. As such, we can reasonably infer that it is very important for national security that the motion detector goes off and, consequently, that the motion detector going off is a *good outcome* in each case.7

What should the responsibility account say about Kominsky et al.’s third experiment? Our account explains the effect of injunctive norms on causal attributions in terms of an agent being responsible for the outcome. But as noted above, we expect people to be more reticent to assign responsibility for a good outcome than a bad outcome. Accepting the logic noted above, in the conjunctive versions of the Motion Detector Case we can infer that the outcome is good when Billy is told to show up and can infer that the outcome is bad when Billy is told to not show up. As such, we would expect causal ratings for Billy to be higher in the condition in which he violates the norm, so we would expect there to be both a cross-agent effect in this condition and a varied agent effect across the two conjunctive conditions. Based on this, I would not be

---

7 The inference that the motion detector going off is important for national security is arguably less clear in the disjunctive scenarios than in the conjunctive scenario where Billy is told to show up, because it is unclear why the boss issues an instruction to Billy, since Suzy is told to show up in either case, and since Suzy showing up is sufficient to set up the motion detector.
surprised to find a superseding effect here. What about for the disjunctive conditions? In the disjunctive versions of the Motion Detector Case we can plausibly infer that the outcome is good in both versions. As such I would expect the cross-agent effect to be weaker in the disjunctive condition where Billy violates the instruction (good outcome) than in the conjunctive condition where Billy violates the instruction (bad outcome). And since my account ties the superseding effect to the cross-agent effect, I would expect the superseding effect to be weaker as well.

In their fourth experiment, Kominsky et al. seek to test the third prediction of the counterfactual sufficiency account—that the superseding effect should occur for any norm violation, including violations of descriptive norms. To do this they looked at a fixed event and a varied event, manipulating the statistical probability of the varied event. Unfortunately, for present purposes it is unclear what conclusions we should draw from this experiment. The issue is that Kominsky et al. didn’t assess ordinary causal attributions, but “because” statements (e.g., “Alex won because of the coin flip”) and it is unclear whether such statements work analogously to causal attributions in this type of scenario. As such, I will leave this experiment to the side in the remainder of the paper.

2. Previous Findings

Kominsky et al. describe the superseding effect as “an exciting discovery,” although they find precedent for it in legal findings (205). While they are the first researchers that I am aware of to systematically describe and attempt to explain this effect, it is worth noting that the superseding effect itself can be found in the previous empirical literature. This is worth noting in large part because the effect is not found as consistently as we would expect based on Kominsky et al.’s

---

8 See Livengood and Machery (2007) and Livengood et al. (2017, fn39) for evidence that “X caused Y” and “Y because X” statements sometimes come apart.
experiments, and it is not found for all types of descriptive norms, which runs counter to their third prediction. In this section I briefly summarize findings from studies in two other articles where Kominsky et al. would expect to see superseding effects.9

2.1 Pen Case

In Sytsma, Livengood, and Rose (2012) we reported results for a number of variations on the Pen Case described above. This included varying whether Professor Smith violates an injunctive norm in taking a pen and whether he violates one of two types of descriptive norms—a population-level statistical norm (faculty members typically do not take pens) or an individual-level statistical norm (Professor Smith typically does not take pens). In contrast, the administrative assistant is described as following the relevant injunctive and descriptive norms in each case. Since these were all conjunctive cases, Kominsky et al. predict that we should find a superseding effect for the administrative assistant across each relevant pair of conditions. To test this, I reanalyzed the data focusing on the superseding effect. Results for these comparisons are summarized in Table 1.

Comparing the cases in which faculty members are described as typically taking pens while varying whether Professor Smith violates a department policy in taking a pen, I found a causal superseding effect that is comparable in size to that reported in Kominsky et al.’s first experiment: causal ratings for the administrative assistant were significantly higher when Professor Smith did not violate an injunctive norm (M=3.53, SD=2.01, N=45) than when he

---

9 In addition, Samland and Waldmann (2016) report a superseding effect in their Experiment 2a with an effect size in the range found by Kominsky et al. (Cohen’s d=0.614). Kirfel and Lagnado (2017) also find a superseding effect for an injunctive norm in their first experiment (decrease in mean ratings of 1.79 points on a 10-point scale across the relevant conditions) and a small superseding effect for violation of an individual-level statistical norm (decrease in mean ratings of 0.73 points on the same scale). In their second experiment, however, they found a reverse effect for violation of an individual-level statistical norm, showing an increase of 1.61 points on the 10-point scale across the relevant conditions. The reverse effect was stronger when they used a cumulative causal structure than a conjunctive causal structure, however.
violated the injunctive norm (M=2.51, SD=1.81, N=59), giving a drop of 1.02 points on the 7-point scale, which corresponds with a medium effect size, t(89.53)=-2.69, p=0.0085, Cohen’s d=-0.54. When we changed the population-level statistical norm so that Professor Smith now violates this descriptive norm as well as the injunctive norm in taking a pen, however, the causal ratings for the administrative assistant were higher than when Professor Smith just violated the injunctive norm (M=2.93, SD=1.68, N=41). While Kominsky et al. would predict a larger superseding effect here—Professor Smith now violating two norms—the difference was no longer significant and the effects size was small, showing a drop of just 0.61 points, t(83.42)=-1.53, p=0.13, Cohen’s d=-0.33. Following Cohen (1977), this comparison is adequately powered to detect an effect size in the range that Kominsky et al. report. They gave effect sizes for three relevant t-tests, showing a Cohen’s d of 0.568 in Experiment 1, 0.691 for the good outcome conditions in Experiment 2, and 0.898 in Experiment 3, for an average value of 0.719 (I’ll refer to this as an average superseding effect below). A power analysis reveals that the test had a power of 0.91 to detect an average superseding effect, with a power of 0.74 to detect an effect with the smallest size they reported and a power of 0.98 to detect an effect with the largest size they reported.

When we switch from the population-level statistical norm to an individual-level statistical norm, however, things get even worse for Kominsky et al.’s account. Comparing the cases in which Professor Smith is described as typically taking pens while varying whether he violates a department policy in taking a pen, the effect runs in the opposite direction: causal ratings for the administrative assistant were higher when Professor Smith violated the injunctive norm (M=4.03, SD=1.86, N=70) than when he did not (M=3.68, SD=1.86, N=41), a rise of 0.35 points, although the effect size is negligible, t(83.95)=0.94, p=0.35, Cohen’s d=0.19. This test had a power of 0.95 to detect an average superseding effect. The size of the reverse effect was larger still when we changed the individual-level statistical norm so that Professor Smith violates
this descriptive norm by taking a pen. When Professor Smith was described as violating both the injunctive norm and the individual-level statistical norm, causal ratings for the administrative assistant were significantly higher \((M=4.81, SD=2.18, N=53)\), a rise of 1.13 points, which corresponds with a medium effect size, \(t(91.08)=2.70, p=0.0083\), Cohen’s \(d=0.55\).

2.2 Lauren and Jane Case

Sytsma et al. (2012) opened with a description of another type of conjunctive case, the Lauren and Jane Case based on a thought experiment given by Knobe (2006).\(^{10}\) In this case, Lauren and Jane are employees of a company that uses a mainframe computer that can be accessed from terminals on different floors of its building. The mainframe has become unstable, however, such that it will crash if more than one person logs in at the same time. Lauren and Jane both log in and the system crashes. As described in our (2012), citing a then unpublished manuscript, when Jane but not Lauren violates an injunctive norm by logging in, people are significantly more likely to affirm that Jane caused the system to crash than Lauren.

In the full set of results reported in Livengood, Sytsma, and Rose (2017), we also tested variations in which there was no injunctive norm in place, allowing for further tests of the superseding effect. Again, we did not conduct those tests in the original article but I have reanalyzed the data. Results are summarized in Table 1. Comparing causal attributions for Lauren between the version in which no injunctive norm was in place \((M=2.70, SD=2.14, N=71)\) and the version in which Jane violated an injunctive norm \((M=2.42, SD=2.04, N=72)\), I found a slight drop of 0.29 points on a 7-point scale, which corresponds with a negligible effect size, \(t(140.46)=-0.82, p=0.41\), Cohen’s \(d=-0.14\). This test had a power of 0.99 to detect an average

\(^{10}\) This was also the basis for the cases used by Reuter et al. 2014, as well as the Email Case used in Kominsky et al.’s second experiment.
superseding effect. When these two cases are run asking each participant about just one agent rather than both, however, the effect runs in the opposite direction, showing a slight rise of 0.053 points (M=2.36, SD=1.56, N=39 vs M=2.41, SD=1.71, N=34), which corresponds with a negligible effect size, \( t(67.55)=0.14, p=0.89, \) Cohen’s \( d=0.032 \). This test had a power of 0.86 to detect an average superseding effect.

