Structure and Norms:
Investigating the Pattern of Effects for Causal Attributions[1]

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 by performing symmetric actions, but with one violating a norm while the other does not, causal ratings are higher for the norm-violating agent. A number of competing explanations of this effect have been offered in the literature. In a pair of recent papers, Kominsky et al. (2015) and Icard et al. (2017) make a strong case for one of these accounts—the counterfactual view—making novel predictions about the pattern of effects seen when the original type of case is expanded to include a contrast case without norms or when the causal structure is changed. They argue that while the counterfactual view is able to explain each of the predicted effects, the alternative accounts are only able to explain some of them. In this paper, I argue that this undersells the competing accounts. Further, I present new evidence suggesting that the expanded pattern of effects is quite different than predicted, and that it in fact coheres better with prominent alternative accounts in the literature than the counterfactual view.

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, Icard et al. 2017, Kominsky and Phillips 2019, Livengood and Sytsma forthcoming).[2] Recent interest in the topic was sparked by a series of findings by Knobe and Fraser (2008) and Hitchcock and Knobe (2009). These papers presented the results of three experiments with a common structure. In each, participants were given a single vignette involving two "agents" performing symmetric

[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, Jonathan Kominsky, and two anonymous reviewers for extremely helpful feedback on earlier drafts 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. 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.

actions that jointly brought about a bad outcome.[3] The causal structure in each was *conjunctive*: if either of the agents had not performed the action, the outcome would not have occurred. The principle difference in the vignettes was that one agent violated a norm and the other did not. And in each case the researchers found that despite the two actions being jointly necessary for the outcome, causal ratings were notably higher for the agent violating the norm than for the agent who did not violate the norm. I'll refer to this as the *cross-agent effect* to highlight that it is comparing two different agents in the same scenario.

Hitchcock and Knobe (2009) offered an ingenious explanation of the cross-agent effect that hinges on the salience of different counterfactuals. This *counterfactual view* has been further developed since then, including in recent papers by Kominsky et al. (2015) and Icard et al. (2017) that focus on several further effects of norms on causal attributions.[4] They argue that the occurrence of these further effects lends support to the counterfactual view and raises problems for alternative accounts that have been put forward in the literature, including the *bias view* (Alicke 1992, 2000; Alicke et al. 2011; Rose 2017), the *pragmatic view* (Samland and Waldmann 2016, Samland et al. 2016), and our *responsibility view* (Sytsma et al. 2012, Livengood et al.2017, Sytsma et al. 2019, Livengood and Sytsma forthcoming).

To elicit these further effects, we begin by adding a second condition in which neither agent violates a norm to experiments like those conducted by Knobe and Fraser (2008) and Hitchcock and Knobe (2009). The result is that the normative status of one agent is varied between the two conditions (the *varied agent* violates a norm in one condition but not the other), while the normative status of the other agent is fixed between the two conditions (the *fixed agent* doesn't violate the norm in either condition). Each of the views noted above predict that causal ratings will be higher

[3] Hitchcock and Knobe's second experiment involved two wires rather than agents, although I will focus on cases involving agents here.
[4] See also Halpern and Hitchock (2015), Phillips et al. (2015), Kominsky and Phillips (2019).

for the varied agent when she violates the norm than when she does not. This effect is sometimes referred to as *abnormal inflation* (e.g., Icard et al. 2017), but for clarity I will refer to it as the *varied agent effect* to emphasize that we are looking at the change in ratings for the varied agent across the two conditions.

More interestingly, Kominsky et al. (2015) note that we can also compare the ratings of the fixed agent between the two conditions, predicting that we will see the reverse effect—causal ratings for the fixed agent being *lower* when the varied agent violates the norm than when the varied agent does not violate the norm. Kominsky et al. term this the *causal superseding effect*, while Icard et al. call it the *supersession effect*. In keeping with the previous two effects, however, I'll refer to it as the *fixed agent effect* to highlight that we are looking at the change in ratings for the fixed agent across the two conditions. Although Kominsky et al. are the first to highlight this effect, it can be found in the prior literature, although as discussed below it is not seen in all cases.

Adding a further wrinkle, we can change the causal structure in these scenarios from *conjunctive* to *disjunctive*: rather than the two actions being jointly necessary for the outcome as in a conjunctive case, in a disjunctive case either action alone is sufficient to bring about the outcome. In such cases, Kominsky et al. predict that there will be no fixed agent effect and Icard et al. predict that the varied agent effect will be reversed. And they present evidence supporting these predictions. Although neither paper notes the cross-agent effect, from the other two effects we can infer that they should predict a *reverse effect* in disjunctive cases where the actions are otherwise symmetric.

This gives an overall pattern of six effects—cross-agent, varied agent, and fixed agent effects in conjunctive cases and their reversal (cross-agent, varied) or absence (fixed agent) in disjunctive cases. Assuming this pattern of effects is accurate, it arguably provides strong support for the counterfactual view, while raising problems for prominent competing accounts. First, it is unclear how the alternative accounts can explain the fixed agent effect in conjunctive cases, since

the fixed agent does not violate a norm in either condition. Second, it is unclear how these views can explain the reverse cross-agent and varied agent effects in disjunctive cases, given that the same normative considerations apply here as do in conjunctive cases.

In this paper I will offer responses to each of these two problems. In response to the first, I offer an alternative explanation of the fixed agent effect that is consistent with the alternative accounts found in the literature: I argue that the fixed agent effect might be nothing more than a context effect, with the varied agent providing the relevant context. While this *context effect hypothesis* is able to explain the occurrence of the fixed agent effect in conjunctive cases, it is not able to simultaneously explain the absence of the effect reported by Kominsky et al. and the presence of the reverse varied agent effect report by Icard et al. for disjunctive cases, raising the second problem. In response, I begin by arguing that this problem does not obviously apply to the alternative accounts. Focusing on the responsibility view, I note that to show the problem applies would require testing responsibility attributions for disjunctive cases, which to the best of my knowledge has not previously been done. I then present evidence that raises doubts about the overall pattern of evidence. Starting with two widely discussed conjunctive cases, I find that the fixed agent effect does not occur reliably for injunctive norms; further, in line with the responsibility view but against the counterfactual view, I find that neither the varied agent nor fixed agent effects occur for one type of descriptive norm, while tending to be *reversed* for a second. I then turn to disjunctive cases, testing two cases used by Kominsky et al. and Icard et al., and failing to find the predicted reverse varied agent effect. Finally, I test responsibility attributions for the same disjunctive cases and find that the results are consistent with alternative accounts. I conclude that the present evidence not only does not clearly favor the counterfactual view but is more consistent with competing views.

Here is how I will proceed. In Section 1, I detail the pattern of effects predicted by the counterfactual view and consider the argument that the full pattern is problematic for alternative accounts. In Section 2, I argue that while alternative accounts are unable to *directly* explain the fixed agent effect, there is a ready alternative explanation—that the fixed agent effect is a context effect—that is compatible with these views. I then further investigate the proposed overall pattern of effects, starting with conjunctive cases (Section 3), before turning to disjunctive cases (Section 4).

1. Pattern of Effects

A large body of research has shown that norms matter for ordinary causal attributions, generating debate over how best to explain these findings. Much of this discussion initially centered on the cross-agent effect. To illustrate, consider the Pen Case given by Knobe and Fraser (2008):

> The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take the pens, but faculty members are supposed to buy their own.
>
> The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly e-mailed them reminders that only administrative assistants are allowed to take the pens.
>
> On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message… but she has a problem. There are no pens left on her desk.

Here we see that two agents perform symmetric actions (taking pens), jointly bringing about a bad outcome (no pens are left for the receptionist). Further, both actions were necessary for the outcome to occur—the outcome would not have occurred if either agent had not taken a pen—making this a *conjunctive* case. The only notable difference between the two actions is that Professor Smith violated an injunctive norm in taking a pen, while the administrative assistant did not.

Knobe and Fraser asked participants two questions about this vignette—whether Professor Smith caused the problem and whether the administrative assistant caused the problem. The results

showed a notable difference, with ratings for Professor Smith being significantly higher than for the administrative assistant. This is the *cross-agent effect*. Similar conjunctive scenarios were subsequently tested by Hitchcock and Knobe (2009)—one involving agents jointly bringing about a good outcome, one involving two parts of a machine jointly bringing about a bad outcome—with similar results: causal ratings for the norm-violating "agent" are significantly higher than for the norm-conforming "agent."

Hitchcock and Knobe offer an explanation of the effect of norms on causal attributions that emphasizes counterfactuals. They explain 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. This *counterfactual view* has subsequently been developed in a number of papers, including by Kominsky et al. (2015) and Icard et al. (2017), with the focus shifting from the cross-agent effect to other effects that can be tested when scenarios like the Pen Case are extended to remove the norm violation and when the causal structure is changed from conjunctive to disjunctive.

1.1 Non-normative Contrast

When giving participants the version of the Pen Case discussed above, the only comparison that can be made is between the two agents. Subsequent work has expanded on this, adding a comparison condition in which neither agent violates a norm. For instance, Sytsma et al. (2012) tested both the

original Pen Case and a version in which the injunctive norm was removed, making it so that there was no departmental policy with regard to taking pens.[5] Looking across these two versions, the normative status of one agent's action is *varied* (in one scenario Professor Smith violates a norm, in the other she does not) while the normative status of the other agent's action is *fixed* (the administrative assistant does not violate a norm in either scenario). There are now two further comparisons that could be made—we could compare ratings for the *varied agent* across the two conditions and we could compare ratings for the *fixed agent* across the two conditions.

Comparing ratings for the varied agent in Sytsma et al. (2012), the mean for Professor Smith is significantly higher when she violated the norm (m=4.05, sd=2.14, n=59) than when she did not (m=3.00, sd=1.80, m=45), $t(100.99)=2.72$, $p=0.0038$, $d=0.53$ ($W=1704.5$, $p=0.0061$).[6] This is the *varied agent effect*. A similar comparison can be carried out for the fixed agent. Given that the fixed agent's action is described the same way in each scenario—she doesn't violate a norm in either case—we might expect that the causal ratings will be roughly the same. But, in fact, there is a comparable difference to that seen for the varied agent, although running in the opposite direction: the mean rating for the administrative assistant was significantly *lower* in the condition where Professor Smith violated the norm (m=2.51, sd=1.81, n=59) compared to the condition where Professor Smith did not violate a norm (m=3.53, sd=2.01, m=45), $t(89.531)=2.69$, $p=0.0043$, $d=0.54$ ($W=1735.5$, $p=0.0030$). This is the *fixed agent effect*.

[5] This was part of a larger set of variations tested where the typicality of the agents' actions was also varied, as discussed in Section 3.

[6] Sytsma et al. did not test the comparisons discussed in this section, so I have reanalyzed the data. Since the effects at issue are all directional—for instance, the fixed agent effect occurring when ratings for the fixed agent are *lower* when the varied agent violates a norm than when the varied agent does not violate a norm—I'll use one-tailed tests throughout this paper. And while it is typical to test for effects of norms on causal attributions using parametric statistics, the distributions are often non-normal, as will be clear below; as such, I will report non-parametric tests (either Wilcoxon signed rank tests or Wilcoxon rank sum tests) in addition to the standard t-tests.

1.2 Counterfactual Sufficiency and Disjunctive Cases

While the fixed agent effect can be seen in Sytsma et al. (2012), we did not note this effect in our paper. To the best of my knowledge, Kominsky et al. (2015) are the first to highlight the effect and to offer an explanation of it—the *counterfactual sufficiency account*. As the name suggests, Kominsky et al.'s explanation follows from the counterfactual view. As with the counterfactual view, the counterfactual sufficiency account starts with the claim 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). This makes the counterfactual in which she 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 fixed agent effect.

What is perhaps most interesting about Kominsky et al.'s account is that it not only makes a prediction about when the fixed agent effect should occur, but also when it should *not* occur.

Specifically, Kominsky et al.'s account predicts that the effect should only occur when the sufficiency condition is threatened. This means that it should occur in *conjunctive cases,* where either action alone is insufficient to bring about the outcome and that it should *not* occur in disjunctive cases where either action alone is sufficient. Kominsky et al. then provide evidence that the fixed agent effect occurs in conjunctive cases and not in disjunctive cases. Intriguingly, this includes an experiment where a statistical norm applies to a non-agent[7], highlighting a further prediction of the counterfactual view: the pattern of effects should be found for both injunctive and descriptive norms.

