The Effects of Single versus Joint Evaluations on Causal Attributions[1]

Justin Sytsma

Abstract: Recent research indicates that norms matter for ordinary causal attributions, although there is a good deal of debate concerning why they matter. One prominent account—that the impact of norms works via the salience of counterfactuals—has received support from a recent paper by Icard et al. (2017) reporting a new effect in cases where two agents perform symmetric actions that are each individually sufficient to bring about an outcome. But in four recent studies (Sytsma under review), I was unable to replicate these findings. In this paper I explore why, investigating a key difference between our studies: Icard et al. asked participants about just one agent (single evaluations), while I asked them about both agents (joint evaluations). I find that this difference helps explain the divergent findings, although the results remain problematic for Icard et al.'s view. Further I identity two evaluation effects: there is a general trend for the causal ratings in these cases to be lower when using single evaluations than when using joint evaluations, and this difference is larger when the agent asked about violates an injunctive norm. I consider four potential explanations of the impact of using single or joint evaluations and argue that determining the correct explanation has important implications for work concerning the effect of norms on causal attributions.

Norms, especially injunctive norms, matter for ordinary causal attributions.[2] But there is a great deal of debate about *why* they matter. One prominent account—the *counterfactual view*—has recently received strong support through a series of new findings from Kominsky et al. (2015) and Icard et al. (2017).[3] These articles identify a fascinating pattern of effects when typical scenarios from the literature are modified to remove the norm violation or to change the causal structure. But the implications of these findings and the reliability of the purported effects has been challenged by Sytsma (under review). Most importantly for present purposes, Icard et al. predict that a common effect in the literature for cases involving two agents performing symmetric actions will be *reversed* when the causal structure is changed from both of the two actions being required to bring about the outcome (*conjunctive*) to the actions being individually sufficient for the outcome (*disjunctive*), but

[1] I want to thank Jonathan Kominsky and Joshua Knobe for suggestions on a previous draft of this paper.

[2] See, for examples, 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), Rose (2017), Kominsky and Phillips (2019), Livengood and Sytsma (forthcoming), among others. By "ordinary causal attributions" I specifically mean the use of language like "X caused Y" (see Sytsma et al. 2019 for further discussion).

[3] See also Hitchcock and Knobe (2009), Halpern and Hitchcock (2015), Phillips et al. (2015), Kominsky and Phillips (2019).

I was unable to replicate this reversed effect across four studies involving two disjunctive cases. In this paper, I further explore this failure.

In Section 1, I detail these divergent findings and identify five differences between the studies conducted by Icard et al. and my studies that might explain the failure of replication. This includes that while Icard et al. asked each participant about just one of the agents (*single evaluations*), I followed the first two experiments from Kominsky et al. (2015) and asked each participant about both agents (*joint evaluations*). These differences are tested in Section 2. I find significant effects for type of evaluation: across two scenarios used by Icard et al., causal ratings were higher when using joint evaluations than when using single evaluations, and this difference was notably larger when the agent violated a norm. While no evidence of the predicted reverse effect was found using joint evaluations, it was seen in one of the two scenarios when using single evaluations. In Section 3, I considering four accounts of the role of single versus joint evaluations in arriving at causal judgments. I conclude by noting that insofar as the evaluation effects in part explain the divergent findings between Icard et al. (2017) and Sytsma (under review), determining the correct explanation of these effects is important for the debates over the impact of norms on ordinary causal attributions.

1. Conflicting Findings

Recent interest in the effect of norms on ordinary causal attributions owes in large part to a series of findings by Knobe and Fraser (2008) and Hitchcock and Knobe (2009). These articles presented the results of three experiments with a common structure: participants were given a vignette in which two agents (or in one study, two wires) perform symmetric actions, jointly bringing about a bad outcome. Each of these cases was *conjunctive*: the two actions together would lead to the outcome, but either action alone would not. The key difference between the actions was that one agent

violated a norm in performing it, while the other did not. Participants were then asked to assess a causal attribution for each agent using joint evaluations. Across these studies the authors found that causal ratings were significantly higher for the norm-violating agent than the norm-conforming agent. In Sytsma (under review), I refer to this as the *cross-agent effect*.

Hitchcock and Knobe (2009) explained the cross-agent effect in terms of norm-violations increasing the salience of counterfactuals on which something more normal was done instead. This *counterfactual view* has since been further developed, including by Kominsky et al. (2015) and Icard et al. (2017). And these papers offer additional support for the view by making a series of new predictions about what would happen when the norm violation was removed from scenarios like those used by Knobe and Fraser (2008) and Hitchcock and Knobe (2009) and/or when the causal structure was changed.

Adding a comparison condition in which neither agent violates a norm, when we look across the conditions we find that the normative status of one agent's action is *varied* (in one condition she violates a norm, in the other she does not), while the normative status of the other agent's action is *fixed* (she doesn't violate a norm in either condition). And it has standardly been found that causal ratings are higher for the varied agent in the condition where she violates the norm than in the condition where she does not. I will refer to this as the *varied agent effect*.[4] Kominsky et al. made a further prediction about such cases—that the reverse effect would be found for the fixed agent, with causal ratings being lower in the condition where the varied agent violates a norm than in the condition where the varied agent does not violate a norm. Further, Kominsky et al. predicted that this *fixed agent effect* would be absent when the causal structure was changed from *conjunctive* to *disjunctive*: rather than the two actions both being required for the outcome to occur as in

[4] Icard et al. (2017) refer to this as "abnormal inflation."

conjunctive cases, in disjunctive cases either action alone is sufficient to bring about the outcome.[5] And Kominsky et al. provided evidence supporting both predictions, with Experiments 1 and 2 using joint evaluations while Experiment 3 used single evaluations. The explanation offered, along with the reliability of the effects, has subsequently been challenged (Sytsma under review), however.