When we include a question about the mainframe and ask each participant about both agents, there was a significant superseding effect, with the mean rating for Lauren dropping 1.20 points between the condition in which no injunctive norm was in place (M=2.74, SD=1.87, N=62) and the condition in which Jane violates an injunctive norm (M=1.54, SD=1.16, N=54), which corresponds with a medium effect size, \( t(103.52)=-4.22, p=5.2e^{-5}, \) Cohen’s \( d=-0.76 \). We also asked participants about Lauren or Jane’s action of logging into the terminal. When we asked each participant about both agents there is a borderline significant superseding effect with a drop of 0.80 points (M=3.33, SD=2.31, N=43 vs. M=2.52, SD=2.06, N=48), which corresponds with a small effect size, \( t(84.83)=-1.75, p=0.084, \) Cohen’s \( d=-0.37 \). This test had a power of 0.92 to detect an average superseding effect. Asking each participant about just one agent, the effect was non-significant with a drop of 0.61 (M=3.12, SD=1.99, N=41 vs. M=2.51, SD=1.94, N=43), which corresponds with a small effect size, \( t(81.58)=-1.42, p=0.16, \) Cohen’s \( d=-0.31 \). This test had a power of 0.90 to detect an average superseding effect.

Finally, we tweaked the vignettes to describe the “instability” of the mainframe as a feature rather than a bug. Asking each participant about both agents there was a significant superseding effect, with a drop of 1.10 points (M=3.18, SD=2.24, N=66 vs. M=2.08, SD=1.83, N=52), which corresponds with a medium effect size, \( t(115.8)=-2.95, p=0.0039, \) Cohen’s \( d=-0.53 \). And asking each participant about just one agent there was again a significant superseding
effect, with a drop of 1.33 points ($M=3.05$, $SD=2.29$, $N=40$ vs. $M=1.72$, $SD=1.37$, $N=36$), which corresponds with a medium effect size, $t(64.7)=3.11$, $p=0.0028$, Cohen’s $d=0.70$.

<table>
<thead>
<tr>
<th>Studies</th>
<th>Norms Violated by Varied Agent</th>
<th>Superseding Effect</th>
<th>p-value</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &amp; 2 (2012)</td>
<td>Injunctive</td>
<td>-1.02, Medium**</td>
<td>0.0085</td>
<td>-0.54</td>
</tr>
<tr>
<td>1 &amp; 2 (2012)</td>
<td>Injunctive, Population-level</td>
<td>-0.61, Small</td>
<td>0.13</td>
<td>-0.33</td>
</tr>
<tr>
<td>3 &amp; 4 (2012)</td>
<td>Injunctive</td>
<td>R: 0.35, Small</td>
<td>0.35</td>
<td>0.19</td>
</tr>
<tr>
<td>3 &amp; 4 (2012)</td>
<td>Injunctive, Individual-level</td>
<td>R: 1.13, Medium**</td>
<td>0.0083</td>
<td>0.55</td>
</tr>
<tr>
<td>1 &amp; 9 (2017)</td>
<td>Injunctive</td>
<td>-0.29, Negligible</td>
<td>0.41</td>
<td>-0.14</td>
</tr>
<tr>
<td>2 &amp; 10 (2017)</td>
<td>Injunctive</td>
<td>R: 0.053, Negligible</td>
<td>0.89</td>
<td>0.032</td>
</tr>
<tr>
<td>3 &amp; 13 (2017)</td>
<td>Injunctive</td>
<td>-1.20, Medium***</td>
<td>0.000052</td>
<td>-0.76</td>
</tr>
<tr>
<td>7 &amp; 11 (2017)</td>
<td>Injunctive</td>
<td>-0.80, Small</td>
<td>0.084</td>
<td>-0.37</td>
</tr>
<tr>
<td>8 &amp; 12 (2017)</td>
<td>Injunctive</td>
<td>-0.61, Small</td>
<td>0.16</td>
<td>-0.31</td>
</tr>
<tr>
<td>5 &amp; 15 (2017)</td>
<td>Injunctive</td>
<td>-1.10, Medium**</td>
<td>0.0039</td>
<td>-0.53</td>
</tr>
<tr>
<td>6 &amp; 16 (2017)</td>
<td>Injunctive</td>
<td>-1.33, Medium**</td>
<td>0.0028</td>
<td>-0.70</td>
</tr>
</tbody>
</table>

Table 1: Summary of comparisons looking for causal superseding effects for the studies in Sytsma et al. (2012) and Livengood et al. (2017); R denotes that the difference in means was the reverse of that predicted.

Summing up, in Sytsma et al. (2012) and Livengood et al. (2017), there is some evidence for a superseding effect for injunctive norms. Leaving out the cases where we manipulated whether the varied agent also violated a descriptive norm, there was a significant superseding effects in four out of nine comparisons. Across these nine comparisons, the effect sizes ranged from negligible to medium, and in two of the comparisons the difference in the means went in the reverse direction of the superseding effect. Looking at the cases where we also manipulated whether the varied agent violated a descriptive norm, the results do not support Kominsky et al.’s third prediction: there was a non-significant superseding effect when we also varied a population-level statistical norm, although the effect size was smaller than when we varied just the injunctive norm, and there was a significant reverse effect when we also varied an individual-level statistical norm.
3. New Findings

Looking across the results for the two conjunctive cases discussed in the previous section, we see variation both in whether the causal superseding effect occurs and the strength of the effect when it does occur. The variability seen for the causal superseding effect suggests that more research is needed. In this section, I describe the results of a series of new experiments testing the superseding effect for five conjunctive cases.

3.1 Study 1: Pen Case

As noted above, our goal in Sytsma et al. (2012) was not to test the causal superseding effect, and our studies were not designed with this in mind. One potential issue for purposes of testing superseding is that in the cases in which Professor Smith did not violate an injunctive norm, we did not specify that either agent was specifically allowed to take pens, instead stating that they were “able” to do so and making no mention of a departmental policy concerning the taking of pens. In contrast, in the vignettes used by Kominsky et al., arguably the normative status of the varied agent’s action is clearer (e.g., Bill buying the bookend in Experiment 1). In addition, we did not test the cases in which Professor Smith violates a descriptive norm but not an injunctive norm. As such, I begin by repeating the studies of the Pen Case with these modifications.

3.1.1 Methods

Participants for each study presented in this paper were recruited through advertising for a free personality test on Google. After answering the causal questions reported below, participants answered basic demographic questions and took a 10-item Big Five personality inventory.

Participants for each study were restricted to native English-speakers, 16 years of age or older,
who completed the survey, and reported that they had not taken a survey through the website previously. Results for Study 1 were collected from 327 participants.\(^\text{11}\)

Participants were given one of eight vignettes featuring a varied agent (Professor Smith) and a fixed agent (administrative assistant), as described in Section 1. Where not fully specified in the text, full vignettes for each of the new scenarios in this paper are given in the online supplemental materials. In the first four vignettes I varied whether Professor Smith violated an injunctive norm and whether he violated a population-level statistical norm; in the last four vignettes I varied whether Professor Smith violated an injunctive norm and whether he violated an individual-level statistical norm. In each vignette, administrative assistants were allowed to take pens, and in the first four vignettes administrative assistants were said to typically take pens, while in the last four vignettes the administrative assistant (now named John) was said to typically take pens. In each condition, participants were asked to rate how much they agreed or disagreed with two causal statements—“Professor Smith caused the problem” and either “the Administrative Assistant caused the problem” (first four conditions) or “John caused the problem” (last four conditions)—on a 7-point scale anchored at 1 with “strongly disagree,” at 4 with “neutral,” and at 7 with “strongly agree.”

3.1.2 Results

Results are shown in Figure 1. For the probes specifying a population-level descriptive norm (first four conditions) I ran two ANOVAs, one looking at agreement ratings for the varied agent and one looking at agreement ratings for the fixed agent, with whether the varied agent violated the injunctive norm and whether he violated the descriptive norm as between-participants factors.

\(^{11}\) 73.7% women (three non-binary), average age 38.3 years, ranging from 16 to 77. Given the large percentage of woman participants, I checked for gender effects, running each of the four ANOVAs described in 3.1.2 with gender (all tests for gender effects were run treating gender as a binary—either women or non-women) as a third between-participants factor. No significant gender effects were found.
Starting with ratings for the varied agent, as expected I found a significant effect for violation of the injunctive norm, $F(1, 158)=34.36, p=2.6e-8$. In line with Sytsma et al. (2012), I found no effect for violation of the population-level statistical norm, $F(1, 158)=0.16, p=0.69$, and no interaction effect, $F(1, 158)=0.72, p=0.40$. Follow-up t-tests showed significant cross-agent effects with medium effect sizes in the condition where Professor Smith violated just the injunctive norm, $t(39)=2.13, p=0.040$, Cohen’s $d=0.56$, and the condition where Professor Smith violated both norms, $t(39)=2.77, p=0.0085$, Cohen’s $d=0.71$, but showed the reverse effect in the condition where he violated just the descriptive norm, $t(41)=-3.49, p=0.0011$, Cohen’s $d=-0.54$, with the mean rating being higher for the administrative assistant than for Professor Smith.