Icard et al. (2017) build on the work of Kominsky et al., offering a sophisticated account of causal judgements that involves probabilistically sampling counterfactual scenarios. This account leads them to a further prediction: the varied agent effect will be *reversed* in disjunctive scenarios (what they term *abnormal deflation*). Again, this prediction begins with the idea that people are more likely to consider counterfactuals on which an abnormal event is replaced with a more normal event. The next step in the prediction calls on considerations of *necessity*: Icard et al. suggest that people will be reluctant to judge that an agent caused an outcome when they recognize that the agent's action was not necessary for the outcome—that is, when they recognize that the outcome would have occurred even if that agent had not acted. And putting these two ideas together generates the prediction: people will be more likely to consider what would have happened if the varied agent had not acted when she violated a norm compared to when she did not violate a norm; and, when people consider this counterfactual in a disjunctive case, they will recognize that the varied agent's action was not necessary for the outcome, making them less likely to judge that she

[7] In Experiment 4, Kominsky et al. looked at a fixed event and a varied event, manipulating the statistical probability of the varied event. Unlike in the their other three experiments, however, 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; 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.

caused it. This prediction is then tested across two experiments, with Icard et al. finding the reverse varied agent effect in disjunctive cases for both injunctive norms (Experiment 1) and descriptive norms (Experiment 2).[8] Further, Kominsky and Phillips (2019) found a comparable effect for disjunctive cases with injunctive norms (Experiment 2). While neither Kominsky et al. or Icard et al. specifically note the cross-agent effect, putting their predictions for disjunctive cases together it appears that they should also expect to find a *reverse* cross-agent effect in scenarios where the agents' actions are symmetric.[9]

Summing up, the counterfactual view predicts a complex pattern of effects: for both injunctive and descriptive norms advocates expect to find cross-agent, varied agent, and fixed agent effects in conjunctive cases and to find that these effects are either reversed (cross-agent, varied agent) or absent (fixed agent) in disjunctive cases. And between Kominsky et al. (2015) and Icard et al. (2017), they provide direct evidence for all but one of these effects: since Kominsky et al. only tested judgments about the fixed agent in disjunctive cases and Icard et al. only tested judgments about the varied agent[10], neither is able to show the occurrence of a reverse cross-agent effect. However, since both papers tested a disjunctive version of the Motion Detector case (discussed in more detail in Section 4) we can indirectly check this prediction and it appears to be borne out.[11]

1.3 Argument for the Counterfactual View

That the counterfactual view can explain the full pattern of effects just noted is taken to be a point in its favor; in contrast, it is argued that competing accounts are only able to explain some of the

[8] As in Kominsky et al.'s Experiment 4, the second experiment asked about "because" statements involving events.
[9] On the standard assumption that there will be no cross-agent effect in the non-normed condition in such scenarios, if there is no fixed agent effect and a reverse varied agent effect, it follows that in the normed condition causal ratings for the varied agent will be *lower* than for the fixed agent.
[10] The same holds for Experiment 2 in Kominsky and Phillips (2019).
[11] Kominsky et al. report of mean of 4.53 for the fixed agent in Experiment 3, while Icard et al. found a mean of 3.24 for the varied agent in Experiment 1.

effects, raising serious problems for such views. I'll focus on two potential problems. The first is that it might seem that the alternative accounts are not able to explain the fixed agent effect in conjunctive cases. The second is that it might seem that the alternative accounts are not able to explain the reverse cross-agent and varied agent effects in disjunctive cases.

To draw out the reasoning behind these two objections, and the responses that I'll offer to them, we need to consider the basic ideas behind the primary alternative accounts that have been put forward. Kominsky et al. (2015, 208) consider two broad types of approach to explaining the impact of norms on ordinary causal attributions:

> Previous work has been divided on whether we should treat moral considerations as (a) playing a role in the operation of people's causal cognition itself or (b) introducing some external bias or pragmatic factor that is skewing the results of what is in fact a purely non-moral cognition system.

Similarly, Icard et al. distinguish their counterfactual view from a broad family of accounts that hold that "the impact of [injunctive] norms arises because people's causal judgments are influenced by judgments that particular agents are blameworthy" (88), and they treat the bias, pragmatic, and responsibility views noted above as all falling under this family of accounts.[12]

To draw out the two objections, consider a simple view falling under this family of accounts: people tend to form negative evaluations of agents when they violate injunctive norms and these

[12] These three alternatives do not exhaust the potential explanations, although they are the most prominent views in the present literature and the ones I will focus on. As an anonymous reviewer helpfully noted, there is a longer history of attribution research that is relevant to these debates. Most importantly, Heider (1958) distinguishes between *impersonal causality* and *personal causality*, with the key distinction being between those behaviors perceived as unintentional, which are explained in terms of the situations people were in, and those perceived as intentional, which are explained in terms of the actor's reasons. Such a distinction might be called on to explain the effect of norms on causal ratings, with the norm-violating agent action being perceived as intentional while the norm-conforming agent's action is perceived as unintentional. It is unclear how straightforwardly this distinction can be applied to scenarios like those presently at issue, however. For instance, it seems that in the original version of the Pen Case, both Professor Smith and the administrative assistant intentionally take pens, and that neither of them had the intention of bringing about the problem for the receptionist. Further, while the work of Heider has often been interpreted as making a simpler distinction between person-causes and situation-causes (see Malle 2011, Section 3), there are serious problems with such an account as Malle notes (see Section 7.2). With regard to the Pen Case, in line with the above considerations, it seems that focusing on the agents' reasons for taking pens, we could treat either as person-causes, but focusing on the external demands on these agents' we could alternatively treat either as agent-causes.

evaluations in turn tend to influence their causal judgments, such that they are more likely to treat a norm-violating agent as having caused a bad outcome. This view can readily explain the cross-agent and varied agent effects in conjunctive scenarios, since the effects lineup with the norm violations. But what about the fixed agent effect? Since the fixed agent doesn't violate a norm in either condition—since the fixed agent doesn't do anything wrong—it seems that the simple view has no direct explanation for this effect. Kominsky et al. do not explicitly consider this worry, but they do note that "if morality functioned as an outside bias that skewed our causal judgment, it would be somewhat surprising if that bias operated on only some causal structures and not others" (2015, 208). As seen above, they provide evidence that the fixed agent effect occurs in conjunctive cases but that it does not occur in disjunctive cases. Since the simple view would not seem to predict a fixed agent effect in either type of case, however, it is only the occurrence of the effect in conjunctive cases that is problematic. This is the first problem.

Turning to the cross-agent and varied agent effects, the issue is reversed for the simple view: while it can explain the occurrence of these effects in conjunctive cases, it does not seem to be able to explain the reverse effects in disjunctive cases, since the same pattern of norm violations occur but the causal attributions are quite different. The simple view explains the effect of norms on causal attributions in terms of norm violations leading to negative evaluations of the agent, but such evaluations do not seem to depend on causal structure. This is the second problem.

The first problem is a live issue not just for the simple view, but for the alternative accounts noted by Icard et al. more generally. It is at best unclear that the same holds for the second problem, however. The reason is that unlike the simple view, the responsibility, bias, and pragmatic views each allow that causal structure will matter for causal attributions. I will focus on explaining why this holds for the responsibility view but will note how similar considerations also apply to the other two views.

1.4 Alternative Accounts and Causal Structure

While there are some clear similarities between the responsibility, bias, and pragmatic views, it is important to note that unlike the latter two the responsibility view does *not* hold that "external bias or pragmatic factors" are "skewing the results of what is in fact a purely non-moral cognitive system." Instead the responsibility view holds that the evaluation of norms is directly relevant to the correct application of the ordinary concept of causation typically at play in causal attributions (at least in English) because that concept itself has a normative component. As such, the responsibility view is better thought of as belonging to Kominsky et al.'s family (a). The basic idea is that causal attributions typically serve to indicate something more than simply that an entity brought about an outcome; they also express a normative evaluation of the action similar to saying that the entity is responsible for or accountable for the outcome.[13] This stands in contrast to the bias view, which holds that the ordinary concept of causation is non-normative but that evaluations of the agents bias the application of this concept. And it stands in contrast to the pragmatic view, which holds that when participants answer questions about causal attributions in studies like those discussed above, they interpret the questions not as asking about causation, but about a separate notion such as accountability or responsibility.

A key point to note here is that unlike the simple view discussed above, the responsibility view does not focus on evaluations of just an agent, but on judgments that the agent caused *an outcome*, taking these to be similar to judgments that the entity is responsible *for the outcome*.

[13] While the cases focused on in this paper involve agents, the responsibility view is not specifically focused on agents, but allows that similar considerations will apply to (what philosophers take to be) non-agentive objects. This contrasts with the way Samland and Waldman think about responsibility and related attributions. They hold that such attributions are specific to agents—"that people, not objects, are held accountable for a negative outcome" (2016, 175). Research by Bloom (2007), Rose (2015), and Rose and Schaffer (2017), among others, suggests that people tend to take an agentive perspective on nature as a whole. This difference between the pragmatic and responsibility views is especially relevant to an important criticism raised by Kominsky and Phillips (2019). In their fourth experiment, they find that violations of norms of proper functioning in objects (artifacts malfunctioning) affect causal attributions similarly to violations of injunctive norms for agents, and argue that this cannot be explained by the pragmatic account, which focuses on *moral* responsibility, assuming that such considerations do not apply to artifacts. These results are consistent with the responsibility view, however, which doesn't assume such a restriction.

While negative evaluations of an agent and judgments that the agent is responsible for a bad outcome will sometimes coincide, as they plausibly do in conjunctive cases, this need not be the case: an agent could arouse negative evaluations without being responsible for a bad outcome, and an agent could be responsible for a bad outcome without necessarily arousing negative evaluations (see Sytsma and Livengood under review). The former is perhaps most clear if we imagine a case where an agent violates an injunctive norm that is unrelated to bringing about the outcome at issue. Imagine for instance, that in the Pen Case scenario, there is an honor bar in the philosophy department and that Professor Smith takes a candy bar without paying for it. Imagine further that there is only one pen on the receptionist's desk, that the administrative assistant takes that pen, and that the problem arises without Professor Smith taking a pen. While the simple view might predict that people would judge that Professor Smith caused the problem, the responsibility view does not: Smith might be a jerk, but she isn't responsible for the issue with the pens. The upshot is that it is imperative for our view that the norm violating action is connected to the outcome. And plausibly the same holds for the bias and pragmatic view.[14]

That the responsibility view focuses on responsibility *for an outcome*, not a more general evaluation of the agent, is important for present purpose because while causal structure *might* be taken to be irrelevant to forming general evaluations of an agent, it is very plausibly relevant to evaluating their responsibility for an outcome. The responsibility view expects that causal attributions will typically be similar to responsibility attributions. As such, we can use responsibility attributions to generate initial predictions for causal attributions. Thus, to get clear on whether

[14] This is relatively straightforward for the pragmatic view, which holds that people (miss)interpret the causal attribution questions as asking about responsibility, and presumably those attributions would be understood in terms of responsibility *for an outcome*. The picture is a bit more complicated for the bias view. While the bias view holds that evaluations of the agent will matter—for instance Alicke et al. (2011, 670) write that "causal attributions reflect a desire to praise or denigrate those whose actions we applaud or deride"—the view allows that a variety of factors will go into this, and this explicitly includes the outcome (one might expect evaluations of an agent to be worse when that agent is to blame *for a bad outcome*).

causal attributions for disjunctive cases are problematic for the responsibility view, we need to test responsibility attributions for such cases, as is done in Section 4. Prior to looking at the empirical results, however, there is reason to expect that responsibility attributions will be different in disjunctive scenarios than in matching conjunctive scenarios. To begin with, this seems intuitively plausible. Responsibility for a bad outcome would seem to be most clear when someone does something wrong and but for that misdeed the outcome would not have occurred (as holds in conjunctive cases). In contrast, responsibility would seem to be at least partially mitigated if the outcome would have occurred regardless of the misdeed (as holds in disjunctive cases).[15]

There is also empirical reason to expect a connection, here. For instance, Gailey and Falk (2008, 674) find that "it appears as though people, when asked to attribute responsibility for wrongdoing, examine four aspects: the cause, prior knowledge, intentions of the actor, and whether the actor appreciated the moral wrongfulness of the act." The questions making up the "cause" factor here did not ask for causal attributions, however, but included questions related to causal structure—whether the agent could have *avoided* the outcome and whether anything else could have *prevented* the outcome. Further, causal structure will often be relevant to inferring the knowledge and intentions of the agents in a causal scenario, and both have been found to impact responsibility and related attributions (e.g., Cushman 2008, Gailey and Falk 2008, Lagnado and Channon 2008, Young and Saxe 2011, Malle et al. 2014). One way to draw this out is to note that the connection between injunctive norm and outcome varies with causal structure: in conjunctive cases, the norm

[15] This is congruent with standard thinking about causation in the law, where the "but-for rule"—but for the action, the harm would not have occurred—is "by far the dominant explicit test for cause in fact in both torts and criminal law" (Moore 2019, Section 2). And the law typically counts conjunctive but not disjunctive, or over-determined, cases as exhibiting cause in fact. As Moore notes, cause in fact is one of "two very different requirements for liability" in the law, with the second being legal or proximate cause. While the concept of proximate cause is less clear, we suspect it is close to the ordinary use of "cause" in causal attributions, and in line with the responsibility view its application is an "evaluative issue." Typically, proximate causation judgments are taken to require establishing both cause in fact and some form of culpability. As a result, causal structure will matter for proximate causation judgments, with it being easier to uphold such judgments in conjunctive cases than in disjunctive cases.

provides a means of preventing the outcome—if the agent had abided by the norm, then the outcome would not have occurred. In disjunctive cases, though, the norm does not do so—whether or not the agent had abided by the norm, the outcome would have occurred. The more closely a norm is connected to the outcome, however, the more plausible the inference that the norm-violating agent either knew or should have known the risk.