Extending this work on disjunctive cases, Icard et al. made a further prediction—that the varied agent effect will be reversed in disjunctive cases.[6] Their first experiment then tested this prediction for four cases involving injunctive norms using single evaluations. And they found evidence of the predicted effect. Further, Kominsky and Phillips (2019) have recently found the same effect. But across four studies, I was unable to find evidence of the reverse varied agent effect in two disjunctive cases matching or adapted from scenarios tested by Kominsky et al.—the Motion Detector Case and the Email Case—despite using larger sample sizes that should be able to detect even a quite small effect. In fact, in the two studies testing the Email Case, I found the opposite effect: there was a significant varied agent effect, rather than the predicted reverse varied agent effect. Further, putting the predictions of Kominsky et al. and Icard et al. together, they should also

[5] Kominsky et al. (2015) refer to the fixed agent effect as "causal superseding," while Icard et al. (2017) call it "supersession." Kominsky et al.'s explanation of this effect begins with the central claim of the counterfactual view—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. Kominsky et al. then 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, their explanation of the fixed agent effect is that in a normed conjunctive scenario, people recognize that the varied agent did something abnormal (violating the norm). This makes the counterfactual in which she does something more normal instead more salient, such that people are more likely to consider this counterfactual. And on this counterfactual, the outcome would not have occurred, which highlights the sensitivity of the sufficiency condition for the fixed agent. Finally, following Woodward, Kominsky et al. argue that when a sufficiency condition is judged to be highly sensitive, people are reluctant to attribute causation.

[6] Icard et al. refer to this as "abnormal deflation." As with Kominsky et al. (2015), Icard et al.'s prediction begins with the core idea behind the counterfactual view—that people are more likely to consider counterfactuals on which an abnormal event is replaced with a more normal one. The next step in predicting the reverse varied agent effect calls on considerations of *necessity*: Icard et al. hold 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 it would have occurred regardless of whether the agent acted). Putting these two ideas together, they predict that 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 caused it.

predict that there will be a *reverse cross-agent effect* in disjunctive cases where the actions are symmetric outside of the norm violation, but I also failed to find evidence of this effect.

What explains these divergent findings? And is there a reverse varied agent effect or not? To begin to answer these questions, it is important to note that there were several differences between the study in Icard et al. (2017) and the studies in Sytsma (under review). In addition to using different sources for our samples, and the possibility of attendant demographic variation, there are five differences that might potentially explain our divergent findings.

First, my replication of the Motion Detector Case was based on Experiment 3 from Kominsky et al., which used just one comprehension check, but Icard et al. added a second testing understanding of the causal structure. Thus, it is possible that my study included participants who misunderstood the causal structure and that their responses dramatically skewed the results. Second, my study for the Email Case used a disjunctive version of the conjunctive scenario tested in Kominksy et al.'s Experiment 2, which did not include a comprehension check, whereas Icard et al. used a different version with two comprehension checks. Thus, it is again possible that the responses of participants who misunderstood the causal structure skewed the results, and it is possible that the alternate wording affected participants' judgments (although it would then be an open question as to which version should be preferred). Third, Icard et al. only looked at participants who were 18 years of age or older, while my samples included participants who were 16 or 17. And it is possible that there is an age effect for such judgments (although it would then be an open question whether the judgments of older participants should be emphasized over those of younger participants). Fourth, Icard et al. did not exclude non-native English-speakers, while I did, following our standard practice in work concerning the responsibility view. Insofar as the studies at issue test judgments concerning the application of the English word "caused," if such judgments differed between native and non-native speakers, the case could be made that we should emphasize the judgments of more fluent

speakers and that this is likely to correlate with being a native speaker.[7] Finally, while Icard et al. asked participants about just the varied agent (single evaluations), I asked participants about both agents (joint evaluations), following previous studies in the literature, including the first two experiments from Kominsky et al. (2015). But it might be that there is an evaluation effect, with participants' judgments changing when they consider both agents compared to when they consider just one.

2. Testing the Differences

To further investigate the reverse varied agent effect, in the present experiment I replicated and expanded upon Icard et al.'s study with much larger sample sizes. To check the first and second differences between the relevant studies in Icard et al. (2017) and Sytsma (under review) noted above, I used the same vignettes and comprehension checks as they did for the two scenarios from my previous studies (Motion Detector Case, Email Case). To check the fifth difference, I varied whether participants were asked about both agents (joint evaluations) or about just one or the other of the two agents (single evaluations). To check the third difference, I included non-native English speakers in the data set. Finally, to check the fourth difference, I tested whether the inclusion of young participants (age 16 or 17) changed the occurrence of the effects.

[7] To test the third and fourth differences, I reanalyzed the data for causal attributions for the four studies using disjunctive scenarios from Sytsma (under review). First, I ran two ANOVAs with whether the participant was 16 or 17 and whether they were a native English-speaker as between-participant factors, in addition to *scenario* (motion detector, email) and *norm* (no violation, violation). For ratings of the varied agent, no significant effects were seen for either factor of interest. For ratings of the fixed agent, however, there was a significant interaction effect between age and scenario, $F(1, 1075)=5.19$, $p=0.023$, $\eta^2=0.005$. Second, I tested the three effects of interest for each of the two scenarios using the restrictions from Icard et al. (excluding 16- and 17-year-olds and including non-native English-speakers). These changes did not notably alter the results. Once again for the Motion Detector Case the difference in the means ran in the opposite direction to the predicted reverse cross-agent effect although the difference was not significant, $t(210)=0.41$, $p=0.34$, $d=0.04$ ($V=16885$, $p=0.37$), and similarly for the predicted reverse varied agent effect, $t(412.47)=1.16$, $p=0.12$, $d=0.11$ ($W=24310$, $p=0.14$), although in line with Kominsky et al. no fixed agent effect was found, $t(411.12)=0.86$, $p=0.20$, $d=0.08$ ($W=24128$, $p=0.17$). And for the Email Case, against the predictions there was still a significant cross-agent effect, $t(177)=5.93$, $p=7.8e^{-9}$, $d=0.60$ ($V=3247$, $p=2.6e^{-8}$), varied agent effect, $t(357.58)=2.26$, $p=0.012$, $d=0.24$ ($W=19356$, $p=0.015$), and fixed agent effect, $t(368.74)=3.57$, $p=0.00020$, $d=0.37$ ($W=20614$, $p=0.00024$). This suggests that the difference in exclusions does not explain the failure to replicate the previous findings.