Turning to the varied agent effect, follow-up t-tests showed significant effects with large effect sizes when the injunctive norm was varied: comparing the condition where Professor Smith violates just the injunctive norm to the condition where he violates neither, $t(74.96)=3.59, p=0.00060$, Cohen’s $d=0.80$; comparing the condition where he violates both norms to the condition where he violates neither, $t(71.85)=4.22, p=7.1e-5$, Cohen’s $d=0.94$; and, comparing the condition where he violates both norms to the condition where he violates just the descriptive norm, $t(69.31)=4.65, p=1.56e-5$, Cohen’s $d=1.04$. By contrast, I did not find significant effects when just the descriptive norm was varied and the effect sizes were negligible. Comparing the condition where Professor Smith violates just the descriptive norm to the condition where he violates neither, $t(79.06)=-0.37, p=0.71$, Cohen’s $d=-0.081$, I found that the effect went in the reverse direction to that predicted by the intervention account, although it went in the predicted direction when comparing the condition where he violated both norms to the condition in which he just violated the injunctive norm, $t(77.28)=0.78, p=0.44$, Cohen’s $d=0.17$. 
Figure 1: Results for first four conditions of Study 1 with histograms above the plot of the means for each condition and showing 95% confidence intervals.

Turning to ratings for the fixed agent, I did not find a significant effect for violation of the injunctive norm, $F(1, 158)=1.52, p=0.22$, or violation of the descriptive norm, $F(1, 158)=0.014, p=0.91$, and I did not find an interaction effect, $F(1, 158)=0.26, p=0.61$. Follow-up
t-tests showed no significant superseding effects. Across the first four conditions, there are five comparisons that are relevant for assessing the causal superseding effect. Each test had a power of 0.89 or higher to detect an average superseding effect. Results for these comparisons are summarized in Table 2. First, I compared ratings for the administrative assistant in the condition where Professor Smith did not violate either norm (M=3.40, SD=1.93, N=40) to the condition in which he violated just the injunctive norm (M=3.17, SD=1.89, N=40). I found a drop of 0.22 points on the 7-point scale; the difference was not significant and the effect size was negligible, \( t(77.97)=-0.53, p=0.60, \) Cohen’s \( d=-0.12. \) Second, I compared ratings for the administrative assistant in the condition where Professor Smith did not violate either norm to the condition in which he violated just the descriptive norm (M=3.60, SD=1.93, N=42). I found a rise of 0.20 points; the difference was not significant and the effect size was negligible, \( t(79.78)=0.46, p=0.65, \) Cohen’s \( d=0.10. \) Third, I compared ratings for the administrative assistant in the condition where Professor Smith did not violate either norm to the condition in which he violated both norms (M=3.05, SD=2.23, N=40). I found a drop of 0.35 points; the difference was not significant and the effect size was negligible, \( t(76.45)=-0.75, p=0.46, \) Cohen’s \( d=-0.17. \) Fourth, I compared ratings for the administrative assistant in the condition where Professor Smith violates just the injunctive norm to the condition in which he violates both norms. I found a drop of 0.13 points; the difference was not significant and the effect size was negligible, \( t(76)=-0.27, p=0.79, \) Cohen’s \( d=-0.060. \) Finally, I compared ratings for the administrative assistant in the condition where Professor Smith violates just the descriptive norm to the condition in which he violates both norms. I found a drop of 0.55 points; the difference was not significant and the effect size was small, \( t(77.09)=-1.18, p=0.24, \) Cohen’s \( d=-0.26. \)
**Figure 2:** Results for last four conditions of Study 1 with histograms above the plot of the means for each condition and showing 95% confidence intervals.

Histograms and means for the causal ratings for the probes specifying an individual-level descriptive norm (last four conditions) are shown in Figure 2. I again ran two ANOVAs as specified above. Starting with ratings for the varied agent, as expected I again found a significant
effect for violation of the injunctive norm, $F(1, 161)=6.67, p=0.011$. In line with Sytsma et al. (2012), I also found a significant effect for violation of the individual-level statistical norm, $F(1, 161)=10.41, p=0.0015$, with the effect running in the opposite direction to that predicted by the causal sufficiency account, as we will see below. No interaction effect was found, $F(1, 161)=0.18, p=0.67$.

Follow-up t-tests showed a significant cross-agent effect with a medium effect size in the condition where Professor Smith violated just the injunctive norm, $t(41)=3.38, p=0.0016$, Cohen’s $d=0.57$. In the condition where Professor Smith violated both norms, however, the mean rating for the administrative assistant was higher than for Professor Smith, although this reverse effect was not significant and the effect size was small, $t(39)=-1.66, p=0.10$, Cohen’s $d=-0.43$. And in the condition where Professor Smith violated just the descriptive norm, the mean rating for the administrative assistant was again higher than for Professor Smith, and this reverse effect was significant with a large effect size, $t(41)=-5.11, p=7.9e-6$, Cohen’s $d=-1.09$.

Turning to the varied agent effect, follow-up t-tests showed one significant effect in the predicted direction in the individual-level typicality conditions and two reverse effects that were significant, with another that was borderline significant. Comparing the condition where Professor Smith violates just the injunctive norm to the condition where he violates neither, the difference in mean agreement with the causal claim for Professor Smith went in the predicted direction, although it was not significant and the effect size was small, $t(80.61)=1.42, p=0.16$, Cohen’s $d=0.31$. Comparing the condition where Professor Smith violates both norms to the condition where he violates neither, the difference in means ran in the reverse direction, although it was not significant and the effect size was negligible, $t(79.00)=-0.46, p=0.64$, Cohen’s $d=-0.10$. Comparing the condition where Professor Smith violates both norms to the condition where he violates just the descriptive norm, I found a significant varied agent effect although the effect
size was small, $t(75.75)=2.18$, $p=0.032$, Cohen’s $d=0.48$. Comparing the condition where Professor Smith violates just the descriptive norm to the condition where he violates neither I found a significant reverse effect with a medium effect size, $t(76.50)=-2.66$, $p=0.0094$, Cohen’s $d=-0.59$. Finally, comparing the condition where Professor Smith violated both norms to the condition in which he just violated just the injunctive norm I found a borderline significant reverse effect with a small effect size, $t(79.69)=-1.91$, $p=0.059$, Cohen’s $d=-0.42$.

Turning to ratings for the fixed agent, I did not find a significant effect for violation of the injunctive norm, $F(1, 161)=2.20$, $p=0.14$. In line with Sytsma et al. (2012), I again found a significant effect for violation of the individual-level statistical norm, $F(1, 161)=8.71$, $p=0.0036$, with the effect running in the opposite direction to that predicted by the causal sufficiency account. I did not find an interaction effect, $F(1, 161)=0.53$, $p=0.47$. Again, there are five relevant comparisons for assessing the causal superseding effect. I found only one significant effect and it ran in the reverse direction of the superseding effect. All tests had a power of 0.89 or higher to detect an average superseding effect. Results are summarized in Table 2. First, I compared ratings for the administrative assistant in the condition where Professor Smith did not violate either norm (M=3.98, SD=2.25, N=41) to the condition in which he violated just the injunctive norm (M=3.24, SD=2.05, N=42). I found a drop of 0.74 points on the 7-point scale; the difference was not significant and the effect size was small, $t(79.85)=-1.56$, $p=0.12$, Cohen’s $d=-0.34$. Second, I compared ratings for the administrative assistant in the condition where Professor Smith did not violate either norm to the condition in which he violated just the descriptive norm (M=4.74, SD=2.16, N=42). I found a rise of 0.76 points; the difference was not significant and the effect size was small, $t(80.67)=1.57$, $p=0.12$, Cohen’s $d=0.35$. Third, I compared ratings for the administrative assistant in the condition where Professor Smith did not violate either norm to the condition in which he violated both norms (M=4.50, SD=2.33, N=40).
I found a *rise* of 0.52 points; the difference was not significant and the effect size was small, $t(78.72)=1.03, p=0.31$, Cohen’s $d=-0.23$. Fourth, I compared ratings for the administrative assistant in the condition where Professor Smith violates just the injunctive norm to the condition in which he violates both norms. I found a *rise* of 1.26 points; the difference was significant and the effect size was medium, $t(77.52)=2.6, p=0.011$, Cohen’s $d=0.58$. Finally, I compared ratings for the administrative assistant in the condition where Professor Smith violates just the descriptive norm to the condition in which he violates both norms. I found a drop of 0.24 points; the difference was not significant and the effect size was negligible, $t(78.80)=-0.48, p=0.63$, Cohen’s $d=-0.011$.

Summing up, in my first study I found no evidence for a general superseding effect in the Pen Case. Across 10 comparisons, I found only one significant effect, and it ran in the reverse direction of the superseding effect. In fact, in four out of the 10 comparisons, the mean agreement rating for the administrative assistant was higher when the varied agent violated one or more norms than when the varied agent violated fewer norms. This included all three comparisons where the varied norms included an individual-level statistical norm, providing further evidence against the third prediction of the causal sufficiency account.