In fact, there is empirical research demonstrating that casual structure matters for people's responsibility attributions (Gerstenberg and Lagnado 2010, 2012; Lagnado et al. 2013; Zultan et al. 2012).[16] As Lagnado and Gerstenberg (2017, 586) summarize, their criticality-pivotality model of responsibility attributions predicts that "people's responsibility judgments increase, the more critical a person's action was perceived to be for the outcome." This is illustrated by considering two simple voting scenarios involving two agents—one with a conjunctive structure (the motion will only pass if both agents vote in favor of it), one with a disjunctive structure (the motion will pass so long as at least one votes in favor of it). Lagnado and Gerstenberg then note that each agent's action is less critical in the disjunctive scenario than in the conjunctive scenario, and hence that their model predicts that the agents will also be judged to be less responsible for the outcome in the disjunctive scenario than in the conjunctive scenario. Pulling these points together, while work specifically on responsibility attributions for disjunctive cases involving injunctive norms is needed for the responsibility view to make more concrete predictions about such cases, we would at minimum predict that causal attributions for the varied agent will tend to be lower when she violates the norm in a disjunctive case compared to when she does so in a matching conjunctive case.

[16] While Gerstenberg and Lagnado often discuss this in terms of *causal* responsibility, the key questions they ask just concern responsibility. And while philosophers often distinguish between causal and moral responsibility, it is unclear whether such a distinction features prominently in ordinary causal attributions. We suspect that the ordinary concept of responsibility at play in responsibility attributions is a normative concept, and that the judgments elicited by Gerstenberg and Lagnado's prompts primarily involve this concept.

Summing up, the pattern of results reported by Kominsky et al. and Icard et al. raise two potential problems for prominent alternatives to the counterfactual view. The first problem is that these views do not seem to be able to offer a direct explanation for the occurrence of the fixed agent effect in conjunctive cases. While I take this to be a clear worry for alternative accounts, I believe there is a ready response: the fixed agent effect can be explained indirectly as a context effect where the relevant context is provided by the differences in causal judgments concerning the varied agent. This is laid out in the next section. The second problem is that the alternative accounts do not seem to be able to explain the reverse cross-agent and varied agent effects in disjunctive cases. As just discussed, it is not clear whether this is a live worry for accounts like the responsibility view, since we would expect causal structure to be relevant to responsibility judgments. Nonetheless, the alternative accounts do not obviously predict the overall pattern of effects, while the counterfactual view does, which all else being equal would suggest in favor of the counterfactual view. To explore this objection further, I further investigate the purported pattern of effects, starting with conjunctive cases in Section 3 and extending this to disjunctive cases in Section 4.

2. Alternative Explanation: Context Effect

As Icard et al. (2017) suggest, the various competing accounts of the effect of norms on ordinary causal attributions can explain the cross-agent and varied agent effects in conjunctive cases. It is less clear that the alternative accounts can directly explain the fixed agent effect in conjunctive cases, however. Nonetheless, while the alternative accounts are arguably unable to offer a direct explanation for the effect, they can offer an indirect explanation. And one explanation readily presents itself—that the fixed agent effect is merely an artefact. Specifically, I suggest that the fixed agent effect is a 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.

One relevant illustration comes from Schwarz, Münkel, and Hippler (1990). In their experiment, 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 "fixed agent 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*).

Consider how such 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 (the varied agent does not violate the norm) or relatively extreme (the varied agent violates the norm). Based on the known 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 what is at issue for the fixed agent effect. As such, it is *prima facie* plausible that the effect is nothing more than a context effect, and such a context effect is consistent with the alternative accounts found in the literature. To illustrate that the fixed agent effect might

simply be a context effect, I ran two studies as a proof of concept, showing that a comparable context effect can arise using questions like those employed in Kominsky et al.'s experiments.

2.1 Study 1: Height Judgments, Two Claims

2.1.1 Methods

Participants for each study in this paper were recruited through advertising for a free personality test on Google. After answering the 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 who were 16 years of age or older. In Study 1, participants were 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 in Kominsky et al.'s experiments. The second claim was the same in each condition—"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 a 7-point scale anchored at 1 with "strongly disagree," at 4 with "neutral," and at 7 with "strongly agree." Results were collected from 133 participants.[17]

2.1.2 Results

Results are shown in Figure 1 and tests for key comparisons are summarized in Table 1. Comparing ratings for Tom Cruise, the mean was significantly lower when the other claim was about Shaquille O'Neil than when it was about Danny DeVito. This is analogous to the fixed agent effect, with variation in the height of one member of the pair affecting height judgments about the other. There was also a significant "cross-agent effect," and a significant "varied agent effect."

[17] 64.7% women (three non-binary), average age 35.1 years, ranging from 16 to 86. One-way ANOVAs showed no effect for gender on either ratings of the "varied agent," $F(2, 130)=0.33$, $p=0.72$, $\eta^2=0.005$, or the "fixed agent," $F(2, 130)=0.045$, $p=0.96$, $\eta^2=0.001$.

2.2 Study 2: Height Judgments, One Claim

Kominsky et al. report a fixed agent effect not just when they asked about both agents, but also when they asked about just the fixed agent. As such, I ran a variation on the previous study asking about just the "fixed agent."

2.2.1 Methods

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 the previous study. 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." Results were collected from 101 participants.[18]

Study	Cross-agent Effect	Varied Agent Effect	Fixed Agent Effect
1	$t(63)=13.40$, $p=2.2e^{-16}$, $d=1.67$	$t(127.39)=16.60$, $p<2.2e^{-16}$, $d=2.85$	$t(130.53)=3.43$, $p=0.00041$, $d=0.59$
	$V=1905.5$, $p=2.6e^{-11}$	$W=4191$, $p<2.2e^{-16}$	$W=1529$, $p=0.00087$
2			$t(94.507)=4.35$, $p=1.7e^{-5}$, $d=0.86$
			$W=711$, $p=4.5e^{-5}$

Table 1: Tests of analogs of cross-agent, varied agent, and fixed agent effects for Studies 1 and 2. All tests one-tailed, significant results in red.

2.2.2 Results

Results are shown in Figure 1. As seen in Table 1, comparing across the two conditions, the mean ratings for Cruise were significantly lower when participants were asked to consider O'Neil than when they were asked to consider DeVito. Again, this is analogous to the fixed agent effect, with variation in the height of one member of the pair affecting height judgments about the other.

[18] 69.3% women (one non-binary), average age 30.4 years, ranging from 16 to 69. A one-way ANOVA showed no effect for gender, $F(2, 98)=0.28$, $p=0.76$, $\eta^2=0.006$.

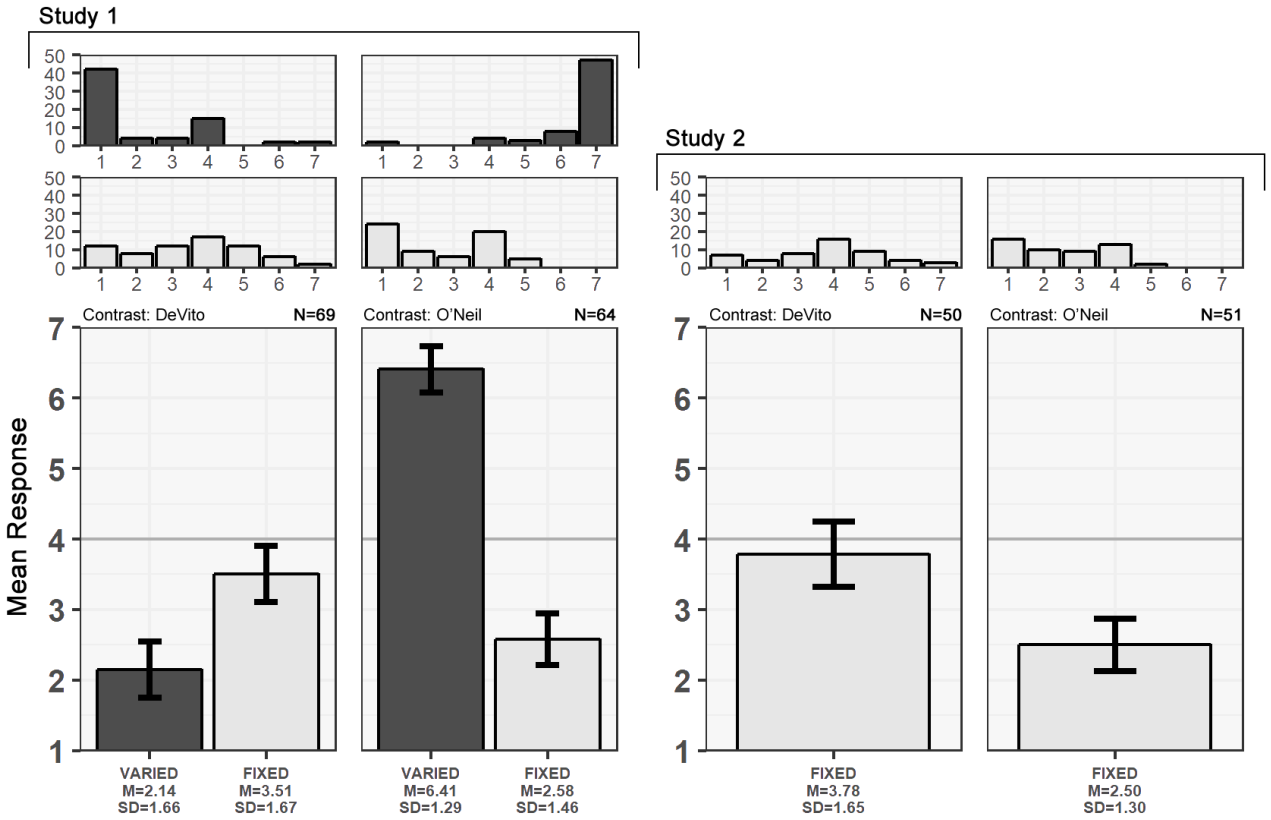

Figure 1: Results for Studies 1 and 2 with histograms above the plot of the means for each condition and showing 95% confidence intervals.

2.3 Discussion

The results for relative agreement ratings about the height of Tom Cruise in Studies 1 and 2 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 a scale standardly employed in research on causal attributions provides proof of concept for the alternative hypothesis that the fixed agent effect is nothing more than a context effect. Given the occurrence of the varied agent effect in the conjunctive cases tested by Kominsky et al., we should not be surprised to also find a context effect for the fixed agent in those cases. I conclude that absent evidence to the contrary we should be warry of concluding that any further explanation of the fixed agent effect is needed when it occurs alongside the varied agent effect.

3. Assessing the Pattern of Effects: Conjunctive Cases

One problem for alternative accounts of the effect of the effect of norms on causal attributions raised by the work of Kominsky et al. (2015) and Icard et al. (2017) is that these views seem unable to directly explain the fixed agent effect in conjunctive cases. In the previous section, I offered a response, arguing that these accounts can offer an indirect explanation of the fixed agent effect when the varied agent effect occurs. Looking at the larger pattern of effects found by Kominsky et al. and Icard et al. across conjunctive and disjunctive cases, however, raises the further question of why causal structure matters. One issue is that the fixed agent effect was found in conjunctive cases, but not in disjunctive cases. Accepting the context effect hypothesis, this is not a problem for the alternative accounts on its own. It becomes problematic, however, when combined with Icard et al.'s finding of a reverse varied agent effect in disjunctive cases. Given a reverse varied agent effect, the context effect hypothesis would plausibly predict that we should also see a reverse fixed agent effect. While the overall pattern of effects is indeed critical for assessing the different accounts of the effect of norms on causal attributions in the literature, previous studies in the literature (Sytsma et al. 2012, Livengood et al. 2017) provide reason to doubt that the pattern is as clear as the studies reported by Kominsky et al. and Icard et al. suggest. In this section, I review and replicate these findings on conjunctive cases, raising doubts about the reliability of the fixed agent effect and whether the fixed agent and varied agent effects occur for all types of norms.

3.1 Pen Case

The counterfactual view holds that the effects of norms on ordinary causal attributions should be seen for all types of norms, including both injunctive and descriptive norms. And Kominsky et al. explicitly extend this to the fixed agent effect. Sytsma et al. (2012) investigated the role of descriptive norms in causal attributions for the Pen Case, distinguishing between two types—

population-level statistical norms (whether the agent acts typically relative to the past actions of a relevant population) and agent-level statistical norms (whether the agent acts typically relative to her past actions). We then derived two predictions from the responsibility view: "First, ordinary causal attributions for the Pen Case are insensitive to population-level statistical norms; second, ordinary causal attributions for the Pen Case are sensitive to agent-level typicality, not atypicality" (816).[19] And we presented the results of a series of studies bearing out these predictions.

As noted above, these studies included testing the original version of the Pen Case as well as a version in which the injunctive norm was removed. And, although we did not note it, the results revealed not only a cross-agent effect, but a varied agent effect and a fixed agent effect. In addition to testing these two conditions, however, we also tested conditions in which one or the other of the two types of descriptive norms were also varied. The subsequent results paint a far less rosy picture concerning the occurrence of the varied agent and fixed agent effects for all types of norms.