2.1 Methods

Participants were given one of four causal scenarios from Icard et al. (2017)—either the norm violation or the no violation vignettes for either their disjunctive version of the Motion Detector Case or their disjunctive version of the Email Case.[8] Finally, I varied whether participants were asked about both agents, just the varied agent (named Billy in both cases), or just the fixed agent (named Suzy in both cases). In the Motion Detector Case the two agents work on a project that is important for national security. Their boss then either tells both of them to arrive at 9am the next morning (no violation) or tells Suzy to arrive at 9am and tells Billy that it is essential that he not arrive at that time (norm violation). It then turns out that a motion detector is installed in their room and that it will go off if at least one person enters. Both Billy and Suzy arrive at 9am. As such, the motion detector goes off. The Email Case involves the two agents working for a company with a central computer. Company policy is either that both are permitted to log into the computer in the morning (no violation) or that Suzy is permitted to log into the computer in the morning while Billy is not (norm violation). It then turns out that if anyone logs into the computer at exactly 9:27am, some important emails will be deleted. Both Billy and Suzy log in at exactly 9:27am. As such, some important emails are deleted.

After reading the vignette, participants were asked to rate their agreement or disagreement with the same causal attributions used by Icard et al. on a 7-point scale anchored at 1 with "strongly disagree," at 4 with "neutral," and at 7 with "strongly agree." For the Motion Detector case these were "Billy caused the motion detector to go off" and "Suzy caused the motion detector to go off"; for the Email Case they were "Billy caused the e-mails to be deleted" and "Suzy caused the e-mails to be deleted." In the joint evaluation conditions, the order of the two statements was randomized. After the evaluation(s), participants were given the two comprehension checks used by Icard et al.

[8] Full text for each vignette used in this paper is given in the supplemental materials.

in fixed order.[9] Participants for each experiment in this paper were recruited through advertising for a free personality test on Google, with ads being targeted to people in the United States. In addition to responding to the above questions, participants answered basic demographic questions—including age and whether they are a native English-speaker—and took a 10-item Big Five personality inventory after answering the target questions. Results were collected from 3678 participants age 16 or older. Of these, 860 failed one or both of the comprehension checks, leaving 2818 responses that were analyzed.[10, 11]

2.2 Results

Results are shown in Figures 1 and 2. An ANOVA looking at ratings for the varied agent with *norm* (no violation, violation), *scenario* (motion detector, email), and *design* (joint evaluations, single evaluations) as between-participants factors showed main effects for norm, $F(1, 1886)=25.04$, $p=6.1e^{-7}$, $\eta^2=0.012$, scenario, $F(1, 1886)=27.54$, $p=1.7e^{-7}$, $\eta^2=0.013$, and design, $F(1, 1886)=64.86$, $p=1.4e^{-15}$, $\eta^2=0.032$. In addition, there were significant interaction effects for norm and design, $F(1, 1886)=13.73$, $p=0.00022$, $\eta^2=0.007$, and norm and scenario, $F(1, 1886)=26.65$, $p=2.7e^{-7}$, $\eta^2=0.013$. A matching ANOVA looking at ratings for the fixed agent showed main effects for design, $F(1, 1826)=14.95$, $p=0.00011$, $\eta^2=0.008$, and scenario, $F(1, 1826)=4.02$, $p=0.045$, $\eta^2=0.002$, but not for

[9] For the Motion Detector Case the questions were "Who was supposed to show up at 9am?" (Suzy, Billy, Both of them) and "The motion detector would go off when it detected how many people?" (1 or more, 2 or more). For the Email Case the questions were "When is Billy allowed to log into the central computer?" (Morning, Afternoon) and "How many people need to log in to the computer at 9:27am to cause the data to be deleted?" (One or more, Two).

[10] 59.3% women (19 non-binary), average age 26.1 years, ranging from 16 to 89. One-way ANOVAs showed no effect for gender on either ratings of the varied agent, $F(2, 1891)=1.54$, $p=0.22$, $\eta^2=0.02$, or the fixed agent, $F(2, 1831)=0.38$, $p=0.68$, $\eta^2=0.00$.

[11] Of those who failed the checks, 10.2% missed both, while 55.5% missed just the first and 34.3% missed just the second. One-way ANOVAs showed no effect for ratings of the varied agent for failing just one or the other of the checks, but there was a significant effect for failing both, $F(1, 2483)=5.06$, $p=0.025$, $\eta^2=0.002$, although the effect size was negligible. Since the first two studies in Sytsma (under review) included the first check, this suggests that differences in the checks do not explain the failure to replicate the reverse varied agent effect.

norm, $F(1, 1826)=1.76$, $p=0.19$, $\eta^2=0.001$. There was also a significant interaction effect for design and scenario, $F(1, 1826)=6.86$, $p=0.0089$, $\eta^2=0.004$.