### 3.2 Study 2: Lauren and Jane Case

As with the Pen Case, our goal in testing the Lauren and Jane Case in Livengood et al. (2017) was not to test the causal superseding effect, and our studies were not designed with that in mind. The same potential issue noted above applies to these studies: in the cases where an injunctive norm was not violated, we did not specify that either agent was allowed to log into the mainframe. This was rectified in the present study by adding a line to the vignettes: “Employees are allowed to log into any of these terminals.”
3.2.1 Methods

Results were collected from 262 participants. Participants were given one of four vignettes featuring a varied agent (Jane) and a fixed agent (Lauren) logging into a mainframe, as described in 2.2. In the first two, the mainframe was described as being unstable, such that it will crash if more than one person is logged in. In the last two, the mainframe was instead described as being designed for a single user. In the first vignette of each pair, both agents were allowed to log in, while in the second vignette Jane violated an injunctive norm by logging in. In each condition, participants were asked to rate agreement/disagreement with each of two causal statements—“Lauren caused the system to crash” and “Jane caused the system to crash”—on the same 7-point scale used previously.

3.2.2 Results

Results are shown in Figure 3; see Table 2 for a summary for the superseding effect. For each pair of conditions, I conducted planned comparisons to assess each of the three effects identified in Section 1. For the first pair, in the condition where Jane violated the injunctive norm there was a significant cross-agent effect with a large effect size, $t(62)=6.64, p=8.8e-9, \text{Cohen’s } d=1.18$. Across the two conditions there was a significant varied agent effect with a large effect size,

---

12 59.2% women (six non-binary), average age 26.5 years, ranging from 16 to 99. I ran four ANOVAs with whether Professor Smith violated the injunctive norm and gender as between-participant factors. I found a significant main effect for gender for ratings of Professor Smith for the first set of probes, $F(1, 118)=7.54, p=0.0070$, and a significant interaction effect for ratings of the administrative assistant, $F(1, 118)=4.07, p=0.046$. The same effects were borderline significant for the second set of probes: $F(1, 136)=2.93, p=0.089; F(1, 136)=3.12, p=0.079$. Looking closer, while there are some differences in the strength of the cross-agent and varied agent effects between the groups on each set of probes (women showing a slightly weaker cross-agent effect and a slightly stronger varied agent effect), the trends were the same for both. There were more notable differences with regard to the superseding effect. Details are given in footnotes below.

13 Among women the mean rating for Jane ($M=4.57, SD=1.92, N=37$) was 1.76 points higher than for Lauren ($M=2.81, SD=1.61$). Among non-women the mean rating for Jane ($M=5.42, SD=1.92, N=26$) was 2.88 points higher than for Lauren ($M=2.54, SD=2.10$).
Finally, looking at the superseding effect, across the two conditions the mean rating for Lauren when Jane violated the norm (M=2.70, SD=1.82, N=63) was 0.28 points lower than when she did not violate the norm (M=2.98, SD=2.06, N=59). The difference was not significant and the effect size was negligible, \( t(115.79)=-0.81, p=0.42, \) Cohen’s \( d=-0.15. \) This test had a power of 0.98 to detect an average superseding effect.

For the second pair of probes, in the condition where Jane violated the injunctive norm there was a significant cross-agent effect with a large effect size, \( t(77)=10.66, p<2.2e-16, \) Cohen’s \( d=2.07. \) Across the two conditions there was a significant varied agent effect with a large effect size, \( t(134.2)=7.69, p=2.7e-12, \) Cohen’s \( d=1.30. \) Finally, looking at the superseding effect, across the two conditions the mean rating for Lauren when Jane violated the norm (M=1.90, SD=1.53, N=78) was 0.99 points lower than when she did not violate the norm (M=2.89, SD=1.73, N=62). The difference was significant and the effect size was medium, \( t(122.67)=-3.54, p=0.00056, \) Cohen’s \( d=-0.61. \) Summing up, I found some further evidence for a general superseding effect in the Lauren and Jane Case. In line with the results of Livengood et

---

14 Among women the mean rating for Jane when she violated the norm (M=4.57, SD=1.92, N=37) was 2.02 points higher than when she did not (M=2.55, SD=1.89, N=40). Among non-women the mean rating for Jane when she violated the norm (M=5.42, SD=1.92, N=26) was 1.69 points higher than when she did not (M=3.74, SD=2.12, N=19).

15 Among women the mean rating for Lauren when Jane violated the norm (M=2.81, SD=1.61, N=37) was 0.21 points higher than when she did not (M=2.60, SD=1.92, N=40); not significant with negligible effect size, \( t(74.35)=0.52, p=0.60, \) Cohen’s \( d=0.12. \) Among non-women the mean rating for Lauren when Jane violated the norm (M=2.54, SD=2.10, N=26) was 1.25 points lower than when she did not (M=3.79, SD=2.18, N=19); borderline significant with medium effect size, \( t(38.16)=-1.93, p=0.061, \) Cohen’s \( d=-0.59. \)

16 Among women the mean rating for Jane (M=5.39, SD=2.23, N=44) was 3.30 points higher than for Lauren (M=2.09, SD=1.76). Among non-women the mean rating for Jane (M=5.85, SD=1.67, N=34) was 4.21 points higher than for Lauren (M=1.65, SD=1.12).

17 Among women the mean rating for Jane when she violated the norm (M=5.49, SD=2.23, N=44) was 2.65 points higher than when she did not (M=2.74, N=1.76, N=34). Among non-women the mean rating for Jane when she violated the norm (M=5.85, SD=1.67, N=34) was 2.42 points higher than when she did not (M=3.43, SD=1.99, N=28).

18 Among women the mean rating for Lauren when Jane violated the norm (M=2.09, SD=1.76, N=44) was 0.56 points lower than when she did not (M=2.74, SD=1.76, N=34). The effect was not significant and the effect size was small, \( t(74.36)=-1.47, p=0.15, \) Cohen’s \( d=-0.33. \) Among non-women the mean rating for Lauren when Jane violated the norm (M=1.65, SD=1.12, N=34) was 1.53 points lower than when she did not (M=3.43, SD=1.99, N=28). The effect was significant and the effect size was large, \( t(42.17)=-3.78, p=0.00049, \) Cohen’s \( d=-1.010. \)
al. (2017), I found only a negligible effect for the probes specifying that the mainframe is “unstable,” but saw a notable effect when this is instead described as a feature of the system.

**Figure 3:** Results of Study 2 with histograms above the plot of the means for each condition and showing 95% confidence intervals.
3.3 Study 3: Dairy Corporation Case

I now turn to a new conjunctive case. The Dairy Corporation Case has a similar set-up to the cases discussed above, with two actions being jointly necessary to produce an outcome. In this case, however, the varied agent (Dairy Corporation) is a corporation rather than an individual, and I varied whether the corporation violated an explicit law or not.

3.2.1 Methods

Results were collected from 147 participants. Participants were given one of two vignettes featuring the varied agent and the fixed agent. In each case, Dairy Corporation sold milk that is contaminated with a fictional chemical, mercurous hypochlorite, which the vignette says is not harmful on its own but which can lead to serious health consequences if combined with a fictional sugar substitute, aspartalose. Geraldine then uses the contaminated milk and aspartalose to bake a cake. Amy eats a piece of the cake, convulses, and dies. The first vignette specifies that current law allows for the presence of mercurous hypochlorite in milk and does not require that it be reported. In contrast, the second vignette specifies that current law prohibits the presence of mercurous hypochlorite in milk. In each condition, participants were asked to rate how much they agreed or disagreed with each of two causal statements—“Dairy Corporation caused Amy’s death” and “Geraldine caused Amy’s death”—on the same 7-point scale used previously.

3.3.2 Results

Results are shown in Figure 4; see Table 2 for a summary for the superseding effect. I conducted planned comparisons to assess each of the three effects identified in Section 1. In the condition

---

19 61.9% women (one non-binary), average age 31.5 years, ranging from 16 to 81. I ran two ANOVAs, with whether Dairy Corporation broke the law and gender as between-participant factors. No significant gender effects were found.
where Dairy Corporation breaks the law there was a significant cross-agent effect with a large effect size, $t(67)=11.54, p<2.2\cdot10^{-16}$, Cohen’s $d=2.25$. Across the two conditions I did not find a significant varied agent effect, $t(118.67)=5.50, p=2.2\cdot10^{-7}$, Cohen’s $d=1.00$. Given the high mean rating for Dairy Corporation when it did not break the law, I also checked for a cross-agent effect in this condition. I found a significant effect with a large effect size, $t(78)=9.28, p=3.1\cdot10^{-14}$.

Finally, looking at the superseding effect, across the two conditions the mean rating for Geraldine when Dairy Corporation broke the law ($M=2.07, SD=1.50, N=68$) was 0.32 points lower than when Dairy Corporation did not break the law ($M=2.39, SD=1.76, N=79$). The difference was not significant and the effect size was negligible, $t(144.98)=-0.32, p=0.24$, Cohen’s $d=-0.19$. This test had a power of 0.99 to detect an average superseding effect.

**Figure 4:** Results of Study 3 with histograms to the right of the plot of the means for each condition and showing 95% confidence intervals.
Summing up, in my third study I did not find evidence of a superseding effect for the Dairy Corporation Case. Interestingly, while there was a cross-agent effect, there was not a varied agent effect. This reflects that agreement ratings for Dairy Corporation were quite high in both conditions. This might indicate that people are inclined to think that Dairy Corporation acted badly in either case, perhaps violating an implicit injunctive norm against selling a product contaminated with a potentially dangerous chemical, whether or not it is against the law to do so. This is in keeping with the responsibility account. Kominsky et al.’s counterfactual sufficiency account, however, is premised on violations of norms making certain counterfactuals more salient, and the explicit legal norm in this case would seem to do exactly that, so they should plausibly still predict that there will be a superseding effect in this case (although they might argue that the implicit norm swamps out the explicit norm).