Looking at the eight conditions from Sytsma et al. in which the fixed agent did not violate a norm, there were ten comparisons in total where Kominsky et al. would expect to see varied agent effects and ten corresponding comparisons where they would expect to see fixed agent effects. Tests for these comparisons are summarized in Table 2. While the varied agent effect occurred reliably when just the injunctive norm was varied (significant effects in all four comparisons), it was not found in two comparisons where just the population-level statistical norm was varied and significant *reverse* effects were found in two comparisons where just the agent-level statistical norm was varied. Turning to the fixed agent effect, it did not occur reliably when just the injunctive norm was

[19] The first prediction is based on the expectation that excuses of the form "everybody was doing it" are not generally taken to mitigate responsibility. The second prediction is based on the expectation that, in contrast, a person's pattern of behaviors will often be relevant to responsibility judgments. Specifically, in cases where an agent could reasonably be expected to know that a bad outcome might result from her behavior, we expect people to be more likely to judge that she is responsible for such an outcome when it occurs if her behavior was not a one-off, but part of a pattern of reckless behavior, since this pattern increases the chance that such a bad outcome would eventually occur. See Sytsma et al. (2012) for extended discussion.

varied (significant result in one of four comparisons, with a borderline significant *reverse* effect in another comparison). Further no effect was found in the two comparisons where just the population-level statistical norm was varied, while a significant *reverse* effect was found in one of two comparisons where just the agent-level statistical norm was varied. Kominsky et al. (2015) gave sizes for the fixed agent effect for three relevant t-tests, showing Cohen's d values of 0.568, 0.691, and 0.898, for an average value of 0.719 (I'll refer to this as an *average fixed agent effect*). Using this value, the tests had a power ranging from 0.950 to 0.989 to detect an average fixed agent effect.

Norms for VA, Condition 1	Norms for VA, Condition 2	Varied Agent Effect	Fixed Agent Effect
Pop-typical, Permissible	Pop-typical, Impermissible	$t(100.99)=2.72$, $p=0.0038$, $d=0.53$	$t(89.53)=1.02$, $p=0.0043$, $d=0.54$
		$W=1704.5$, $p=0.0061$	$W=1735.5$, $p=0.0030$
Pop-atypical, Permissible	Pop-atypical, Impermissible	$t(83.927)=3.29$, $p=0.00074$, $d=0.71$	$t(83.727)=0.19$, $p=0.43$, $d=0.04$
		$W=1302.5$, $p=0.00043$	$W=913$, $p=0.54$
Pop-typical, Permissible	Pop-atypical, Permissible	R: $t(87.913)=0.52$, $p=0.30$, $d=0.11$	$t(87.727)=1.28$, $p=0.10$, $d=0.27$
		R: $W=931.5$, $p=0.25$	$W=1180.5$, $p=0.084$
Pop-typical, Permissible	Pop-atypical, Impermissible	$t(83.689)=2.81$, $p=0.0030$, $d=0.61$	$t(83.419)=1.52$, $p=0.066$, $d=0.33$
		$W=1249$, $p=0.0021$	$W=1075$, $p=0.091$
Pop-typical, Impermissible	Pop-atypical, Impermissible	$t(95.506)=0.057$, $p=0.48$, $d=0.011$	R: $t(90.938)=1.19$, $p=0.12$, $d=0.24$
		$W=1216.5$, $p=0.48$	R: $W=1000.5$, $p=0.066$
Agent-typical, Permissible	Agent-typical, Impermissible	$t(68.991)=4.78$, $p=4.8e^{-6}$, $d=1.00$	R: $t(83.949)=0.94$, $p=0.17$, $d=0.19$
		$W=2126.5$, $p=7.8e^{-6}$	R: $W=1322.5$, $p=0.24$
Agent-atypical, Permissible	Agent-atypical, Impermissible	$t(94.655)=1.83$, $p=0.035$, $d=0.37$	R: $t(86.412)=1.60$, $p=0.057$, $d=0.33$
		$W=1401.5$, $p=0.033$	R: $W=969$, $p=0.072$
Agent-typical, Permissible	Agent-atypical, Permissible	R: $t(80.953)=3.54$, $p=0.00033$, $d=0.77$	R: $t(79.429)=0.76$, $p=0.22$, $d=0.16$
		R: $W=525.5$, $p=0.00026$	R: $W=802$, $p=0.19$
Agent-typical, Permissible	Agent-atypical, Impermissible	R: $t(87.724)=1.77$, $p=0.040$, $d=0.37$	R: $t(91.076)=2.70$, $p=0.0041$, $d=0.55$
		R: $W=849.5$, $p=0.031$	R: $W=738.5$, $p=0.0035$
Agent-typical, Impermissible	Agent-atypical, Impermissible	R: $t(93.539)=7.28$, $p=0.5.0e^{-11}$, $d=1.38$	R: $t(101.85)=2.09$, $p=0.0.019$, $d=0.39$
		R: $W=690.5$, $p=7.3e^{-10}$	R: $W=1415.5$, $p=0.011$

Table 2: Tests of the varied agent and fixed agent effects for relevant comparisons for Pen Case results from Sytsma et al. (2012) where the fixed agent does not violate a norm. All tests one-tailed, reverse effects indicated by "R," significant results in red.

Our goal in Sytsma et al. (2012) was not to test the varied agent and fixed agent effects, however, and our studies were not designed with this in mind. One potential issue for this purpose is that in the conditions in which Professor Smith did not violate an injunctive norm, we did not state

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. As such, in Study 3 I replicated the eight versions of the Pen Case just considered, changing "able" to "allowed" in the relevant conditions.

3.1.1 Methods

Participants for Study 3 were given one of eight Pen Case vignettes following Sytsma et al. (2012). Full text of all vignettes is in the supplemental materials. The first four vignettes varied whether Professor Smith violated an injunctive norm and whether she violated a population-level statistical norm; the last four varied whether Professor Smith violated an injunctive norm and whether she 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 (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 the same 7-point scale used in the previous studies. Results were collected from 327 participants.[20]

3.1.2 Results

Results are shown in Figure 2. Looking at the ten pairs of comparisons noted above, the results were similar to those found by Sytsma et al. and are summarized in Table 3. While the varied agent effect

[20] 73.7% women (three non-binary), average age 38.3 years, ranging from 16 to 77. One-way ANOVAs showed no effect for gender on either ratings of the varied agent, $F(2, 324)=0.019$, $p=0.98$, $\eta^2=0$, or the fixed agent, $F(2, 324)=0.11$, $p=0.90$, $\eta^2=0.001$.

again occurred reliably when just the injunctive norm was varied (with significant effects in three of four comparisons and a borderline significant effect in the fourth), it was not found in two comparisons where just the population-level statistical norm was varied and *reverse* effects were found in the two comparisons where just the agent-level statistical norm was varied. Turning to the fixed agent effect, again it did not occur reliably when just the injunctive norm was varied (no significant effects in the four comparisons, although one comparison was borderline significant). Further no effect was found in the two comparisons where just the population-level statistical norm was varied, while a significant *reverse* effect was found in one of two comparisons where just the agent-level statistical norm was varied (and a borderline significant reverse effect in the other). The tests had a power ranging from 0.939 to 0.945 to detect an average fixed agent effect.

Norms for VA, Condition 1	Norms for VA, Condition 2	Varied Agent Effect	Fixed Agent Effect
Pop-typical, Permissible	Pop-typical, Impermissible	$t(74.957)=3.59$, $p=0.00030$, $d=0.80$	$t(77.958)=0.53$, $p=0.30$, $d=0.12$
		$W=1138.5$, $p=0.00044$	$W=861$, $p=0.28$
Pop-atypical, Permissible	Pop-atypical, Impermissible	$t(69.313)=4.65$, $p=7.8e^{-6}$, $d=1.04$	$t(77.094)=1.18$, $p=0.12$, $d=0.26$
		$W=1254.5$, $p=4.5e^{-5}$	$W=1000$, $p=0.066$
Pop-typical, Permissible	Pop-atypical, Permissible	**R:** $t(79.06)=0.37$, $p=0.36$, $d=0.08$	**R:** $t(79.78)=0.46$, $p=0.32$, $d=0.10$
		R: $W=812.5$, $p=0.40$	**R:** $W=785$, $p=0.30$
Pop-typical, Permissible	Pop-atypical, Impermissible	$t(71.851)=4.22$, $p=3.5e^{-5}$, $d=0.94$	$t(76.453)=0.75$, $p=0.23$, $d=0.17$
		$W=1179.5$, $p=9.7e^{-5}$	$W=913.5$, $p=0.13$
Pop-typical, Impermissible	Pop-atypical, Impermissible	$t(77.277)=0.78$, $p=0.22$, $d=0.17$	$t(76)=0.27$, $p=0.61$, $d=0.06$
		$W=887$, $p=0.20$	$W=862.5$, $p=0.73$
Agent-typical, Permissible	Agent-typical, Impermissible	$t(80.611)=1.42$, $p=0.079$, $d=0.31$	$t(79.846)=1.56$, $p=0.061.$, $d=0.34$
		$W=1009.5$, $p=0.085$	$W=1025$, $p=0.064$
Agent-atypical, Permissible	Agent-atypical, Impermissible	$t(75.75)=2.18$, $p=0.016$, $d=0.48$	$t(78.801)=0.48$, $p=0.32$, $d=0.11$
		$W=1046.5$, $p=0.025$	$W=877$, $p=0.36$
Agent-typical, Permissible	Agent-atypical, Permissible	**R:** $t(76.503)=2.66$, $p=0.0047$, $d=0.59$	**R:** $t(80.668)=1.57$, $p=0.060$, $d=0.35$
		R: $W=605$, $p=0.0084$	**R:** $W=695.5$, $p=0.063$
Agent-typical, Permissible	Agent-atypical, Impermissible	**R:** $t(78.997)=0.46$, $p=0.32$, $d=0.10$	**R:** $t(78.722)=1.03$, $p=0.15$, $d=0.23$
		R: $W=771$, $p=0.32$	**R:** $W=709.5$, $p=0.15$
Agent-typical, Impermissible	Agent-atypical, Impermissible	**R:** $t(79.685)=1.91$, $p=0.030$, $d=0.42$	**R:** $t(77.52)=2.6$, $p=0.0056$, $d=0.58$
		R: $W=639.5$, $p=0.030$	**R:** $W=576$, $p=0.0065$

Table 3: Tests of the varied agent and fixed agent effects for Study 3. All tests one-tailed, reverse effects indicated by "R," significant results in red.

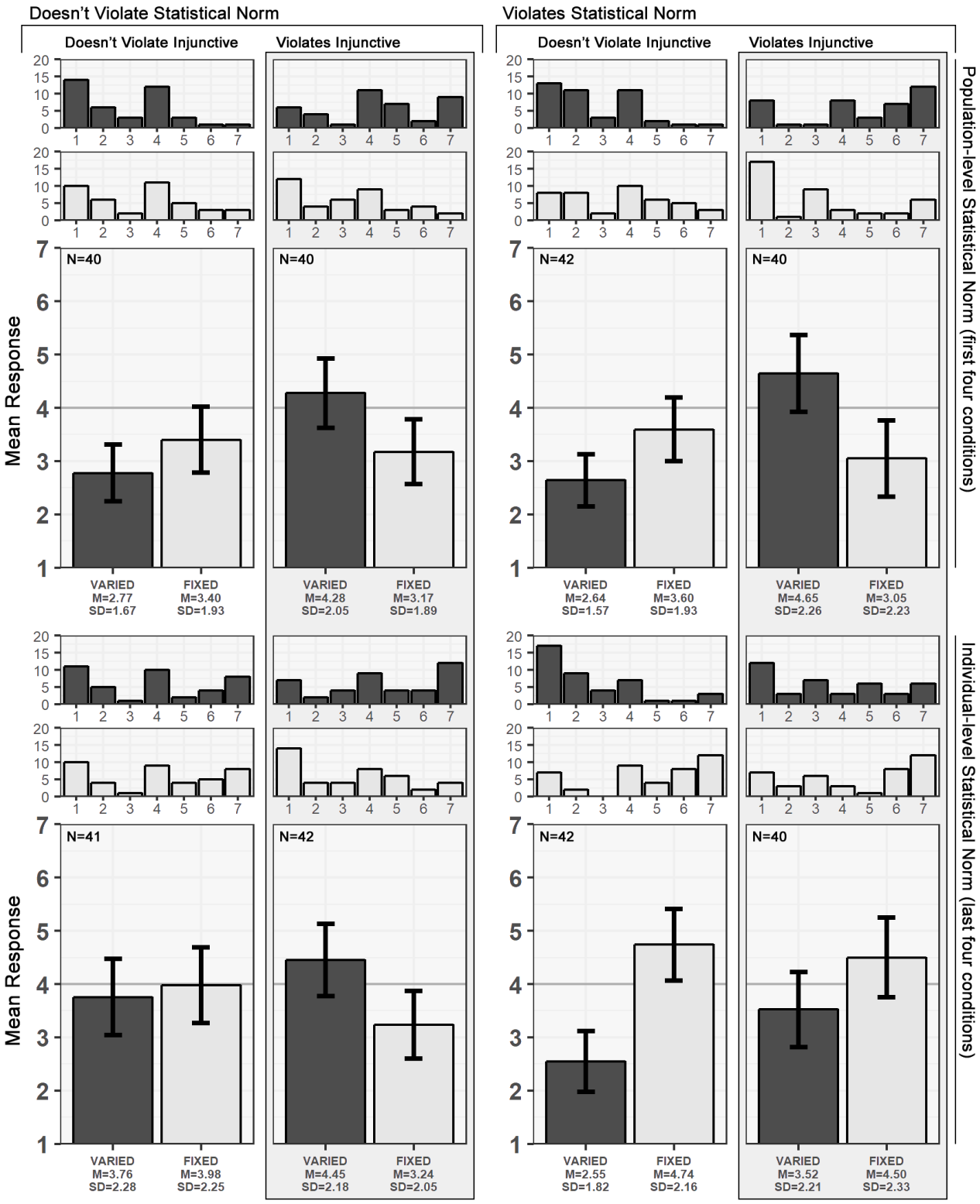

Figure 2: Results of Study 3 with histograms above the plot of the means for each condition and showing 95% confidence intervals.