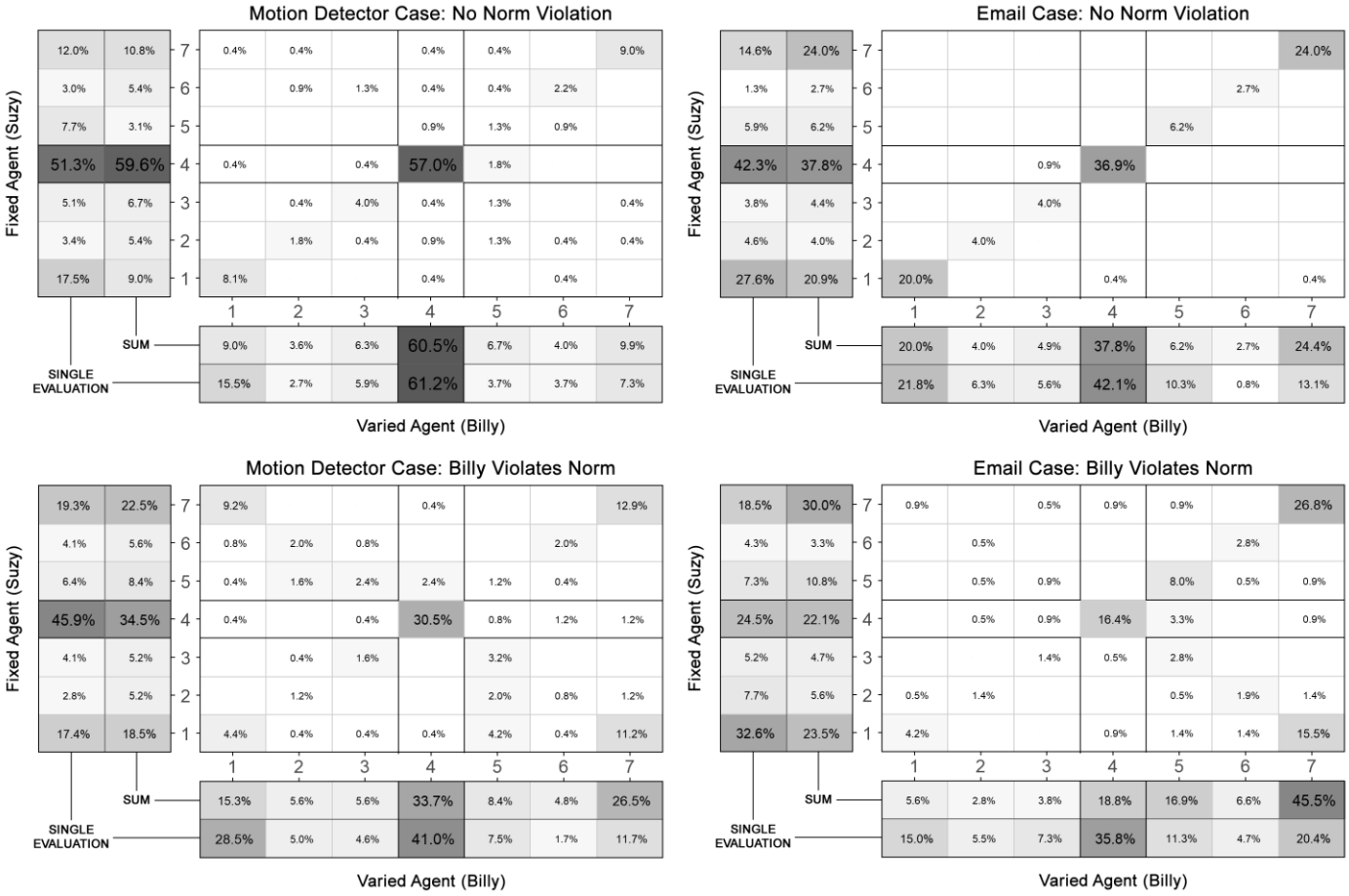

Figure 1: Heat maps showing distribution of patterns of responses for the two agents in the joint evaluation conditions (for example, 9.2% answered 7 for Suzy and 1 for Billy for the Motion Detector Case when Billy violated the norm). Additional columns show the summed percentages for each agent and the corresponding percentages for the single evaluation conditions.

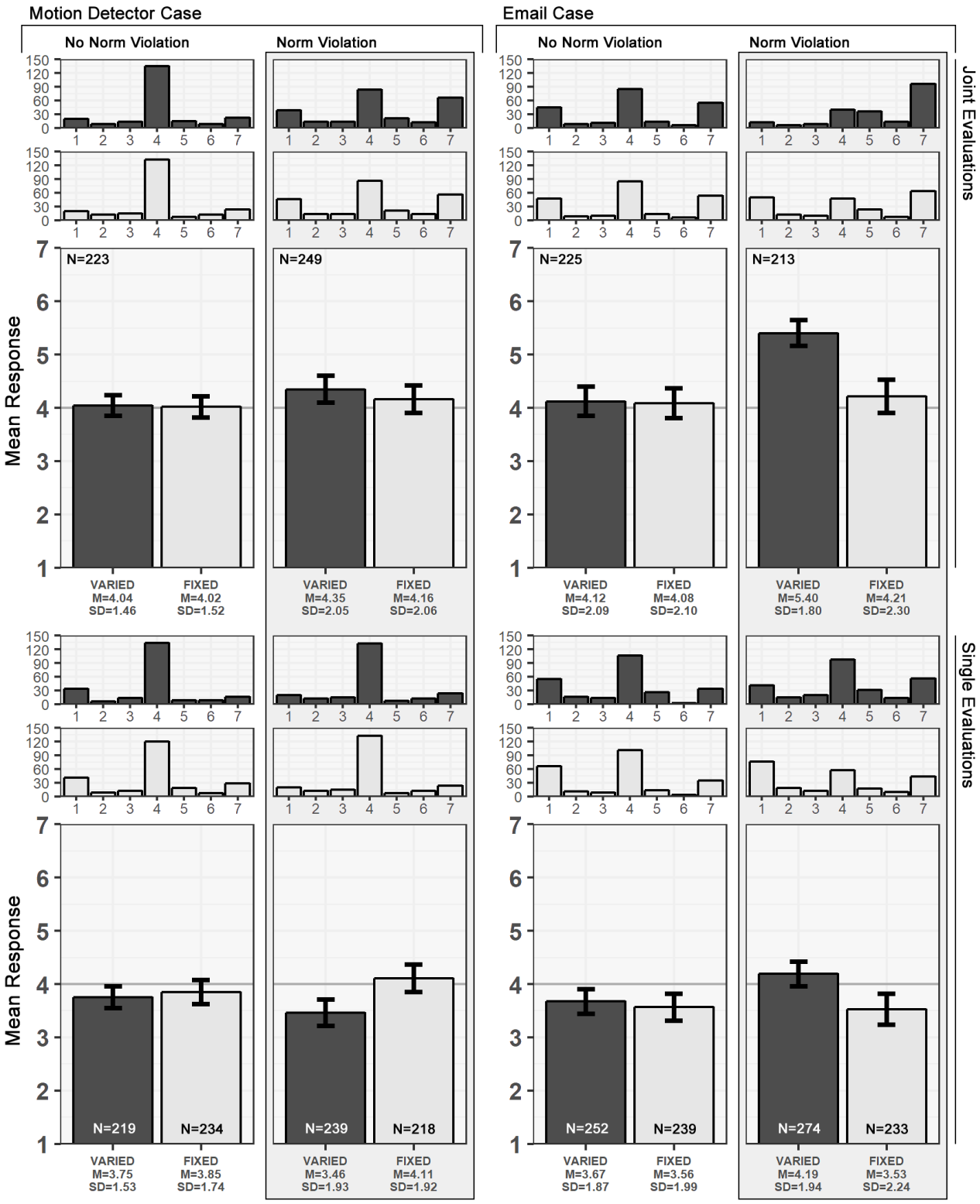

Figure 2: Results with histograms above the plots of the means for each condition and showing 95% confidence intervals.