3.4 Study 4: Trash Case

In my fourth study, I consider another case not been previously tested in the literature, but was presented as a thought experiment in Knobe (2006). The Trash Case was given as a thought experiment in Knobe (2006). Like the previous cases it is conjunctive, but unlike the previous cases this time the agents bring about the outcomes through omissions.

---

20 To check this, I ran a follow up study replacing “caused” with “is responsible for” in each of the causal claims. Results were collected from 165 participants (63.6% women (one non-binary), average age 27.7 years, ranging from 16 to 65). While the responsibility account neither holds that the language of causal attribution and the language of responsibility attributions work in exactly the same way nor holds that people are univocal in using the language of causal attribution in the way we lay out, our account does predict that causal attributions and responsibility attributions should show similar trends in cases like these. And that is what I found. Combining the two studies, ANOVAs with whether Dairy Corporation broke the law and term (cause, responsible) as between-participants factors did not show a significant main effect for term for either agreement ratings for Dairy Corporation, $F(1, 308)=1.65, p=0.20$, or Geraldine, $F(1, 308)=1.18, p=0.28$. Nor did I find a significant interaction effect for either attribution, $F(1, 308)=1.10, p=0.30$, $F(1, 308)=0.78, p=0.38$. But, as expected, I did find a significant main effect for whether Dairy Corporation broke the law for each attribution, $F(1, 308)=6.23, p=0.013$, $F(1, 308)=7.12, p=0.0080$. 
3.4.1 Methods
Results were collected from 164 participants.\textsuperscript{21} Participants were given one of four vignettes featuring a varied agent (George) and a fixed agent (Harry) who do not take out the trash in an office building, resulting in either a bad outcome (an odor fills the building) in the first pair of vignettes or a good outcome (an important document is saved) in the second pair of vignettes. In the first vignette in each pair, George is described as a janitor who is supposed to take out the garbage but neglects to do so on this day. In the second vignette in each pair, George is instead described as taking care of the computers, and it is stated that it is not his job to take out the garbage. In each of the four vignettes, Harry is described as taking care of the mail, and it is stated that it is not his job to take out the garbage. In the first pair of conditions, participants were asked to rate how much they agreed or disagreed with each of two causal statements—“George caused the bad odor to fill the building” and “Harry caused the bad odor to fill the building”—on the same 7-point scale used in the previous studies. In the second pair of conditions, the two causal statements were changed to “George caused the document to be recovered” and “Harry caused the document to be recovered.”

3.4.2 Results
Results are shown in Figure 5; see Table 2 for a summary for the superseding effect. I ran two ANOVAs, one looking at agreement ratings for the fixed agent, one for the varied agent, with whether the varied agent violated the injunctive norm and outcome valence as between-participants factors. Starting with ratings for the varied agent, as expected I found a significant main effect for violation of the injunctive norm, $F(1, 160)=17.48$, $p=4.8e-5$. Against what we

\textsuperscript{21} 72.0% women (one non-binary), average age 31.3 years, ranging from 16 to 87. I ran the four ANOVAs described below with gender added as a between-participant factor. No significant gender effects were found.
would predict based on the intervention account, I also found a significant main effect for outcome valence, $F(1, 160)=15.42, p=0.00013$. I did not find an interaction effect, $F(1, 160)=2.45, p=0.12$. Follow-up t-tests showed a significant cross-agent effect with a medium effect size, $t(39)=3.19, p=0.0028$, Cohen’s $d=0.60$. I also found a borderline significant varied agent effect with a small effect size, $t(79.67)=1.95, p=0.055$, Cohen’s $d=0.43$.

Turning to the ratings for the fixed agent, I did not find a significant main effect for violation of the injunctive norm, $F(1, 160)=0.49, p=0.48$, although I found a significant interaction effect, $F(1, 160)=4.68, p=0.032$. And against what we would predict based on the intervention account, I again found a significant main effect for outcome valence, $F(1, 160)=11.75, p=0.00077$. A follow-up t-test failed to find a significant superseding effect in the conditions where the outcome was bad, $t(79.67)=-1.03, p=0.31$, Cohen’s $d=-0.23$. This test had a power of 0.90 to detect an average superseding effect. The agreement ratings for Harry were lower when George violated the injunctive norm ($M=3.75, SD=1.96, N=40$) than when George did not violate the injunctive norm ($M=4.19, SD=1.93, N=42$), a drop of 0.44 points, which corresponds with a small effect size. This pattern was reversed when the outcome was good, however, with agreement ratings for Harry being higher when George violated the injunctive norm ($M=3.37, SD=2.08, N=41$) than when George did not violate the injunctive norm ($M=2.46, SD=1.98, N=41$), a rise of 0.90 points. This difference was significant although the effect size was small, $t(79.78)=2.01, p=0.048$, Cohen’s $d=0.44$.

Summing up, in my fourth study I failed to find evidence of a notable superseding effect for the George and Harry Case. Further, against the predictions of the intervention account, I found that outcome valence showed a significant effect on ratings for both the varied agent and the fixed agent. And against the third prediction of the causal sufficiency account—that “causal superseding should occur even for outcomes that are in no way bad” (199)—I not only failed to
find a superseding effect when the outcome was good, but found a significant effect running in the reverse direction.

**Figure 5:** Results of Study 4 with histograms above the plot of the means for each condition and showing 95% confidence intervals.
3.5 Study 5: Motion Detector Case

As described in Section 2, in Kominsky et al.’s third experiment they gave participants one of a pair of scenarios in which two employees show up at work at the same time and set off a motion detector. They then asked whether the fixed agent caused the motion detector to go off, finding a significant superseding effect. In my fifth study, I repeated the conjunctive version of this scenario asking each participant about both the fixed agent and the varied agent.

3.5.1 Methods

Results were collected from 82 participants. Participants were given one or the other of the two conjunctive vignettes from Kominsky et al.’s third experiment. In the first vignette the varied agent (Billy) does not violate an injunctive norm, in the second vignette he does. In each vignette the fixed agent (Suzy) does not violate an injunctive norm. In each condition, participants were asked to rate how much they agreed or disagreed with each of two causal statements—“Billy caused the motion detector to go off” and “Suzy caused the motion detector to go off”—on the same 7-point scale used in our previous studies. Following Kominsky et al., participants were then given a comprehension check. They were asked “Who was supposed to show up at 9am?” and were able to select either “Billy,” “Suzy,” or “Both of them.” Two participants failed the comprehension check and were excluded from the subsequent analysis.

3.5.2 Results

Results are shown in Figure 6; see Table 2 for a summary for the superseding effect. I conducted planned comparisons to assess each of the three effects identified in Section 1. There was a

---

22 57.3% women, average age 30.1 years, ranging from 16 to 79. I ran two ANOVAs with whether Billy violated the injunctive norm and gender as between-participant factors. No significant gender effects were found.
significant cross-agent effect with a large effect size, $t(39)=7.47, p=5.0e-9$, Cohen’s $d=2.02$. There was also a significant varied agent effect with a large effect size, $t(77.93)=5.15, p=1.9e-6$, Cohen’s $d=1.15$. Most importantly, in line with Kominsky et al.’s results, I found a significant superseding effect with a large effect size, although this was somewhat smaller than in the original experiment, $t(74.90)=-3.28, p=0.0016$, Cohen’s $d=-0.733$. The mean rating for Suzy when Billy violated the injunctive norm (M=2.00, SD=1.55, N=40) was 1.28 points lower than when Billy did not violate the norm (M=3.27, SD=1.91, N=40).