3.2 Computer Case

Sytsma et al. (2012) opened with a description of another type of conjunctive case—the Computer Case—based on a thought experiment given by Knobe (2006).[21] 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 at the same time 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, participants were significantly more likely to affirm that Jane caused the system to crash than Lauren, illustrating the cross-agent effect.

In the full set of results reported in Livengood et al. (2017), we also tested variations in which there was no injunctive norm in place, allowing for further tests of the varied agent and fixed agent effects. We tested seven relevant pairs of cases—the initial scenario with each participant asked about both Lauren and Jane (Studies 1 and 9) or about just one of the two agents (Study 2 and combining Studies 10 and 14); a variation in which participants were also asked about the mainframe (Studies 3 and 13); variations in which participants were asked whether the agents' *actions* caused the outcome, with each participant either being asked about both actions (Studies 7 and 11) or just one of the two actions (Studies 8 and 12); and variations in which the issue with the mainframe is described as a feature rather than a bug, again with each participant either being asked about both agents (Studies 5 and 15) or just one of the two agents (Studies 6 and 16). This gives seven comparisons where we would expect to find a varied agent effect and seven comparisons where the counterfactual view predicts we should find a fixed agent effect. We did not test these comparisons in the original article, but I have reanalyzed the data. The results are summarized in

[21] 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 and Icard et al.'s first experiment.

Table 4. A significant varied agent effect was found in each of the seven comparisons, while a significant fixed agent effect was found in four of the seven (with a fifth being borderline significant). The three tests that did not show a significant effect had powers ranging from 0.948 to 0.996 to detect an average fixed agent effect.

Again, our goal in testing the Computer Case was not to test the varied agent and fixed agent effects, and our studies were not designed with that in mind. The same potential issue noted for the Pen Case applies to these studies: in the cases where an injunctive norm was not violated, we did not specify that either agent was specifically allowed to log into the mainframe. As such, in Study 4 I replicated two of the seven comparisons from Livengood et al.—the initial scenarios and the scenarios where the instability was described as a feature rather than a bug, asking each participant about both agents—adding a line to the non-normed vignettes making clear that employees were allowed to log into any of the terminals.

3.2.1 Methods

Participants for Study 4 were given one of four vignettes featuring a varied agent (Jane) and a fixed agent (Lauren) logging into a mainframe, as described above. 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 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. Results were collected from 262 participants.[22]

[22] 59.2% women (six non-binary), average age 26.5 years, ranging from 16 to 99. One-way ANOVAs showed no effect for gender on ratings of the fixed agent, $F(2, 259)=0.28$, $p=0.76$, $\eta^2=0.002$, but did show an effect for the varied agent,

3.2.2 Results

Results are shown in Figure 3 and tests are summarized in Table 4. Replicating the results from Livengood et al. (2017), in the first pair of conditions (mainframe unstable) there was a significant varied agent effect but not a fixed agent effect, despite the test having a power of 0.989 to detect an average fixed agent effect; and, in the second pair of conditions (single user) there was both a significant varied agent effect and a significant fixed agent effect.

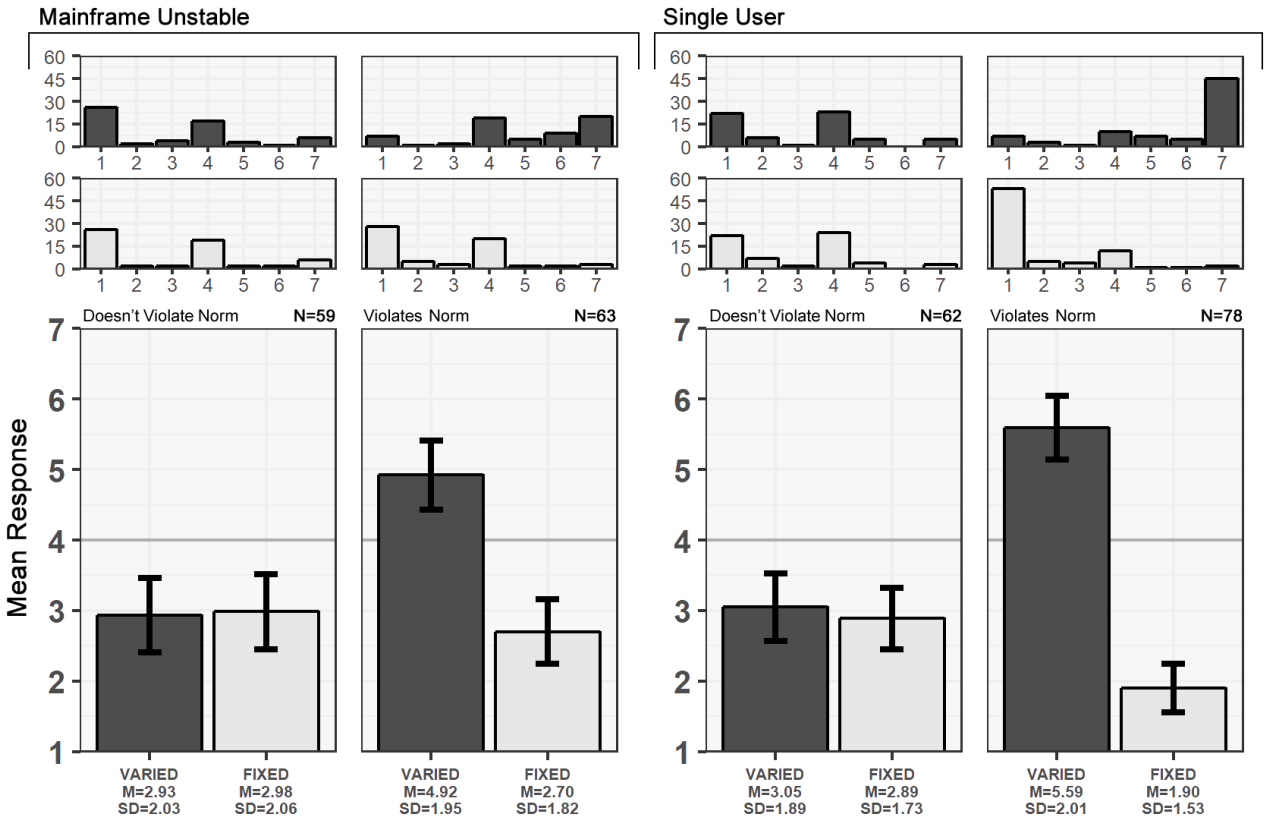

Figure 3: Results of Study 4 with histograms above the plot of the means for each condition and showing 95% confidence intervals.

$F(2, 259)=4.97$, $p=0.0077$, $\eta^2=0.037$. Looking closer, there were notable differences with regard to the fixed agent effect. For the first pair of conditions (mainframe unstable), there was a non-significant reverse effect for women, $t(74.345)=0.52$, $p=0.30$, $d=0.12$ ($W=659.5$, $p=0.19$), and a significant effect for men, $t(38.519)=2.07$, $p=0.023$, $d=0.64$ ($W=310.5$, $p=0.016$). For the second pair of conditions (single user), for women the t-test was borderline significant, $t(74.36)=1.47$, $p=0.073$, $d=0.33$, and the Wilcoxon rank sum test was significant, $W=939.5$, $p=0.018$, while both tests were significant for men, $t(37.561)=3.91$, $p=0.00019$, $d=1.09$ ($W=618$, $p=0.00024$).

Condition 1 (no violation)	Condition 2 (norm violation)	Varied Agent Effect	Fixed Agent Effect
Livengood et al. Study 9	Livengood et al. Study 1	$t(140.98)=6.91$, $p=7.8e^{-11}$, $d=1.16$	$t(140.46)=0.82$, $p=0.21$, $d=0.14$
		$W=3961.5$, $p=2.1e^{-9}$	$W=2701$, $p=0.26$
Livengood et al. Study 10+14	Livengood et al. Study 2	$t(49.509)=6.30$, $p=3.9e^{-8}$, $d=1.44$	R: $t(56.815)=0.77$, $p=0.22$, $d=0.16$
		$W=2426$, $p=1.2e^{-8}$	R: $W=1215$, $p=0.19$
Livengood et al. Study 13	Livengood et al. Study 3	$t(98.766)=6.29$, $p=4.3e^{-9}$, $d=1.19$	$t(103.52)=4.22$, $p=2.6e^{-5}$, $d=0.76$
		$W=2656.5$, $p=9.3e^{-9}$	$W=2238.5$, $p=0.00018$
Livengood et al. Study 11	Livengood et al. Study 7	$t(85.157)=4.83$, $p=2.9e^{-6}$, $d=1.02$	$t(84.83)=1.75$, $p=0.042$, $d=0.37$
		$W=1545$, $p=9.9e^{-6}$	$W=1242.5$, $p=0.037$
Livengood et al. Study 12	Livengood et al. Study 8	$t(67.721)=4.20$, $p=4.0e^{-5}$, $d=0.98$	$t(81.583)=1.42$, $p=0.080$, $d=0.31$
		$W=1046.5$, $p=7.5e^{-5}$	$W=1041$, $p=0.066$
Livengood et al. Study 15	Livengood et al. Study 5	$t(111.78)=5.55$, $p=9.9e^{-8}$, $d=1.02$	$t(115.8)=2.95$, $p=0.0019$, $d=0.53$
		$W=2602.5$, $p=3.7e^{-7}$	$W=2171.5$, $p=0.0039$
Livengood et al. Study 16	Livengood et al. Study 6	$t(63.956)=4.82$, $p=4.56e^{-6}$, $d=1.16$	$t(64.7)=3.11$, $p=0.0014$, $d=0.70$
		$W=904$, $p=1.6e^{-5}$	$W=955.5$, $p=0.0032$
Replication, Unstable	Replication, Unstable	$t(118.67)=5.50$, $p=1.1e^{-7}$, $d=1.00$	$t(115.79)=0.81$, $p=0.21$, $d=0.15$
		$W=2792.5$, $p=4.3e^{-7}$	$W=1969$, $p=0.27$
Replication, Single User	Replication, Single User	$t(134.2)=7.69$, $p=1.4e^{-12}$, $d=1.30$	$t(122.67)=3.54$, $p=0.00028$, $d=0.61$
		$W=3929.5$, $p=2.6e^{-11}$	$W=3252.5$, $p=6.1e^{-5}$

Table 4: Tests of the varied agent and fixed agent effects for relevant comparisons for the Computer Case results from Livengood et al. (2017) and Study 4. All tests one-tailed, reverse effects indicated by "R," significant results in red.

3.3 Discussion

Whether looking at Sytsma et al.'s (2012) original results for the Pen Case or the replication, there is consistent evidence of the varied agent effect for the injunctive norm, while the results are mixed for the fixed agent effect. In line with the responsibility view and against the counterfactual view, neither effect is found for population-level statistical norms while the effects tend to be reversed for agent-level statistical norms. And despite finding the expected cross-agent and varied agent effects across seven pairs of probes for the Computer Case in Livengood et al. (2017), a significant fixed agent effect was only found in four of the pairs. Two pairs of conditions were replicated in Study 4, and showed comparable results—there was a significant varied agent effect in each pair, but a significant fixed agent effect was only found in one of the two pairs.

Overall, a significant fixed agent effect was only found in six out of 17 comparisons where an injunctive norm was varied. Both the counterfactual view and context effect hypothesis are able to explain the effects when they occurred, given that a significant varied agent effect was also found in each of these cases. But neither readily explains why the fixed agent effect was only found in a minority of the comparisons. Further work is needed here. Nonetheless, insofar as the counterfactual view directly predicts the occurrence of a fixed agent effect, while the alternative views only do so only indirectly, rather than supporting the counterfactual view over the alternatives, if anything the current results suggest the reverse.

4. Assessing the Pattern of Effects: Disjunctive Cases

We have seen that the overall pattern of effects across both conjunctive and disjunctive cases is important for assessing the explanations of the impact of norms on causal attributions that have been put forward. In the previous section, however, we saw that part of the supposed pattern doesn't occur as systematically as predicted by the counterfactual view. In this section, I extend the investigation to disjunctive cases. I begin by replicating both the conjunctive and disjunctive versions of Kominsky et al.'s Motion Detector Case but asking participants about both the varied agent and the fixed agent; I then extend this to test responsibility attributions and blame attributions. Finally, I test a disjunctive version of Kominsky et al.'s Email Case.