The first thing to note is that there is a significant difference in responses across the two scenarios. This is especially pronounced for ratings of the varied agent, most notably when the varied agent violates the norm. In the joint evaluation conditions the mean rating for the varied agent when he violates the norm was 1.05 points higher for the Email Case than for the Motion Detector Case, $t(459.73)=5.89$, $p=3.8e^{-9}$, $d=0.54$ ($W=34498$, $p=3.9e^{-9}$). In the single evaluation conditions the mean rating was 0.73 points higher, $t(502.12)=4.26$, $p=1.2e^{-5}$, $d=0.38$ ($W=39389$, $p=3.9e^{-5}$).[12] And these differences are reflected in divergent findings between the scenarios for the reverse cross-agent and reverse varied agent effects, as discussed below.

The second thing to note is that design—whether participants were asked about both agents or just one—had a significant effect on ratings for both the varied agent and the fixed agent. Further, there was a significant interaction between design and norm for the varied agent. To investigate these evaluation effects further, I conducted planned analyses for each of the three effects noted above (cross-agent effect, varied agent effect, and fixed agent effect) for both the joint evaluation conditions and the single evaluation conditions for each of the two cases. Results of these tests are summarized in Table 1. In brief, for the Motion Detector Case, while the reverse cross-agent effect and reverse varied agent effect were not found comparing the joint evaluation conditions (in fact, there was a significant varied agent effect), significant effects were found comparing the single evaluation conditions. In contrast, for the Email Case, the reverse cross-agent effect and reverse varied agent effect were not found for either set of conditions; in fact, in each case there was a significant effect in the opposite direction. Importantly, though, the effect sizes were smaller in the single evaluation conditions than in the joint evaluation conditions. These results suggest that the

[12] Since the primary effects at issue are all directional—for instance, the reverse varied agent effect occurring when ratings for the varied 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 except where otherwise noted. And while it is typical to test for effects of norms on causal attributions using parametric statistics, the distributions are often non-normal, as is clear from Figure 2; 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.

reverse effects are more likely to be found using single evaluation (as in Icard et al.) than joint evaluation (as in my previous studies).

Case	Design	Cross-agent Effect	Varied Agent Effect	Fixed Agent Effect
Motion Detector	Joint	$t(248)=0.96$, $p=0.17$, $d=0.09$	$t(448.01)=1.90$, $p=0.030$, $d=0.17$	**R:** $t(453.89)=0.86$, $p=0.19$, $d=0.08$
		$V=3666$, $p=0.18$	$W=30628$, $p=0.020$	**R:** $W=26083$, $p=0.11$
	Single	**R:** $t(451.59)=3.58$, $p=1.9e^{-4}$, $d=0.34$	**R:** $t(447.19)=1.78$, $p=0.038$, $d=0.16$	**R:** $t(437.52)=1.48$, $p=0.07$, $d=0.14$
		R: $W=21494$, $p=3.3e^{-4}$	**R:** $W=24014$, $p=0.050$	**R:** $W=23646$, $p=0.076$
Email	Joint	$t(212)=6.80$, $p=5.1e^{-11}$, $d=0.58$	$t(431.96)=6.90$, $p=9.5e^{-12}$, $d=0.66$	**R:** $t(426.97)=0.60$, $p=0.27$, $d=0.058$
		$V=3075$, $p=5.0e^{-10}$	$W=32490$, $p=1.2e^{-11}$	**R:** $W=22940$, $p=0.21$
	Single	$t(462.72)=3.53$, $p=2.3e^{-4}$, $d=0.32$	$t(522.82)=3.10$, $p=0.0010$, $d=0.27$	$t(460.8)=0.19$, $p=0.42$, $d=0.017$
		$W=37580$, $p=2.1e^{-4}$	$W=39502$, $p=0.0015$	$W=28286$, $p=0.38$

Table 1: Tests of the cross-agent, varied agent, and fixed agent effects. All tests one-tailed, reverse effects indicated by "R," significant results in red.

To test the evaluation effects more directly, I analyzed the difference between the joint evaluation ratings and the single evaluation ratings for each of the four pairs of conditions, as summarized in Table 2. In each pair the mean rating was higher in the joint evaluation condition, and it was significantly higher in four of the six conditions. Further, the difference in the means for the varied agent when he violates the norm was notably larger than for the other three conditions for each scenario—over five times larger than the average of the other three for the Motion Detect Case and over twice as large for the Email Case—and this is reflected in the difference in effect sizes.[13] This indicates that there are two evaluation effects for these scenarios. First, we find a general trend for causal ratings to be higher when using joint evaluation; second, we find that this difference is greater for the varied agent when he violates the norm.

[13] For the Motion Detector Case there was a small effect for the varied agent when he violates the norm, while each of the other effect sizes was negligible (with the former being over four times larger than the average of the latter). For the Email Case there was a medium effect for the varied agent when he violates the norm, while the other effects were small (with the former being over two times larger than the average of the latter).

Case	Norm	Varied Agent	Fixed Agent
Motion Detector	No Violation	$t(438.08)=2.04$, $p=0.021$, $d=0.02$	$t(451.68)=1.10$, $p=0.14$, $d=0.10$
		$W=26785$, $p=0.022$	$W=26908$, $p=0.26$
	Violation	$t(485.81)=4.93$, $p=5.6e^{-7}$, $d=0.45$	$t(463.18)=0.30$, $p=0.38$, $d=0.028$
		$W=36801$, $p=1.4e^{-6}$	$W=27688$, $p=0.35$
Email	No Violation	$t(452.03)=2.44$, $p=0.0076$, $d=0.23$	$t(455.89)=2.73$, $p=0.0033$, $d=0.25$
		$W=31509$, $p=0.014$	$W=30504$, $p=0.0044$
	Violation	$t(470.16)=7.14$, $p=1.8e^{-12}$, $d=0.65$	$t(437.71)=3.17$, $p=8.1e^{-4}$, $d=0.30$
		$W=39710$, $p=9.5e^{-13}$	$W=29017$, $p=7.6e^{-4}$

Table 2: Tests of evaluation effects, comparing joint evaluation to single evaluation condition. All tests one-tailed, reverse effects indicated by "R," significant results in red.