![Figure 6: Results for Study 5 with histograms to the right of the plot of the means for each condition and showing 95% confidence intervals.](image)
<table>
<thead>
<tr>
<th>Study, Sub-set</th>
<th>Varied Norm Violation</th>
<th>Constant Norm Violation</th>
<th>Superseding Effect</th>
<th>p-value</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, First Four</td>
<td>Injunctive</td>
<td>-0.23, Negligible</td>
<td>0.60</td>
<td>-0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Population-level</td>
<td>R: 0.20, Negligible</td>
<td>0.65</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Injunctive, Population-level</td>
<td>-0.35, Negligible</td>
<td>0.46</td>
<td>-0.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Population-level</td>
<td>Injunctive</td>
<td>-0.13, Negligible</td>
<td>0.79</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>Injunctive, Population-level</td>
<td>-0.55, Small</td>
<td>0.24</td>
<td>-0.26</td>
<td></td>
</tr>
<tr>
<td>1, Second Four</td>
<td>Injunctive</td>
<td>-0.74, Small</td>
<td>0.12</td>
<td>-0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Individual-level</td>
<td>R: 0.76, Small</td>
<td>0.12</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Injunctive, Individual-level</td>
<td>R: 0.52, Small</td>
<td>0.31</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Individual-level</td>
<td>Injunctive</td>
<td>R: 1.26, Small*</td>
<td>0.011</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Injunctive, Individual-level</td>
<td>-0.24, Negligible</td>
<td>0.63</td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td>2, First Set</td>
<td>Injunctive</td>
<td>-0.28, Negligible</td>
<td>0.42</td>
<td>-0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>R: 0.21, Negligible</td>
<td>0.60</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-women</td>
<td>-1.25, Medium</td>
<td>0.061</td>
<td>-0.59</td>
<td></td>
</tr>
<tr>
<td>2, Second Set</td>
<td>Injunctive</td>
<td>-0.99, Medium***</td>
<td>0.00056</td>
<td>-0.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>-0.56, Small</td>
<td>0.15</td>
<td>-0.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-women</td>
<td>-1.53, Large***</td>
<td>0.00049</td>
<td>-1.01</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Injunctive</td>
<td>-0.32, Negligible</td>
<td>0.34</td>
<td>-0.19</td>
<td></td>
</tr>
<tr>
<td>4, Bad Outcome</td>
<td>Injunctive</td>
<td>-0.44, Small</td>
<td>0.31</td>
<td>-0.23</td>
<td></td>
</tr>
<tr>
<td>4, Good Outcome</td>
<td>Injunctive</td>
<td>R: 0.90, Small*</td>
<td>0.048</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Injunctive</td>
<td>-1.28, Medium**</td>
<td>0.0016</td>
<td>-0.73</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**: Summary of comparisons looking for causal superseding effects, Studies 1-5; **R** denotes that the difference in means was the reverse of that predicted.

**4. Alternative Explanation: Context Effect**

Across the cases just surveyed, we have seen a great deal of variation both in whether the causal superseding effect occurs and the strength of the effect when it does occur. Some of these results raise doubts about whether the effect occur consistently for certain types of descriptive norms (an issue for the third prediction of Kominsky et al.’s counterfactual sufficiency account), while others raise doubts about whether the effect occur consistently when bad behaviors lead to a good outcome (an issue for the first prediction of Kominsky et al.’s counterfactual sufficiency account). While the superseding effect neither looks to be as regular nor as strong as Kominsky et al.’s experiments would suggest, there remains reason to think that an effect does occur, at
least in some cases. The full range of cases considered, however, raises doubts about whether the counterfactual sufficiency account can adequately explain the effect. In its place I offer an alternative, deflationary explanation. I suggest that the causal superseding is a byproduct of the cross-agent effect. My hypothesis is that the superseding effect is a type of context effect that arises from relative differences in the perceived causal strength of the agents affecting how participants interpret the scale used in eliciting their causal ratings.

The occurrence of context effects in survey research is well recognized (Sudman, Bradburn, and Schwarz 1996, Chapter 4). And this includes ratings of stimuli. Schwarz (1996, 77) summarizes:

As numerous studies demonstrate, respondents use the most extreme stimuli to anchor the endpoints of a rating scale. As a result, a given stimulus will be rated as less extreme if presented in the context of a more extreme one, than if presented in the context of a less extreme one. In Parducci’s [1983] model, this impact of the range of stimuli is referred to as the ‘range effect’.

While Parducci (1983) was dealing with psychophysical experiments in which participants were interacting with perceptible objects, similar effects are found in questionnaire studies. For instance, Daamen and de Bie (1992) show that participant ratings for the probabilities of target events occurring were higher if they were preceded by a set of events with a lower probability of occurring than if they were preceded by a set of events with a higher probability of occurring. Or consider an experiment conducted by Schwarz, Münkel, and Hippler (1990). Participants were asked to rate how “typically German” four drinks were on a 9-point scale anchored at 1 with “not at all typical” and at 9 with “very typical.” The first drink in the list was varied between a prototypically German drink (beer) and an atypical drink (vodka). The remaining three drinks were fixed (wine, coffee, and milk). Schwarz and colleagues found that the mean typicality rating for the fixed drinks was significantly lower when preceded by beer (M=4.42) than when preceded by vodka (M=5.40). This is essentially a “superseding effect” for typicality judgments:
the extent to which the target drinks (fixed drinks) are judged to be typically German is affected by the typicality status of the other drink (varied drink).

Such experiments lead Sudman et al. (1996, 96) to conclude that “a given stimulus will receive different ratings depending on the extremity, frequency, and distribution of context stimuli introduced in the questionnaire.” While the occurrence of context effects is generally recognized, it is also recognized that they do not always occur and are often hard to predict. As Schwarz and Sudman (1992, 3-4) put it, “context effects have sometimes been observed and other times not observed, and predictions about when they would and would not occur have often been wrong.” That context effects can be fickle in this way fits with the range of results we have seen for causal superseding across the previous sections.

Consider how a context effect might arise in the studies we have looked at. Accepting that injunctive norm violations increase judgments of causal strength, in these studies participants are asked to assess a causal statement about the fixed agent either in the context of a comparison agent whose causal strength is relatively moderate (varied agent does not violate the norm) or relatively extreme (varied agent violates the norm). Based on the occurrence of context effects alone, we would expect the ratings for the fixed agent to be lower when contrasted with a varied agent who violates a norm than when contrasted with a varied agent who does not violate a norm. But that is exactly the effect at issue for superseding. As such, I contend that it is prima facie plausible that the superseding effect is nothing more than a context effect. To help make the case, I ran three further studies. The first two suggest that the superseding effect could be a context effect. The third study suggests that the superseding effect is in fact a context effect.
4.1 Study 6: Height Judgments, Two Claims

4.1.1 Methods

Results were collected from 133 participants. Participants were simply given one of two sets of two claims. The first claim was either “Danny DeVito is tall” or “Shaquille O’Neil is tall.” This is analogous to the varied agent. The second claim was the same for both sets—“Tom Cruise is tall.” This is analogous to the fixed agent. In each condition participants were asked to rate their agreement with the claims using the same 7-point scale as the previous studies.

4.1.2 Results

Results are shown in Figure 7. Comparing ratings for Tom Cruise across the two conditions, I found a significant difference with a medium effect size, \( t(130.53) = -3.43, p = 0.00081 \), Cohen’s \( d = -0.59 \). The mean rating was lower when the other claim was about Shaquille O’Neil (\( M = 2.58, SD = 1.46, N = 64 \)) than when it was about Danny DeVito (\( M = 3.51, SD = 1.67, N = 69 \)), a drop of 0.93 points. This is analogous to the superseding effect, with variation in the height of one member of the pair affecting height judgments about the other. I also found a significant “cross-agent effect” with a large effect size, \( t(63) = 13.40, p < 2.2e-16 \), Cohen’s \( d = 2.78 \), and a significant “varied agent effect” with a large effect size, \( t(127.39) = 16.60, p < 2.2e-16 \), Cohen’s \( d = 2.85 \).

23 64.7% women (three non-binary), average age 35.1 years, ranging from 16 to 86. I ran an ANOVA with whether the first claim was about DeVito or O’Neil and gender as between-participant factors. While there was not a significant main effect for gender, \( F(1, 129) = 0.0015, p = 0.97 \), there was a significant interaction effect, \( F(1, 129) = 8.091, p = 0.0052 \). Looking closer, I found a notable difference between the two groups with regard to the “superseding effect.” Details noted below.

24 Among women the mean rating for Cruise when O’Neil was the comparison (\( M = 3.23, SD = 1.74, N = 43 \)) was 0.37 points lower than when DeVito was the comparison (\( M = 2.86, SD = 1.46, N = 43 \)). The effect was not significant and the effect size was small, \( r(81.43) = -1.074, p = 0.29 \), Cohen’s \( d = -0.23 \). Among non-women the mean rating for Cruise when O’Neil was the comparison (\( M = 3.96, SD = 1.46, N = 26 \)) was 1.96 points lower than when DeVito was the comparison (\( M = 2.00, SD = 1.30, N = 21 \)). The effect was significant and the effect size was large, \( r(44.48) = -4.87, p = 1.5e-5 \), Cohen’s \( d = -1.41 \).
4.2 Study 7: Height Judgments, One Claim

Kominsky et al. find a superseding effect not just when they ask about both agents, but also when they ask about just the fixed agent. As such, I ran a variation on the previous study in which I just asked about the “fixed agent.”

4.2.1 Methods

Results were collected from 101 participants. Participants were asked to consider one of two sets of celebrities before rating their agreement with a claim about one of them (“Tom Cruise is tall” in each case) using the same 7-point scale used in my previous studies. In the first condition, participants were told to “Consider the following two celebrities: former NBA player Shaquille O’Neil and actor Tom Cruise.” In the second condition, “former NBA player Shaquille O’Neil” was replaced with “actor Danny DeVito.”

4.2.2 Results

Results are shown in Figure 7. Comparing across the two conditions I found that mean ratings for Cruise were lower when participants were asked to consider O’Neil (M=2.50, SD=1.30, N=50) than when they were asked to consider DeVito (M=3.78, SD=1.65, N=51), a drop of 1.28 points. This difference was significant and the effect size was large, t(94.51)=−4.35, p=3.5e-5, Cohen’s d=−0.86. Again, this is analogous to the causal superseding effect, with variation in the height of one member of the pair affecting height judgments about the other.