4.1 Motion Detector Case

In Kominsky et al.'s third experiment they tested the fixed agent effect for both conjunctive and disjunctive versions of the same case. In these scenarios, a varied agent (Billy) and a fixed agent (Suzy) work on a project that is important for national security. In the non-normed conditions, both are supposed to arrive at work at 9am; in the normed conditions, Billy is prohibited from arriving in

the morning. In each case, both arrive at 9am. It turns out that there is a motion detector installed in the room they arrive at, which will either go off if *more than one person* arrives (conjunctive) or if *at least one person* arrives (disjunctive). Kominsky et al. gave participants one of the resulting four vignettes and asked them to assess the statement "Suzy caused the motion detector to go off." As they predicted, there was a significant fixed agent effect in the conjunctive conditions, but no effect in the disjunctive conditions. The same scenarios were also used in Icard et al.'s first experiment, but this time soliciting causal judgments about just the varied agent. As they predicted, there was a significant varied agent effect in the conjunctive conditions, and a significant *reverse* varied agent effect in the disjunctive conditions.

This pair of results poses an interesting challenge for alternative accounts of the effect of norms on causal attributions. First, while the context effect hypothesis can explain the occurrence of a fixed agent effect when there is a varied agent effect, it does not explain why we would see a reverse varied agent effect without a corresponding reverse fixed agent effect. That said, as seen in the previous section, the fixed agent effect is not reliably found in conjunctive cases, including in cases where both the counterfactual view and the context effect hypothesis would expect it to occur. As such, the absence of the fixed agent effect in a single disjunctive case does not strongly support the counterfactual view over competitors. Second, it has been suggested that the alternative accounts should make the same predictions about conjunctive and disjunctive versions of the same cases, with the result that the occurrence of the varied agent effect in the conjunctive conditions and the *reverse* varied agent effect in the disjunctive conditions looks quite problematic. As noted in Section 1, however, the alternative accounts allow that causal structure will matter for causal attributions, and based on previous work at minimum views focusing on responsibility for an outcome predict that attributions for the varied agent will tend to be lower when she violates the norm in a disjunctive case compared to when she does so in a matching conjunctive case. And such

an effect appears to be present across Kominsky et al. and Icard et al.'s studies, with the mean response for Billy in the normed conjunctive condition (6.00) being notably higher than in the normed disjunctive condition (3.24).[23]

To address the challenge posed by results for the Motion Detector Case, in my fifth study I replicated then extended these studies. I had three goals. First, given the variability seen for the fixed agent effect in conjunctive cases, I wanted to further test that there is reliably no effect in disjunctive cases. Second, to investigate whether the results are consistent with the context effect hypothesis, I wanted to solicit ratings for both the varied agent and the fixed agent. Finally, to assess alternative accounts for disjunctive cases, I wanted to solicit further relevant attributions.

4.1.1 Methods

Participants in Study 5 were given one of the four vignettes from Kominsky et al.'s third experiment, varying the causal structure (conjunctive, disjunctive) and whether the varied agent violated an injunctive norm. Participants were asked to rate one of three pairs of causal statements with the order of the two questions randomized—either a pair of causal attributions ("Billy caused the motion detector to go off," "Suzy caused the motion detector to go off"), a pair of responsibility attributions ("Billy is responsible for the motion detector going off," "Suzy is responsible for the motion detector going off"), or a pair of blame attributions ("Billy is to blame for the motion detector going off," "Suzy is to blame for the motion detector going off"). Participants assessed each statement using the same 7-point scale used in the previous studies. Each participant was then given the comprehension check used by Kominsky et al., being asked "Who was supposed to show

[23] Thanks to Jonathan Kominsky for supplying the means for these cases.

up at 9am?" Results were collected from 507 participants, excluding 23 participants who failed the comprehension check.[24]

4.1.2 Results

Results are shown in Figure 4. An ANOVA looking at ratings for the varied agent with *norm* (neutral, normed), *structure* (conjunctive, disjunctive), and *term* (cause, responsibility, blame) as between-participant factors showed main effects for norm, $F(1, 495)=66.34$, $p=3.1e^{-15}$, $\eta^2=0.11$, and structure, $F(1, 495)=5.69$, $p=0.017$, $\eta^2=0.010$, but not for term, $F(2, 495)=1.08$, $p=0.34$, $\eta^2=0.004$. In addition, there was an interaction effect for norm and structure, $F(1, 495)=17.74$, $p=3.0e^{-5}$, $\eta^2=0.030$. A matching ANOVA looking at ratings for the fixed agent showed the same pattern of effects: main effects for norm, $F(1, 495)=18.63$, $p=1.9e^{-5}$, $\eta^2=0.034$, and structure, $F(1, 495)=19.16$, $p=1.5e^{-5}$, $\eta^2=0.035$, but not for term, $F(2, 495)=2.08$, $p=0.13$, $\eta^2=0.008$, and an interaction effect for norm and structure, $F(1, 495)=4.24$, $p=0.040$, $\eta^2=0.008$.

While the effects seen for norm and structure are potentially consistent with the pattern predicted by the counterfactual view, Figure 4 paints a different picture. Planned comparisons testing each of the three predicted effects for causal attributions in conjunctive and disjunctive cases reveal that while the predictions are born out for the conjunctive conditions (there are significant cross-agent, varied agent, and fixed agent effects), the results run counter to two of the three predictions for the disjunctive conditions: rather than finding a *reverse* cross-agent effect, there is a borderline significant cross-agent effect, and rather than finding a *reverse* varied agent effect, the mean is *higher* when the varied agent violates the norm, although the difference is not significant; as predicted, however, no effect was found for the fixed agent (despite having a power of 0.941 to

[24] 71.6% women (two non-binary), average age 27.9 years, ranging from 16 to 99. One-way ANOVAs showed no effect for gender on either ratings of the varied agent, $F(2, 504)=2.29$, $p=0.10$, $\eta^2=0.009$, or the fixed agent, $F(2, 504)=0.12$, $p=0.88$, $\eta^2=0$.

find an average fixed agent effect). Results were similar for responsibility attributions (although here the varied agent effect was significant for the disjunctive conditions) and blame attributions (with borderline significant varied agent and fixed agent effects in the disjunctive conditions), as summarized in Table 5.

Finally, against the contention that alternative accounts are unable to explain the differences in results for conjunctive and disjunctive cases, there were no effects for term: whether participants assessed causal attributions, responsibility attributions, or blame attributions did not show a significant effect. And while the causal structure did affect ratings, it did so similarly for each type of attribution. As such, these results are consistent with each of the alternative accounts discussed in Section 1. Further, in line with the prediction derived from previous work on responsibility attributions, in the normed conditions causal ratings were significantly higher for the varied agent in the conjunctive scenario than in the disjunctive scenario, $t(74.693)=2.38$, $p=0.0099$, $d=0.53$ ($W=1010$, $p=0.016$), and similarly for responsibility ratings, $t(78.927)=2.37$, $p=0.010$, $d=0.51$ ($W=1269$, $p=0.013$), and blame ratings, $t(81.822)=2.70$, $p=0.0042$, $d=0.58$ ($W=1214$, $p=0.0082$).

Term	Structure	Cross-agent Effect	Varied Agent Effect	Fixed Agent Effect
Cause	Conjunctive	$t(39)=7.47$, $p=2.5e^{-9}$, $d=1.18$	$t(77.931)=5.15$, $p=9.4e^{-7}$, $d=1.15$	$t(74.904)=3.28$, $p=7.9e^{-4}$, $d=0.73$
		$V=416$, $p=3.5e^{-6}$	$W=1262$, $p=2.1e^{-6}$	$W=1115.5$, $p=0.00055$
	Disjunctive	$t(39)=1.39$, $p=0.086$, $d=0.22$	$t(72.514)=0.94$, $p=0.18$, $d=0.21$	$t(72.687)=0.84$, $p=0.20$, $d=0.19$
		$V=160.5$, $p=0.056$	$W=940$, $p=0.12$	$W=912.5$, $p=0.18$
Resp.	Conjunctive	$t(47)=6.32$, $p=4.4e^{-8}$, $d=0.91$	$t(91.611)=5.7$, $p=7.2e^{-8}$, $d=1.18$	$t(91.985)=2.30$, $p=0.012$, $d=0.47$
		$V=704.5$, $p=3.43e^{-6}$	$W=1748$, $p=2.2e^{-7}$	$W=1413$, $p=0.0065$
	Disjunctive	$t(41)=1.74$, $p=0.045$, $d=0.27$	$t(74.322)=2.11$, $p=0.019$, $d=0.46$	$t(74.339)=0.30$, $p=0.38$, $d=0.06$
		$V=198$, $p=0.029$	$W=1064.5$, $p=0.027$	$W=917.5$, $p=0.30$
Blame	Conjunctive	$t(42)=7.83$, $p=4.9e^{-10}$, $d=1.19$	$t(80.83)=5.15$, $p=8.9e^{-7}$, $d=1.13$	$t(74.427)=2.90$, $p=0.0025$, $d=0.64$
		$V=612$, $p=3.0e^{-7}$	$W=1354$, $p=1.6e^{-6}$	$W=1142.5$, $p=0.0026$
	Disjunctive	$t(43)=2.25$, $p=0.015$, $d=0.34$	$t(80.095)=1.51$, $p=0.067$, $d=0.32$	$t(83.488)=1.31$, $p=0.096$, $d=0.28$
		$V=276$, $p=0.017$	$W=1073.5$, $p=0.091$	$W=1089$, $p=0.065$

Table 5: Tests of the cross-agent, varied agent, and fixed agent effects for Study 5. All tests one-tailed, reverse effects indicated by "R," significant results in red.

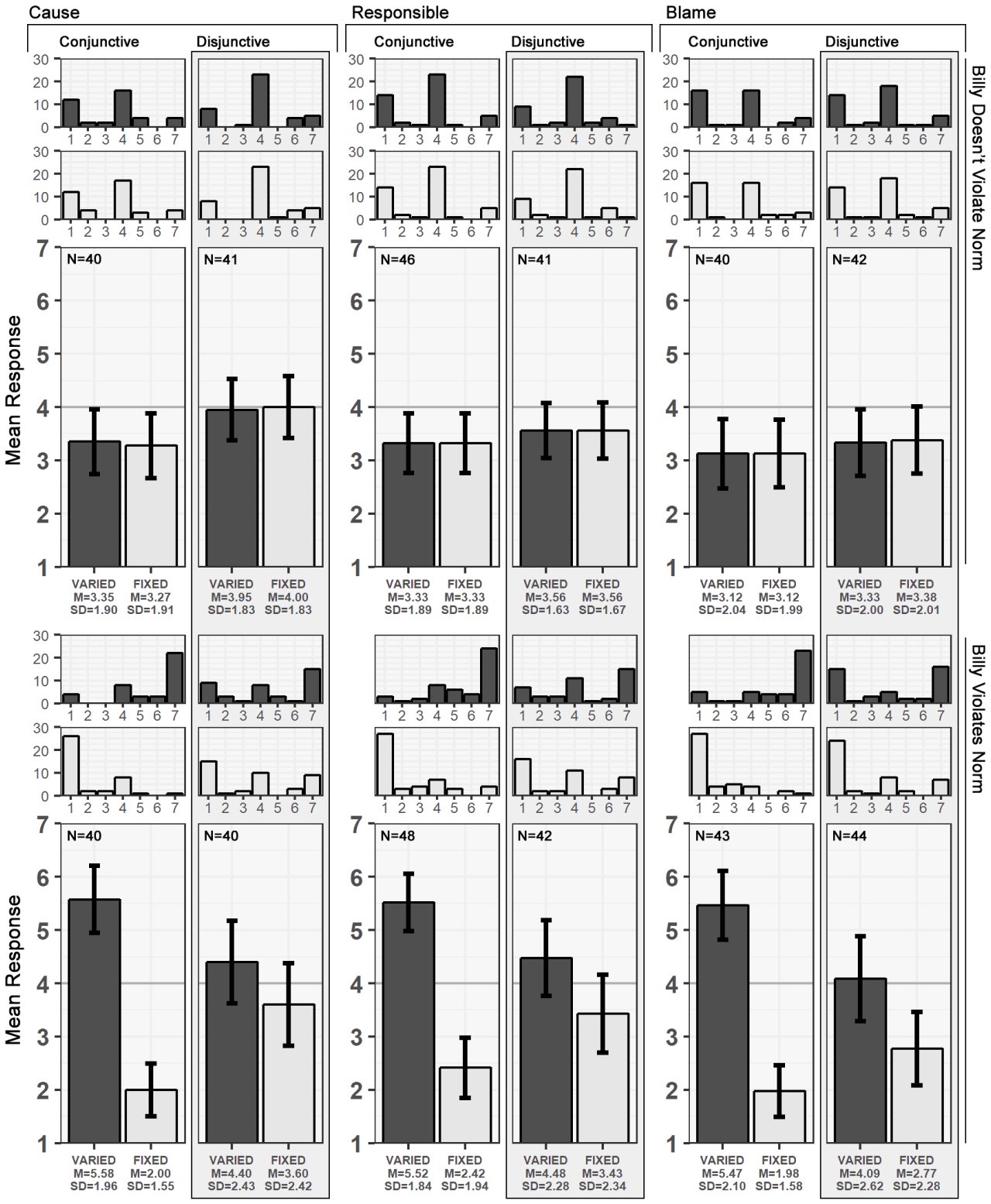

Figure 4: Results for Study 5 with histograms above the plots of the means for each condition and showing 95% confidence intervals.