The results of the present experiment suggest that the divergent results found by Icard et al. (2017) and Sytsma (under review) are at least in part explained by evaluation effects. Further, the results for the joint evaluation conditions are largely in line with my previous findings—each study using the Motion Detector Case failing to find the predicted reverse cross-agent and reverse varied agent effects, while each study using the Email Case instead found significant cross-agent and varied agent effects. This indicates that the failure of replication was not due to the differences in comprehension checks, the wording of the Email Case, or the exclusion of non-native English-speakers.[14]

The results reported above, however, differ from those reported by Icard et al. in including 16- and 17-year-olds. As such, I repeated the ANOVAs with the excluded age group as a fourth factor. No significant effects on ratings of the varied agent were found for this factor, indicating that the failure to replicate the reverse varied agent effect is not due to this inclusion. For ratings of the fixed agent, however, there was a significant main effect, $F(1, 1818)=4.76$, $p=0.029$, $\eta^2=0.003$, although the effect size was negligible. Further, the effects were comparable when excluding this

[14] Two differences are worth noting: first, in the previous studies I found no effects for the Motion Detector Case, while this time I found a significant varied agent effect; second, in the previous studies I found a significant fixed agent effect for the Email Case, but no effect was found for the fixed agent in the present experiment.

age group, as summarized in Table 3, except that now a significant reverse fixed agent effect is found in the joint evaluation conditions, which runs counter to the prediction made by Kominsky et al. (2015).

Case	Design	Cross-agent Effect	Varied Agent Effect	Fixed Agent Effect
Motion Detector	Joint	R: $t(99)=0.25$, $p=0.40$, $d=0.04$	$t(175.94)=1.33$, $p=0.093$, $d=0.19$	R: $t(176.79)=1.76$, $p=0.040$, $d=0.26$
		R: $V=492$, $p=0.39$	$W=30628$, $p=0.020$	R: $W=3334.5$, $p=0.029$
	Single	R: $t(251.18)=3.31$, $p=5.4e^{-4}$, $d=0.41$	R: $t(250.66)=2.17$, $p=0.016$, $d=0.27$	R: $t(230.05)=0.77$, $p=0.22$, $d=0.10$
		R: $W=6506.5$, $p=4.9e^{-4}$	R: $W=6732.5$, $p=0.010$	R: $W=6413.5$, $p=0.23$
Email	Joint	$t(133)=5.81$, $p=2.2e^{-8}$, $d=0.63$	$t(258.26)=5.60$, $p=2.7e^{-8}$, $d=0.68$	R: $t(264.73)=0.22$, $p=0.41$, $d=0.03$
		$V=1370$, $p=1.8e^{-7}$	$W=12375$, $p=5.2e^{-8}$	R: $W=8867$, $p=0.35$
	Single	$t(329.98)=2.59$, $p=0.0050$, $d=0.28$	$t(312.41)=2.06$, $p=0.020$, $d=0.23$	R: $t(330.79)=0.14$, $p=0.44$, $d=0.02$
		$W=16125$, $p=0.0053$	$W=13890$, $p=0.027$	R: $W=13790$, $p=0.47$

Table 3: Tests of the cross-agent, varied agent, and fixed agent effects, removing 16- and 17-year-olds. All tests one-tailed, reverse effects indicated by "R," significant results in red.

2.3 Discussion

To summarize, of the five differences I noted between the studies in Icard et al. (2017) and Sytsma (under review), the results of the present experiment suggest that only the use of single versus joint evaluations is a likely explanation of the divergent findings with regard to the reverse varied agent effect. In fact, the experiment provides evidence of two evaluation effects: first, there was a *general effect*, with causal ratings tending to be higher when using joint evaluations; second, there was a *specific effect*, with this difference to be greater for the varied agent when he violates the norm. Focusing on the reverse varied agent effect, while it was not seen in the joint evaluation conditions, it was found when comparing the single evaluation conditions for the Motion Detector Case. Even using single evaluations, however, the effect was not found for the Email Case (in fact the opposite effect was found). And, it should be noted that in the one set of conditions where the reverse varied agent effect was found, the effect size was negligible (d=0.16) and the difference in the means was notably smaller than the difference found by Icard et al. (0.29 versus 1.07).

The upshot is that only the results for the Motion Detector Case using single evaluations provide any support for the prediction made by Icard et al. (2017), and even here the effect is slight. In contrast, the results for the Email Case and the results for the Motion Detector Case using joint evaluations suggest against the counterfactual view. As such, the present experiment only offers support for the counterfactual view if there is reason to discount both the use of joint evaluations *and* the use of the Email Case—despite results using joint evaluations and results using the Email Case being part of the evidence previously offered for the counterfactual view. Correspondingly, the results raise two explanatory issues: first, the difference between the two scenarios needs to be explained; second, the evaluation effects need to be explained. In the remainder of this paper, I'll focus on the second issue.

3. Explaining the Evaluation Effects

In the previous section we saw evidence for two evaluation effects in disjunctive cases: a *general effect* (across the board ratings were higher when employing joint evaluations than when employing single evaluations) and a *specific effect* (the difference was larger for the varied agent when he violates the norm). How are we to explain these effects? In this section, I'll consider four possible accounts.

4.1 Equal Treatment

The first account I'll consider was suggested by Jonathan Kominsky. Call this the *equal treatment* account. Focusing on the norm violation conditions in the motion detector case, the idea is that the counterfactual view explains the results for the single evaluation conditions: participants tend to focus on the counterfactual in which the varied agent does not violate the norm; then, noting that the outcome would still occur on that counterfactual, causal ratings for the varied agent are deflated,

producing the reverse varied agent effect. The equal treatment account then argues that this effect is broken in the joint evaluation conditions due to pragmatic pressures. The idea is that when asked about both agents, participants will tend to note that it is strange to treat the agent who violated the norm as *less causal* than the agent who did not violate the norm; after all, their actions were symmetric outside of the norm-violation.