25 69.3% women (one non-binary), average age 30.4 years, ranging from 16 to 69. I ran an ANOVA with whether the contrasting agent and gender as between-participant factors. No significant gender effects were seen.
Figure 7: Results for Studies 6 and 7 with histograms above the plot of the means for each condition and showing 95% confidence intervals.

4.3 Study 8: Varying the Scale

The results for relative agreement ratings about the height of Tom Cruise are most naturally explained in terms of a context effect, with the context provided by the “varied agent” affecting
responses with regard to the “fixed agent.” That a context effect can be generated using the same scale as in my previous studies suggests that the superseding effect might just be a context effect. But while this is suggestive, it is not direct evidence. In my eighth study, I attempt to test the context effect hypothesis more directly. I reason that if the superseding effect is a context effect, then we would expect the size of the effect to vary with the size of the scale used. The idea is that a larger scale would provide more room for the context effect to occur—more room for the perceived causal strength of the varied agent to push around ratings for the varied agent.

4.3.1 Methods

Results were collected from 249 participants.26 As seen in Table 2, I found the strongest superseding effect in Study 5 using the conjunctive version of Kominsky et al.’s Motion Detector Case with a bad outcome. I expanded on these results by testing the same pair of conditions, but this time varying the size of the scale. I tested three further pairs of conditions—the first using an 11-point scale, the second a 9-point scale, and the third a 5-point scale. As in the original study using a 7-point scale, the low extreme was anchored with “strongly disagree,” the mid-point was anchored with “neutral,” and the high extreme was anchored with “strongly agree.” The comprehension check used by Kominsky et al. was also included. Eight participants failed the comprehension check and were excluded from the subsequent analysis.

4.3.2 Results

The results from the bad outcome conditions from Study 5 were combined with the results from Study 8. To facilitate analysis of the effect of the size of the scale on the superseding effect, I

26 63.5% women, average age 28.4 years, ranging from 16 to 73. I ran an ANOVA with the size of the scale, whether the varied agent violated a norm, and gender as between-participant factors. No significant gender effects were seen.
converted the ratings to POMP scores (percentage of maximum possible score). I ran an ANOVA looking at POMP scores for the fixed agent, with whether the varied agent violated an injunctive norm and the size of the scale as between participant factors. As expected, I found a significant main effect for whether the varied agent violated an injunctive norm, $F(1, 317)=66.14$, $p=9.6e-15$, indicating that there was generally a superseding effect. I did not find a significant main effect for the size of the scale, $F(1, 317)=2.40$, $p=0.122$, but in line with the context effect hypothesis I did find a significant interaction effect, $F(1, 317)=4.97$, $p=0.027$.

Follow-up t-tests show a significant superseding effect for each size of scale, but with the size of the effect varying with the size of the scale. For the 11-point scale, the average POMP score for the fixed agent was 33.5 percentage points higher when the varied agent did not violate the norm (M=50.3, SD=27.9, N=40) than when he did (M=16.8, N=23.8, N=40). This difference was statistically significant and the effect size was large, $t(76.13)=-5.78$, $p=1.5e-7$, Cohen’s $d=-1.29$. For the 9-point scale, the average POMP score was 35.4 percentage points higher when the varied agent did not violate the norm (M=55.1, SD=31.6, N=42) than when he did (M=19.7, SD=28.1, N=40). This difference was statistically significant and the effect size was large, $t(79.65)=-5.36$, $p=8.0e-7$, Cohen’s $d=-1.18$. For the 7-point scale, the average POMP score was 21.3 percentage points higher when the varied agent did not violate the norm (M=37.9, SD=31.8, N=40) than when he did (M=16.7, SD=25.9, N=40). This difference was statistically significant and the effect size was medium, $t(74.90)=-3.28$, $p=0.0016$, Cohen’s $d=-0.73$. And for the 5-point scale, the average POMP score was 16.3 percentage points higher when the varied agent did not violate the norm (M=37.5, SD=33.5, N=40) than when he did (M=21.2, SD=31.7, N=39). This difference was statistically significant and the effect size was medium, $t(76.93)=-2.23$, $p=0.029$, Cohen’s $d=-0.50$. Going from the 5-point scale to the 11-point scale, then, we find that both the difference in average POMP score and the effect size of the superseding effect more than double.
One might worry that the effect of the size of the scale on the superseding effect is a more
general effect, with the cross-agent effect of norm violations also varying with the scale. To test
this, I ran a second ANOVA looking at the difference in POMP scores for agreement ratings for
the varied agent and agreement ratings for the fixed agent. As expected, I found a significant
main effect for norm violation, $F(1, 317)=192.02$, $p<2.2e-16$, but did not find a significant main
effect for the size of the scale, $F(1, 317)=0.022$, $p=0.88$, or a significant interaction effect, $F(1,$
$317)=0.071$, $p=0.79$. Follow-up t-tests show a significant cross-agent effect with a comparable
effect size for each size of scale in the conditions where the varied agent violated the norm. For
the 11-point scale, the average POMP score for the varied agent was 58.0 percentage points
higher than for the fixed agent. This difference was statistically significant and the effect size
was large, $t(71.39)=9.09$, $p=1.5e-13$, Cohen’s $d=2.03$. For the 9-point scale, the average POMP
score for the varied agent was 62.5 percentage points higher than for the fixed agent. This
difference was statistically significant and the effect size was large, $t(77.88)=9.73$, $p=4.2e-15$,
Cohen’s $d=2.18$. For the 7-point scale, the average POMP score for the varied agent was 59.6
percentage points higher than for the fixed agent. This difference was statistically significant and
the effect size was large, $t(74.12)=9.04$, $p=1.4e-15$, Cohen’s $d=2.02$. For the 5-point scale, the
average POMP score for the varied agent was 58.3 percentage points higher than for the fixed
agent. This difference was statistically significant and the effect size was large, $t(75.80)=7.91$,
$p=1.6e-11$, Cohen’s $d=1.79$.

5. Disjunctive Cases
Summing up, we have seen that while the causal superseding effect occurs with some regularity,
it is quite variable across cases. We’ve seen mixed results concerning whether it occurs for both
good and bad outcomes, which puts stress on the first prediction of Kominsky et al.’s causal
sufficiency account, with the results indicating that at the very least it does not occur with equal strength for both good and bad outcomes, which contradicts Hitchcock and Knobe’s prediction in laying out their intervention account. We’ve also seen evidence that the effect does not occur for all types of norm violation, which contradicts Kominsky et al.’s third prediction. And, finally, we saw evidence in Studies 6 and 7 that the superseding effect could be a context effect, and we saw more direct evidence that it is a context effect in Study 8.

But what about Kominsky et al.’s second prediction—that the superseding effect should not occur when sufficiency is not threatened, such as in disjunctive cases? This prediction is rather specific to the causal sufficiency account, and in their third experiment Kominsky et al. did not find a superseding effect in the disjunctive version of the Motion Detector Case. As noted in Section 1, however, the failure of the superseding effect in some disjunctive cases is also compatible with the responsibility account. First, if I am correct that the superseding effect is a context effect, then we shouldn’t be surprised that it does not occur in some cases. More generally, given that we have seen a good bit of variation in whether the effect occurs in conjunctive cases, we should be wary of generalizing on a single failure. Second, I expect the superseding effect to be less pronounced for good outcomes, and we saw above that in the disjunctive versions of the Motion Detector Case we can plausibly infer that the outcome is good. Third, on my account we should not expect to see a superseding effect when there is not a cross-agent effect. Since Kominsky et al. only asked participants about the fixed agent in their third experiment, we cannot tell whether there is a cross-agent effect. As such, I conducted three final studies to further test the superseding effect for disjunctive cases. I began by repeating Kominsky et al.’s third experiment, but this time asking about both the fixed agent and the varied agent. I then tested two different disjunctive cases where the outcome is more clearly bad.
5.1 Study 9: Disjunctive Motion Detector Case

5.1.1 Methods

Results were collected from 86 participants. Participants were given one of the two disjunctive versions of the Motion Detector Case from Kominsky et al., following the same procedure used in my fifth study. The only difference was that in this study the motion detector was described as being triggered if at least one person appeared in the room. Five participants failed the comprehension check and were excluded from the subsequent analysis.

5.1.1.3 Results

Results are shown in Figure 8. I conducted planned comparisons to assess each of the three effects identified in Section 1. I did not find a significant cross-agent effect, \( t(39)=1.39, p=0.17 \), Cohen’s \( d=0.17 \). Nor did I find a significant varied agent effect, \( t(72.51)=0.94, p=0.35 \), Cohen’s \( d=0.21 \). And in line with Kominsky et al.’s results, I did not find a significant superseding effect, \( t(72.69)=-0.84, p=0.40 \), Cohen’s \( d=-0.19 \). The mean rating for Suzy when Billy violated the injunctive norm (M=3.60, SD=2.42, N=40) was 0.40 points lower than when Billy did not violate the norm (M=4.00, SD=1.83, N=41). This is a somewhat larger difference than found by Kominsky et al., however, who saw a drop of just 0.09 points. This test had a power of 0.89 to detect an average superseding effect.