4.2 Motion Detector Case, Replication

Against the contention that the alternative accounts cannot explain the pattern of effects found across conjunctive and disjunctive cases, the results of Study 5 showed no effect for type of attribution. And against the predictions of the counterfactual view, the reverse effects for disjunctive cases were not seen. Given that the causal ratings for the varied agent in the disjunctive conditions run counter to those found by Icard et al. (2017) and Kominsky and Phillips (2019), some care is needed here, however. As such, in Study 6 I directly replicated the cause and responsibility conditions from the previous study with larger sample sizes (roughly N=200 per condition).

4.2.1 Methods

Methods were identical to Study 5, except that the two conditions soliciting blame attributions were excluded. Results for Study 6 were collected from 1628 participants, excluding 81 participants who failed the comprehension check.[25]

4.2.2 Results

Results are shown in Figure 5. As in Study 5, an ANOVA looking at ratings for the varied agent with *norm* (neutral, normed), *structure* (conjunctive, disjunctive), and *term* (cause, responsibility) as between-participant factors showed main effects for norm, $F(1, 1619)=135.01$, $p<2.2e^{-16}$, $\eta^2=0.069$, and structure, $F(1, 1619)=69.30$, $p<2.2e^{-16}$, $\eta^2=0.036$, but not for term, $F(1, 1619)=0.55$, $p=0.46$, $\eta^2=0.000$. There was, however, a significant interaction effect between term and structure, $F(1, 1619)=6.20$, $p=0.013$, $\eta^2=0.003$, in addition to the interaction effect seen in Study 5 between norm and structure, $F(1, 1619)=115.77$, $p<2.2e^{-16}$, $\eta^2=0.059$. Interestingly, a matching ANOVA looking at

[25] 74.3% women (17 non-binary), average age 32.4 years, ranging from 16 to 86. One-way ANOVAs showed no effect for gender on either ratings of the varied agent, $F(2, 1624)=0.49$, $p=0.61$, $\eta^2=0.001$, or the fixed agent, $F(2, 1624)=0.70$, $p=0.50$, $\eta^2=0.001$.

ratings for the fixed agent showed a main effect for term, $F(1, 1619)=8.97$, $p=0.0028$, $\eta^2=0.005$, in addition to the effects seen in Study 5—main effects for norm, $F(1, 1619)=44.59$, $p=3.3e^{-11}$, $\eta^2=0.025$, and structure, $F(1, 1619)=89.63$, $p<2.2e^{-16}$, $\eta^2=0.050$, and an interaction effect for norm and structure, $F(1, 1619)=21.89$, $p=3.1e^{-6}$, $\eta^2=0.012$.

While the results show a difference between the causal ratings and the responsibility ratings, the effect sizes for term were negligible. This is consistent with the responsibility view, which takes causal attributions and responsibility attributions to have a family resemblance, not that the two terms are pure synonyms. More importantly, as summarized in Table 6, we see the same overall pattern of effects for both causal ratings and responsibility ratings. As in Study 5, planned comparisons found the three predicted effects for causal ratings in the conjunctive conditions (there are significant cross-agent, varied agent, and fixed agent effects), and the same effects are also found for responsibility ratings. For the disjunctive conditions, however, for causal ratings the results again run counter to two of the three predictions made by the counterfactual view. No significant effects were seen here, despite the large sample sizes. And the same held for responsibility ratings. Finally, in line with the prediction derived from work on responsibility attributions, in the normed conditions causal ratings were significantly higher for the varied agent in the conjunctive scenario than in the disjunctive scenario, $t(396.04)=9.87$, $p<2.2e^{-16}$, $d=0.98$ ($W=31458$, $p<2.2e^{-16}$), and similarly for responsibility ratings, $t(365.87)=8.72$, $p<2.2e^{-16}$, $d=0.89$ ($W=26125$, $p=3.7e^{-16}$).

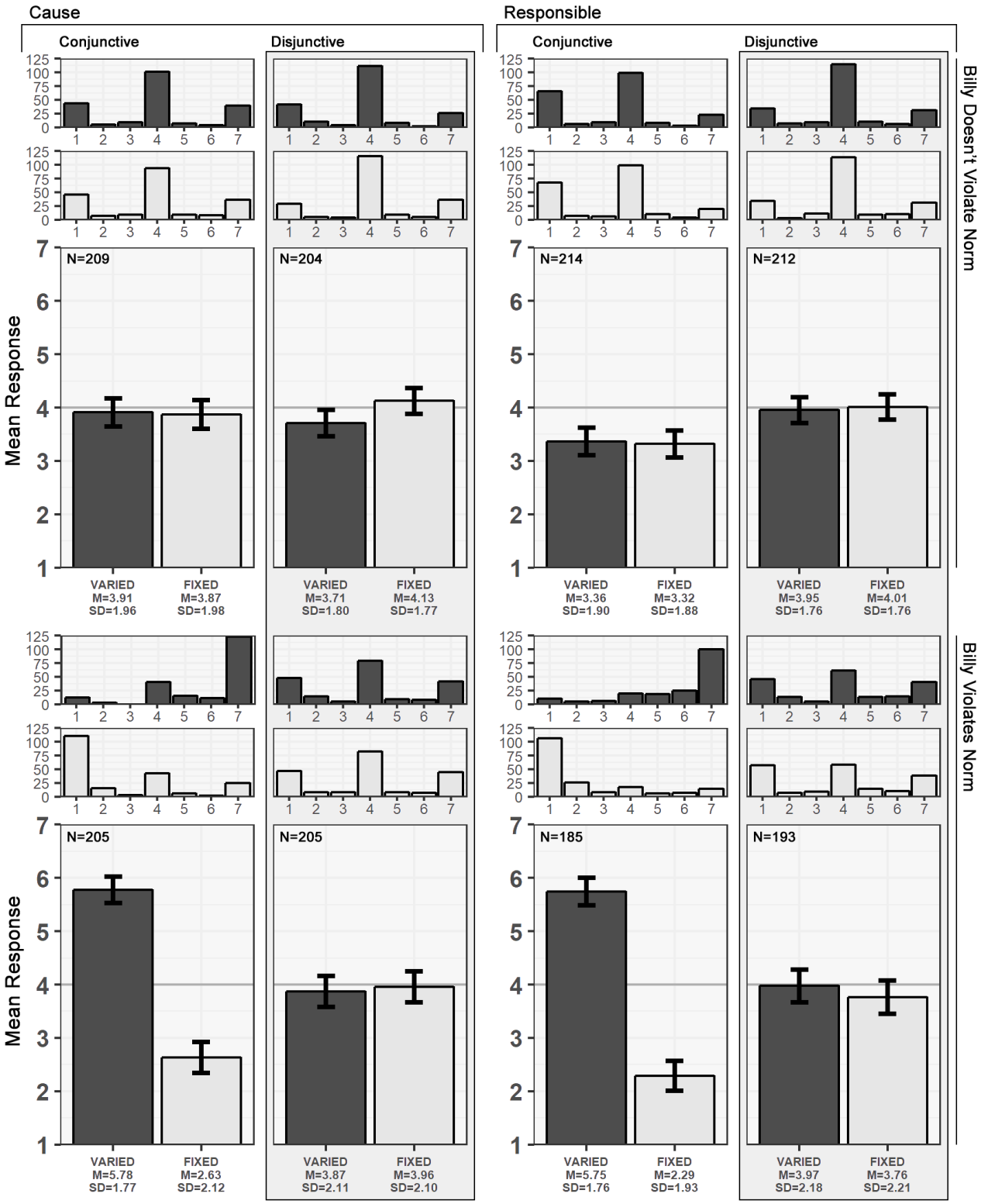

Figure 5: Results for Study 6 with histograms above the plots of the means for each condition and showing 95% confidence intervals.

Term	Structure	Cross-agent Effect	Varied Agent Effect	Fixed Agent Effect
Cause	Conjunctive	$t(204)=14.6$, $p<2.2e^{-16}$, $d=1.02$	$t(409.33)=10.2$, $p<2.2e^{-16}$, $d=1.00$	$t(408.77)=6.14$, $p=9.8e^{-10}$, $d=0.60$
		$V=8406.5$, $p<2.2e^{-16}$	$W=32446$, $p<2.2e^{-16}$	$W=28614$, $p=2.7e^{-10}$
	Disjunctive	R: $t(204)=0.39$, $p=0.35$, $d=0.03$	$t(397.39)=0.86$, $p=0.19$, $d=0.09$	$t(395.87)=0.87$, $p=0.19$, $d=0.09$
		R: $V=1779$, $p=0.34$	$W=21770$, $p=0.22$	$W=21930$, $p=0.18$
Resp.	Conjunctive	$t(184)=15.3$, $p<2.2e^{-16}$, $d=1.13$	$t(394.99)=13.0$, $p<2.2e^{-16}$, $d=1.30$	$t(385.44)=5.36$, $p=7.3e^{-8}$, $d=0.54$
		$V=10853$, $p<2.2e^{-16}$	$W=32228$, $p<2.2e^{-16}$	$W=25844$, $p=1.2e^{-8}$
	Disjunctive	$t(192)=0.86$, $p=0.20$, $d=0.06$	$t(369.84)=0.11$, $p=0.46$, $d=0.01$	$t(367.32)=1.26$, $p=0.10$, $d=0.13$
		$V=2436$, $p=0.22$	$W=20813$, $p=0.38$	$W=21668$, $p=0.13$

Table 6: Tests of the cross-agent, varied agent, and fixed agent effects for Study 6. All tests one-tailed, reverse effects indicated by "R," significant results in red.

4.3 Disjunctive Email Case

In both Study 5 and Study 6 tests failed to find two of the three effects for causal attributions in disjunctive scenarios predicted by the counterfactual view: against the predictions the reverse cross-agent effect and reverse varied agent effect were absent, but in line with the predictions the fixed agent effect was also absent. The absence of the fixed agent effect is also consistent with the context effect hypothesis, however, given the lack of a varied agent effect. To further investigate the effects for disjunctive cases, in Study 7 I tested a disjunctive version of the Email Case from Kominsky et al.'s second experiment. Subsequent to running this study, I found that in Icard et al.'s first experiment they tested causal ratings for the varied agent in a similar disjunctive version of this case, with slightly different wording, finding the predicted reverse varied agent effect.

4.3.1 Methods

Participants were given one or the other of two disjunctive 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. And 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. Results were collected from 85 participants.[26]

4.3.2 Results

Results are shown in Figure 6. I tested each of the three effects identified in Section 1, as summarized in Table 7. None of the three predictions made by the counterfactual view were borne out: against the prediction that there would be a *reverse* cross-agent effect, there was a significant effect in the opposite direction; against the prediction that there would be a *reverse* varied agent effect, there was again a significant effect in the opposite direction; and, against the prediction that there would be no fixed agent effect, there was a significant difference.

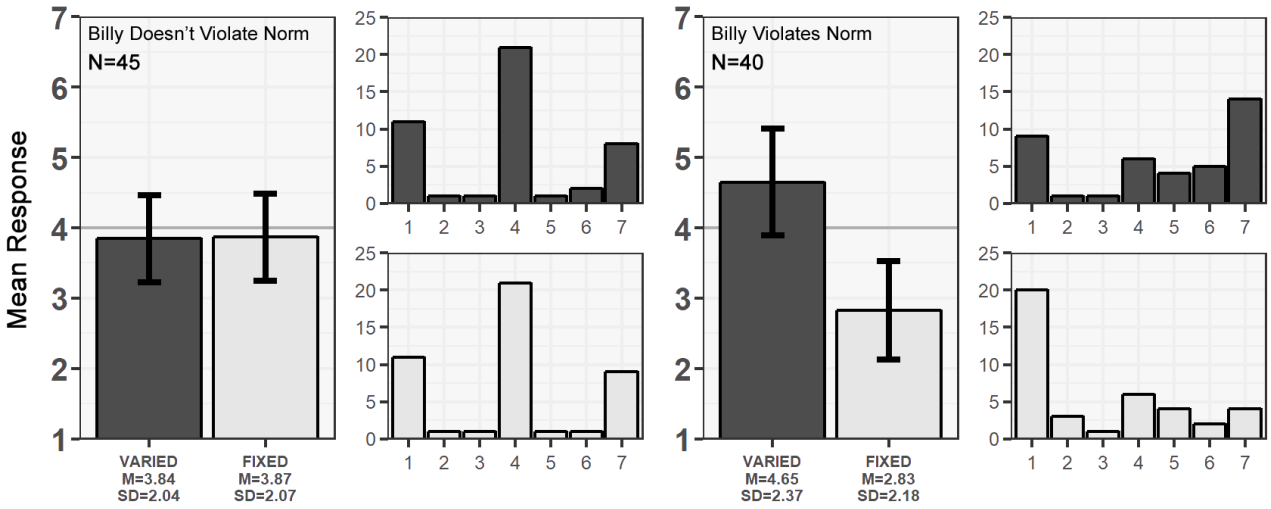

Figure 6: Results for Study 7 with histograms to the right of the plot of the means for each condition and showing 95% confidence intervals.

[26] 55.3% women, average age 30.0 years, ranging from 16 to 83. One-way ANOVAs showed no effect for gender on either ratings of the varied agent, $F(1, 83)=0.022$, $p=0.88$, $\eta^2=0$, or the fixed agent, $F(1, 83)=1.38$, $p=0.24$, $\eta^2=0.016$.