While the equal treatment account offers a plausible suggestion concerning how using joint evaluations might impact participants' thinking about the task, the data from the present experiment does not bear out this explanation. First, looking at the heat maps for the Motion Detector Case in Figure 1, we see that when Billy does not violate the norm, 83.4% of participants treat the two agents equivalently. This is significantly higher than in the condition where Billy does not violate the norm, with just 53.8% of participants treating the two agents equivalently in that condition, χ^2=45.84, p=6.4e^{-12}. Thus, it does not appear to be the case that the failure to find the reverse varied agent effect when using joint evaluations is due to participants thinking it is strange to treat the norm-violating agent as less causal than the norm-conforming agent. Instead, the primary difference appears to be that in the norm violation condition a significantly lower percentage of participants give equivalent neutral ratings for both agents (30.5% vs. 57.0%, χ^2=32.45, p=6.1e^{-9}), with a corresponding increase in participants giving extremely unequal responses (11.2% versus 0% answering 7 for Billy and 1 for Suzy, χ^2=27.59, p=7.5e^{-8}; 9.2% versus 0.4% answering 1 for Billy and 7 for Suzy, χ^2=17.05, p=1.8e^{-5}). This is not readily explained in terms of there being a pragmatic pressure to treat the two agents equivalently in the norm violation condition when using joint evaluation. In fact, quite the reverse.

A similar pattern is found for the corresponding conditions for the Email Case. Once again there was a significant decrease in the percentage of participants treating the two agents equally from the condition in which Billy does not violate a norm (97.8%) to the condition where he

violates the norm (61.0%), χ^2=89.74, p<2.2e^{-16}. Further, the primary shift was away from treating both agents as completely non-causal (20.0% to 4.2%, χ^2=23.75, p=5.5e^{-7}) or both agents as neutral (36.9% to 16.4%, χ^2=22.24, p=1.2e^{-6}), with the largest increase being in participants answering 7 for Billy and 1 for Suzy (0.4% to 15.5%, χ^2=9.57, p=9.8e^{-4}).

Finally, it is unclear how the equal treatment account would explain the general effect. Insofar as the proposed mechanism produces the specific effect by breaking the reverse varied agent effect, it would not apply to the fixed agent or to the varied agent when he does not violate the norm. As such, this account seems unable to explain the general difference in causal ratings seen between joint and single evaluations.

4.2 Reflectivity

The second account I'll consider was suggested by Joshua Knobe. I will refer to this as the *reflectivity account*. The basic idea is that asking participants about both agents promotes thinking about the role of each in the scenario, leading to more reflective judgments; in contrast, asking about just one of the agents won't exert this pressure and participants will tend to give more intuitive judgments. This could then be spelled out in different ways depending on what one thinks the intuitive judgments look like.

Starting with the counterfactual view, one might suggest that intuitive judgments follow counterfactual saliency, focusing on the Motion Detector Case instead of the Email Case. It would then be intuitive to judge that Billy is less causal, relative to a baseline, when he violates a norm. The counterfactual view is not clear on how the baseline is established, but would predict that when Billy violates the norm, ratings for Suzy will be at baseline, and similarly for both agents when there is no norm violation. This version of the account would then argue that when asked about both agents, participants will reflect further on the situation, considering the role of each agent. One

possibility is that this would prompt participants to think about the agents jointly, perhaps noting that if neither had performed the action, the outcome would not have occurred. Participants might then partition the role in causing the outcome between the agents, perhaps assigning a greater role to Billy when he violates the norm. This would then serve to elevate ratings in general (explaining the general effect), and to further elevate ratings for the varied agent when he violates the norm (explaining the specific effect).

Alternatively, one could start with a different account of the intuitive judgments. For instance, one might focus on the norm-violation conditions of the Email Case instead of the Motion Detector Case, finding the default tendency to be to treat the norm-violating agent as more causal, then applying the same reasoning as above. This version of the reflectivity account would also be able to explain both the general effect and the specific effect, but it would not be in line with the counterfactual view.

4.2.1 Time Spent

One source of evidence with regard to the reflectivity account is the length of time that participants spent answering the questions in the above experiment. If asking participants about both agents promotes thinking more deeply about the scenario, then we would expect them to take longer in the joint evaluation conditions. And if thinking more deeply about the scenario explains the evaluation effects, we would expect to see a positive correlation between time spent and responses across conditions (general effect), with a larger correlation for the varied agent in the conditions where he violates the norm (specific effect). It should be noted, however, that even if the reflectivity account is false, there would be reason to expect participants to take slightly longer in the joint evaluation conditions (since they are asked an additional question in these conditions that would take some time to read and answer). Further, based on the observed evaluation effects in conjunction with the

expected difference in time spent, we would expect time spent to show a slight positive correlation with participants' responses.

To assess the predictions of the reflectivity account, I looked at the timing data for the experiment reported in Section 2. Since time spent is susceptible to outliers, I removed participants who were outside 1.5 times the interquartile range above the upper quartile or bellow the lower quartile for each condition (145 participants in total, 5.1%). The average time spent on the probes for the remaining participants was 90.5 seconds, although this varied across the conditions.[15] When Billy violated the norm in the Motion Detector Case, participants spent 6.4% (5.8s) longer for joint evaluation than single evaluation, $t(493.12)=2.18$, $p=0.015$, $d=0.17$. To put this in context, counting the words in the vignettes and questions, the joint evaluation text is 5.3% longer than the single evaluation text. When neither agent violated the norm, participants spent 8.4% (6.5s) longer, $t(379.02)=2.62$, $p=0.0045$, $d=0.23$. (Text 6.9% longer.) When Billy violated the norm in the Email Case, participants spent 15.9% (14.2s) longer for joint evaluation than single evaluation, $t(330.09)=3.21$, $p=1.7e^{-5}$, $d=0.38$. (Text 4.3% longer.) And when neither agent violated the norm, participants spent 16.3% (14.6s) longer, $t(349.12)=4.54$, $p=3.9e^{-6}$, $d=0.41$. (Text 4.7% longer.) Overall, given the difference in text length, not to mention the need to then select an answer for a further question, while participants did spend longer in the joint evaluation conditions for the Motion Detector Case, the difference is not large enough to make a compelling case for the reflectivity view. The prospects are slightly more promising for the Email Case, however.