\(^{27}\) 64.0% women (two non-binary), average age 31.1 years, ranging from 16 to 77. I ran two ANOVAs, with whether the varied agent violated the norm and gender as between-participant factors. No significant gender effects were found.
Figure 8: Results for Study 9 with histograms to the right of the plot of the means for each condition and showing 95% confidence intervals.

5.2 Study 10: Lauren, Jane, and Helen

Kominsky et al.’s account predicts that we will not see a superseding effect in disjunctive cases, and I did not find a significant effect in the disjunctive version of the Motion Detector Case. In line with the responsibility account I tentatively predict that superseding effects will sometimes occur in disjunctive cases but tie the occurrence of a superseding effect to the occurrence of a cross-agent effect. And in the previous study I did not find a significant cross-agent effect, although I did see a difference in mean ratings corresponding with a small effect size, as is clear in looking at the histograms in Figure 8. As such, it is not surprising (on my account) that there was not a significant superseding effect. That said, as the histograms make clear, there was a
shift in responses between the two conditions, and the shift is suggestive of a superseding effect. To better test the causal sufficiency and responsibility accounts, in my tenth experiment I used a disjunctive case where the outcome is more clearly bad—a version of the Lauren and Jane Case in which a third employee also logs into the mainframe.

5.2.1 Methods

Results were collected from 86 participants. Participants were given one of two modified versions of the Lauren and Jane Case in which I added a third employee, Helen, who also logs in. In the updated vignettes, Helen is the fixed agent and does not violate the injunctive norm. Lauren and Jane are both varied agents: in the first scenario neither violates the injunctive norm, in the second scenario they both violate the injunctive norm. As previously, in each scenario it is specified that the computer system will crash if more than one person logs in. Since there are three people who log in, however, the system would still have crashed even if any given one of them did not. This means that sufficiency is not obviously threatened (although see below for a worry). In both conditions participants were asked to rate how much they agreed or disagreed with each of three causal statements—“Helen caused the system to crash,” “Lauren caused the system to crash,” and “Jane caused the system to crash”—on the same 7-point scale used previously.

5.2.1.3 Results

Results are shown in Figure 9. I conducted planned comparisons to assess each of the three effects identified in Section 1. I found both a significant cross-agent effect between Lauren and

---

28 66.3% women, average age 30.1 years, ranging from 16 to 71. I ran the three ANOVAs, with whether the varied agents violated the norm and gender as between-participant factors. No significant gender effects were found, although there was a borderline significant main effect for gender on agreement ratings for the fixed agent, $F(1, 82)=3.12, p=0.081$. Given the low mean rating for the fixed agent and the large effect sizes detailed below, this was treated as irrelevant for present purposes.
Helen, $t(43)=10.28$, $p=3.7\cdot 10^{-13}$, Cohen’s $d=2.30$, and between Jane and Helen, $t(43)=10.40$, $p=2.63\cdot 10^{-13}$, Cohen’s $d=2.32$. I also saw a significant varied agent effect for both Lauren, $t(83.67)=4.69$, $p=1.0\cdot 10^{-5}$, Cohen’s $d=1.01$, and Jane, $t(83.66)=5.16$, $p=1.7\cdot 10^{-6}$, Cohen’s $d=1.11$.

And I found a significant superseding effect for Helen with a large effect size, $t(69.44)=-4.48$, $p=2.9\cdot 10^{-5}$, Cohen’s $d=-0.98$. The mean rating for Helen when both Lauren and Jane violated the injunctive norm (M=1.55, SD=1.37, N=44) was 1.74 points lower than when neither violated the norm (M=3.29, SD=2.13, N=42).

**Figure 9**: Results for Study 10 with histograms to the right of the plot of the means for each condition and showing 95% confidence intervals.
5.3 Study 11: Disjunctive Email Case

The occurrence of a superseding effect in the Lauren, Jane, and Helen Case pushes against prediction three of the causal sufficiency account. Further, it is problematic for Hitchcock and Knobe’s account, which the causal sufficiency account builds off. To see why, recall that the intervention account is based around the idea that people arrive at causal attributions by considering counterfactuals on which the outcome does not occur. Hitchcock and Knobe then argue that norms play a role in this process because people are more likely to consider counterfactuals in which an abnormal event is replaced with a more normal one. But in the Lauren, Jane, and Helen Case, the outcome will occur regardless of whether either one of the norm-violating agents acted or not. That is, restricting counterfactuals to manipulations of individual actions, there does not seem to be a relevant counterfactual for the account to call on to explain the impact of norms on this case.

It might be argued, however, that the counterfactuals people consider are not restricted to individual actions, even in cases like the Helen, Lauren, and Jane Case where we have discrete agents acting independently of one another. As such, it might be claimed that the relevant counterfactual in this case is the one on which neither Lauren nor Jane violate the norm. And on such a counterfactual, the outcome would not have occurred, since Helen alone logging in would not produce the outcome. While this response saves the intervention and causal sufficiency accounts, it raises new questions about when and why discrete agents are bundled together in a single counterfactual. This calls out for explanation. And the need for explanation is pressing since we might wonder why the same thing does not occur for Lauren and Jane in the original scenario, for instance. It might be argued that the injunctive norm plays a further role—that rule breakers get lumped together, such that Lauren and Jane form a pair in the revised case but not in the original. But this too requires some prior motivation. Rather than pursue the point further
here, however, I turn to a disjunctive case involving just two agents, testing a disjunctive version of the Email Case from Kominsky et al.’s second experiment.

5.3.1 Methods

Results were collected from 85 participants. Participants were given one or the other of the two modified versions of the Email Case. In the first vignette the varied agent (Billy) does not violate an injunctive norm, in the second vignette he does. In each vignette the fixed agent (Suzy) does not violate an injunctive norm. In each vignette it is specified that important emails will be deleted if anyone logs into a central computer and that both agents log in. In both conditions, participants were asked to rate how much they agreed or disagreed with each of two causal statements—“Billy caused the e-mails to be deleted” and “Suzy caused the e-mails to be deleted”—on the same 7-point scale used previously.

5.3.1.3 Results

Results are shown in Figure 10. I conducted planned comparisons to assess each of the three effects identified in Section 1. I found both a significant cross-agent effect, \( t(39)=3.83, p=0.00045 \), Cohen’s \( d=0.80 \), and a borderline significant varied agent effect, \( t(77.57)=1.67, p=0.099 \), Cohen’s \( d=0.37 \). And I found a significant superseding effect, \( t(80.66)=02.25, p=0.027 \), Cohen’s \( d=-0.49 \). The mean rating for Suzy when Billy violated the injunctive norm (\( M=2.83, SD=2.18, N=40 \)) was 1.04 points lower than when Billy did not violate the norm (\( M=3.87, SD=2.07, N=45 \)). Unlike for the previous result, this effect cannot be explained in terms of a counterfactual changing multiple actions, at least not in a way that is consistent with the

\[^{29}\] 55.3% women, average age 30.0 years, ranging from 16 to 83. I ran two ANOVAs, with whether the varied agent violated the norm and gender as between-participant factors. No significant gender effects were found.
intervention and causal sufficiency accounts. As such, the results of this study put severe pressure on these accounts.

![Figure 10: Results for Study 11 with histograms to the right of the plot of the means for each condition and showing 95% confidence intervals.](image)

6. Conclusion

Current evidence indicates that norms play a role in ordinary causal attributions, with causal ratings typically being higher when an agent violates a norm than when she doesn’t. One leading view, Hitchcock and Knobe’s intervention account, explains these findings in terms of normative information having an effect on the counterfactuals that people consider. Kominsky et al. extend this view to explain a further effect of norms—the causal superseding effect—in which the normative status of one agent’s action affects causal attributions for another agent. And they
provide evidence supporting their account, which in turn provides support for the intervention account. We have seen, however, that while there is reason to think that the superseding effect does occur in some cases, it is also quite variable. Further the results I have reported suggest that it does not occur with equal strength for cases involving good outcomes and bad outcomes, that it doesn’t occur for all types of norm violations, and that it sometimes occurs for disjunctive cases as well as for conjunctive cases. These findings run counter to the predictions of Kominsky et al.’s causal sufficiency account as well as the intervention account.

In contrast, my results are in line with the predictions of the alternative responsibility account, at least when coupled with a deflationary explanation of the superseding effect. The responsibility account contends that the effect of norms on ordinary causal attributions reflects that people are employing a concept with a normative component when they make these judgments, such that their causal attributions are more akin to asserting that the agent is responsible for the outcome than simply that the agent brought about the outcome. My alternative deflationary explanation of the superseding effect is that it is a context effect arising from relative differences in the perceived causal strength of the agents affecting how participants interpret the scale used to elicit their causal attributions. Together, the responsibility account and this context effect hypothesis predict that the superseding effect will be variable across cases, that it will not typically occur with equal strength for cases involving good outcomes and bad outcomes, that it will not occur for all types of norm violations, and that it will sometimes occur for disjunctive cases. And my results support each of these predictions.
References


Livengood, J., and J. Sytsma (under review). “Actual causation and compositionality.”


Sytsma, J. and J. Livengood (under review). “Intervention, bias, responsibility… and the trolley problem.”