4.4 Disjunctive Email Case, Replication

Given that the finding of a varied agent effect in Study 6 runs counter to the results reported by Icard et al. (2017) and Kominsky and Phillips (2019), some care is again needed here. As such, in Study 8 I directly replicated the study with larger sample sizes (roughly N=200 per condition).

4.4.1 Methods

Methods were identical to Study 7. Results were collected from 393 participants.[27]

4.4.2 Results

Results are shown in Figure 7. As in the previous study, I tested each of the three effects and the results are summarized in Table 7. Results were similar, showing a significant cross-agent, varied agent, and fixed agent effect. Once again, this runs counter to the predicted pattern of effects given by the counterfactual view.

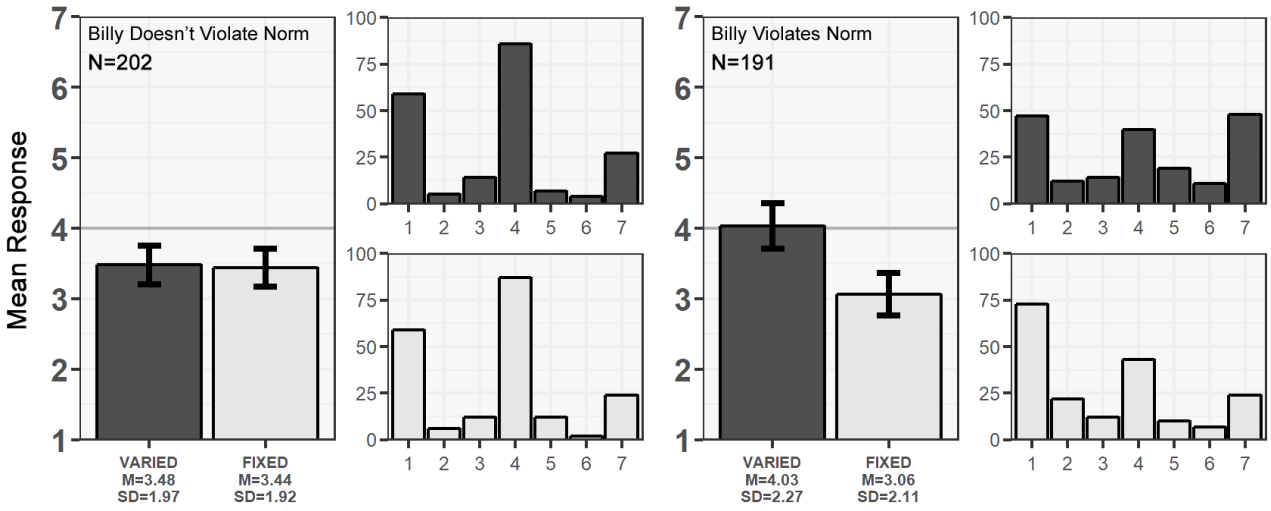

Figure 7: Results for Study 8 with histograms to the right of the plot of the means for each condition and showing 95% confidence intervals.

[27] 77.4% women (four non-binary), average age 29.4 years, ranging from 16 to 100. One-way ANOVAs showed no effect for gender on either ratings of the varied agent, $F(2, 390)=0.62$, $p=0.54$, $\eta^2=0.003$, or the fixed agent, $F(2, 390)=0.98$, $p=0.38$, $\eta^2=0.005$.

Study	Cross-agent Effect	Varied Agent Effect	Fixed Agent Effect
7	$t(39)=3.83$, $p=0.00023$, $d=0.61$	$t(77.57)=1.67$, $p=0.050$, $d=0.37$	$t(80.664)=1.25$, $p=0.014$, $d=0.49$
	$V=206.5$, $p=0.00068$	$W=1106.5$, $p=0.030$	$W=1131$, $p=0.017$
8	$t(190)=4.58$, $p=4.1e^{-6}$, $d=0.33$	$t(376.11)=2.57$, $p=0.0053$, $d=0.26$	$t(382.49)=1.86$, $p=0.032$, $d=0.19$
	$V=3524$, $p=5.2e^{-6}$	$W=22092$, $p=0.0051$	$W=21532$, $p=0.019$

Table 7: Tests of the cross-agent, varied agent, and fixed agent effects for Studies 7 and 8. All tests one-tailed, reverse effects indicated by "R," significant results in red.

5. Conclusion

A number of competing explanations of the effect of norms on ordinary causal attributions have been put forward in the literature. In a recent pair of papers, Kominsky et al. (2015) and Icard et al. (2017) have offered strong support for one of these accounts—the counterfactual view—making several novel predictions about the overall pattern of effects found when expanding on the initial type of scenarios tested in the literature by including non-normative contrast cases and varying the causal structure. Insofar as the predicted pattern of effects holds, and insofar as the counterfactual view is able to explain each of these effects while alternative accounts (such as our responsibility view) cannot, this is strong evidence in favor of the counterfactual view. We have seen that there are difficulties with both of these assumptions, however.

In this paper I have presented evidence casting doubt on the proposed pattern of effects. First, for conjunctive cases while the predicted cross-agent and varied agent effects reliably occur for injunctive norms in the cases tested, the fixed agent effect does not, only being found in a minority of comparisons for the Pen Case and Computer Case. Second, while the counterfactual view predicts that these three effects should occur for both injunctive and descriptive norms, in conjunctive versions of the Pen Case they were not found for one type of descriptive norm (population-level statistical norm) and tended to be *reversed* for a second type of descriptive norm (individual-level statistical norm). Third, for disjunctive versions of the Motion Detector Case,

across two studies only one of the three predictions made by the counterfactual view were borne out: against the prediction that there would be a reverse cross-agent effect, no effect was found, and against the prediction that there would be a reverse varied-agent effect, again no effect was found; but, in line with the prediction that the fixed agent effect would be absent, no effect was found. Further, for disjunctive versions of the Email Case *none* of the three predictions made by the counterfactual view were borne out: significant cross-agent effects, varied agent effects, and fixed agent effects were found in each of two studies.

I've also raised doubts about whether the alternative accounts in the literature would actually be unable to explain the proposed pattern of effects, if the pattern were borne out. It is standardly accepted that these alternative accounts are able to explain the cross-agent and varied agent effects for conjunctive cases. Given that, I've argued that they are also able to indirectly explain the fixed agent effect. I made the case that this effect *might* be nothing more than a kind of artifact—a context effect with the varied agent providing the relevant context. And I reported on two studies showing that such a context effect can arise using the scales typically employed in work on causal attribution. While this does not establish that the fixed agent effect is simply a context effect, it does establish that it is a live possibility. With regard to disjunctive cases, I've argued that while a simple view focusing on just evaluations of the agents would lack the resources to explain why causal attributions vary with causal structure, the alternative accounts at issue are not like this. Rather they either focus on judgments that the agents are responsible *for the outcome* or to blame *for the outcome* or allow that such judgments are relevant to assessing the agent. As such, to test the alternative accounts, what is needed is to test further attributions in disjunctive cases. This was done in the two studies, and the results for responsibility and blame attributions were strikingly similar to those found for causal attributions.

References

Alicke, M. (1992). "Culpable causation." *Journal of Personality and Social Psychology*, 63: 368–378.

Alicke, M., Rose, D., and Bloom, D. (2011). "Causation, Norm Violation and Culpable Control." *Journal of Philosophy*, 108: 670–696.

Bloom, P. (2007). "Religion is Natural." *Developmental Science*, 10: 147–151.

Cushman, F. (2008). "Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment." Cognition, 108: 353–380.

Danks, D., D. Rose, and E. Machery (2014). "Demoralizing Causation." *Philosophical Studies*, 171: 251–277.

Gailey, J. and F. Falk (2008). "Attribution of Responsibility as a Multidimensional Concept." *Sociological Spectrum*, 28: 659–680.

Gerstenberg, T. and D. Lagnado (2010). "Spreading the blame: The allocation of responsibility amongst multiple agents." *Cognition*, 115: 166–171.

Gerstenberg, T. and D. Lagnado (2012). "When contributions make a difference: Explaining order effects in responsibility attribution." *Psychonomic Bulletin & Review*, 19: 729–736.

Halpern, J. and C. Hitchcock (2015). "Graded Causation and Defaults." *British Journal for the Philosophy of Science*, 66: 413–457.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Hilton, D. and B. Slugoski (1986). "Knowledge-based causal attribution: The abnormal conditions focus model." *Psychological Review*, 93: 75–88.

Hitchcock, C. and J. Knobe (2009). "Cause and Norm." *The Journal of Philosophy*, 106: 587–612.

Icard, T., J. Kominsky, and J. Knobe (2017). "Normality and actual causal strength." *Cognition,* 161: 80–93.

Knobe, J. (2006). *Folk Psychology, Folk Morality*. Dissertation.

Knobe, J. and B. Fraser (2008). "Causal judgments and moral judgment: Two experiments." In W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality*, pp. 441–447, Cambridge: MIT Press.

Kominsky, J. and J. Phillips (2019). "Immoral Professors and Malfunctioning Tools: Counterfactual Relevance Accounts Explain the Effect of Norm Violations on Causal Selection." *Cognitive Science*, 43(11): e12792.

Kominsky, J., J. Phillips, T. Gerstenberg, D. Lagnado, and J. Knobe (2015). "Causal superseding." *Cognition*, 137: 196–209.

Lagnado, D. and S. Channon (2008). "Judgments of cause and blame: The effects of intentionality and foreseeability." *Cognition*, 108: 754–770.

Lagnado, D. and T. Gerstenberg (2017). "Causation in Legal and Moral Reasoning." In M. Waldmann (ed.), *The Oxford Handbook of Causal Reasoning*, pp. 565–601, London: Oxford University Press.

Lagnado, D., T. Gerstenberg, and R. Zultan (2013). "Causal Responsibility and Counterfactuals." *Cognitive Science*, 37: 1036-1073.

Livengood, J. and E. Machery (2007). "The Folk Probably Don't Think What You Think They Think: Experiments on Causation by Absence." *Midwest Studies in Philosophy*, 31: 107–127.

Livengood, J., and J. Sytsma (forthcoming). "Actual causation and compositionality." *Philosophy of Science*.

Livengood, J., J. Sytsma, and D. Rose (2017). "Following the FAD: Folk attributions and theories of actual causation." *Review of Philosophy and Psychology*, 8(2): 274–294.

Malle, B. (2011). "Time to Give Up the Dogmas of Attribution: An Alternative Theory of Behavior Explanation." In M. Zanna and J. Olson (eds.), *Advances in Experimental Social Psychology, Volume 44*, pp. 297–352, San Diego: Elsevier.

Malle, B., S. Guglielmo, and A. Monroe (2014). "A Theory of Blame." *Psychological Inquiry*, 25: 147–186.

Moore, M. (2019). "Causation in the Law." In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2019 Edition)*, plato.stanford.edu/archives/win2019/entries/causation-law

Phillips, J., J. Luguri, and J. Knobe (2015). "Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities." *Cognition*, 145: 30–42.

Reuter, K., L. Kirfel, R. van Riel, and L. Barlassina (2014). "The good, the bad, and the timely: How temporal order and moral judgment influence causal selection." *Frontiers in Psychology*, 5: 1336.

Rose, D. (2015). "Persistence Though Function Preservation." *Synthese*, 192: 97–146.
Rose, D. (2017). "Folk Intuitions of Actual Causation: A Two-pronged Debunking Explanation." *Philosophical Studies*, 174(5): 1323–1361.

Rose, D. and Schaffer, J. (2017). "Folk mereology is teleological." *Noûs*, 51(2): 238–270.

Samland, J. and M. R. Waldmann (2016). "How prescriptive norms influence causal inferences." *Cognition*, 156: 164–176.

Samland, J., M. Josephs, M. Waldmann, and H. Rakoczy (2016). “The Role of Prescriptive Norms and Knowledge in Children’s and Adults’ Causal Selection.” *Journal of Experimental Psychology: General*, 145(2): 125–130.

Schwarz, N. (1996). “Survey Research: Collecting Data by Asking Questions.” In G. Semin and K. Fiedler (eds.), *Applied Social Psychology*, pp. 65–90, London: SAGE Publications.

Schwarz, N., T. Münkel, and H. Hippler (1990). “What determines a ‘perspective’? Contrast effects as a function of the dimension tapped by preceding questions.” *European Journal of Social Psychology*, 20: 357–361.

Sudman, S., N. Bradburn, and N. Schwarz (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers.

Sytsma, J., J. Livengood, and D. Rose (2012). “Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions.” *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43: 814–820.

Sytsma, J. and J. Livengood (under review). “Causal Attributions and the Trolley Problem.”

Sytsma, J., R. Bluhm, P. Willemsen, and K. Reuter (2019). “Causal Attributions and Corpus Analysis.” In E. Fischer and M. Curtis (eds.), *Methodological Advances in Experimental Philosophy*, London: Bloomsbury Press.

Woodward, J. (2006). “Sensitive and insensitive causation.” *The Philosophical Review*, 115(1): 1–50.

Young, L. and R. Saxe (2011). “When ignorance is no excuse: Different roles for intent across moral domains.” Cognition, 120: 202–214.

Zultan, R., T. Gerstenberg, and D. Lagnado (2012). “Finding fault: Counterfactuals and causality in group attributions.” *Cognition*, 125(3), 429–440.