Turning to the correlations between time spent and participants' responses, aggregating across conditions there was a significant positive correlation for ratings of the varied agent, although the size was negligible, $r=0.059$, $t(1799)=2.51$, $p=0.0060$. The correlation was not significant for ratings of

[15] An ANOVA looking at time spent showed significant main effects for norm, $F(1, 2795)=7.02$, $p=0.0081$, $\eta^2=0.002$, scenario, $F(1, 2795)=22.24$, $p=2.5e^{-6}$, $\eta^2=0.008$, and whether participants were asked about both agents, just the fixed agent, or just the varied agent, $F(2, 2795)=10.11$, $p=4.2e^{-5}$, $\eta^2=0.007$; no interaction effects were seen.

the fixed agent, r=0.022, t(1735)=0.022, p=0.17. While the correlation is larger for the varied agent, which is in line with the prediction made concerning the specific effect, neither correlation would appear sufficient to explain the evaluation effects, especially given that we would expect slight correlations even if the reflectivity account were false. Further, the correlations for individual conditions ran in both directions (nine positive, seven negative) and were generally negligible, with none of them being significant using two-tailed tests. Overall, the timing data suggests against the reflectivity account.

4.3 Counterfactual Salience

The third account I'll consider offers a variation on the counterfactual view, although it is unclear how readily this can be fit together with the view given by Kominsky et al. (2015) and Icard et al. (2017). The basic idea is that being asked about just the one agent might make the counterfactual on which that agent doesn't act more salient. In contrast, when people are asked about both agents, this effect on counterfactual salience is cancelled out such that neither counterfactual is made more salient. This might then explain the general evaluation effect. In the joint evaluation conditions, there would be no change in the saliency of the different counterfactuals, and thus no effect on causal ratings. By contrast, in single evaluation conditions, the counterfactual on which the agent asked about did not act would be more salient, and since the outcome would still occur on that counterfactual, causal ratings would be lower.

What about the specific effect? Here we might note that counterfactual saliency is not an all-or-nothing thing and that different factors could make a given counterfactual more or less salient.[16] The counterfactual view holds that norm violations will make the counterfactual on which the norm

[16] Alternatively, this explanation could be spelled out in terms of different factors making a given counterfactual more likely to be salient to participants. Since the same explanation would follow on either version, I'll leave this to the side.

violation didn't occur more salient. So, in the conditions where the varied agent violates the norm, the counterfactual on which he didn't act is expected to be more salient. Accepting that asking about both agents won't have an independent effect on counterfactual saliency, in the conditions where the varied agent violates the norm it would only be this norm violation that has an impact. In the corresponding single evaluation conditions where participants are asked about the varied agent, however, both the norm violation and the question would serve to make the counterfactual on which that agent didn't act more salient, further reducing causal ratings.

If we think of factors affecting counterfactual salience as being simply additive, the explanation derived from the counterfactual view will not explain the specific effect. The reason is that the impact of the norm violation on ratings of the varied agent would be the same whether asking about both agents or just the varied agent, and this would match the general effect. To make this clearer, let's call the impact that the norm violation has on counterfactual saliency **N** and the impact that using single evaluations has on counterfactual saliency **E**. If these effects are additive, then as seen in Table 4, the evaluation effect would correspond to that produced by **E** for each pair of conditions, including for the varied agent when he violates the norm. So, we would have a general effect, but not a specific effect. Further, this account would predict the same reverse cross-agent and reverse varied agent effects in each relevant pair of conditions, whether using joint or single evaluations, and this would correspond to the effect produced by **N**. But the goal was to explain the differences with regard to these effects.

One need not assume that the effects on counterfactual salience would simply be additive, however. It might be suggested that when multiple factors affect the salience of a counterfactual, these effects reinforce each other, operating in a synergistic fashion. And this would enable the view to explain the specific effect. Let's call the interaction between the two effects on counterfactual saliency **I**. As seen in Table 4, this would mean that comparing the joint and single evaluation

conditions for the varied agent when he violates the norm, we would now predict an evaluation effect that corresponds with **E+I**, while the other comparisons would remain as above. Thus, by positing a further interaction effect between the factors affecting counterfactual saliency, this account is able to explain both the general and specific effects. Further, this would also lead to a difference in the expected effects for studies using joint versus single evaluations: we would expect the reverse cross-agent effect and reverse varied agent effect to be larger by an amount corresponding with the impact of **I** when using single evaluations. While this explanation would still predict that we should find these effects using joint evaluations, since the effect would be smaller it could be argued that it is *less* surprising that they were not found in those studies (if still surprising given the sample sizes).

		Effect on Saliency of Counterfactual	
		Varied Agent	**Fixed Agent**
No Norm Violation	Joint Evaluation		
	Single Evaluation	E	E
Evaluation Effect: Difference Between Joint and Single		**E**	**E**
Norm Violation: Additive	Joint Evaluation	N	
	Single Evaluation	N + E	E
Evaluation Effect: Difference Between Joint and Single		**E**	**E**
Norm Violation: With Interaction	Joint Evaluation	N	
	Single Evaluation	N + E + I	E
Evaluation Effect: Difference Between Joint and Single		**E + I**	**E**

Table 4: Effects on counterfactual saliency and corresponding evaluation effects for the explanation derived from the counterfactual view.

4.4 Singling-out

The final account I'll consider is based on our competing explanation of the effect of norms on causal attributions—the *responsibility view*.[17] I'll refer to this as the *singling-out account*. To set-up this account, I need to first briefly lay out the responsibility view and detail how this view applies to disjunctive cases.

The core idea of the responsibility view is that the ordinary concept of causation typically at play in causal attributions has a normative component, with the result that causal attributions are generally akin to responsibility attributions. Thus, causal attributions—claims that X caused Y—typically serve to indicate something more than that an entity brought about an outcome; they also express a normative evaluation similar to saying that the entity is responsible for or accountable for the outcome. As such, while we expect that there will be some differences between causal attributions and responsibility attributions, we take responsibility attributions to be a good guide to what causal attributions will look like. And in the two studies that tested responsibility attributions in Sytsma (under review), causal ratings and responsibility ratings for the Motion Detector Case were remarkably similar. As I argue there, it is at best unclear that the occurrence of the reverse cross-agent and reverse varied agent effects in disjunctive cases would be problematic for our view.

What should the responsibility view say about causal ratings for disjunctive cases? Thinking about such cases in terms of responsibility attributions, it is not clear cut who we should take to be responsible for the outcome, with competing considerations pointing in different directions. The first consideration starts by noting that the two agents' actions are symmetric outside of the norm violation. When neither does anything wrong, this suggests that they are equally responsible. This is

[17] See Sytsma et al. (2012), Livengood et al. (2017), Sytsma et al. (2019), Livengood and Sytsma (forthcoming), Sytsma and Livengood (under review), Sytsma (under review). While I will focus on the responsibility view here, similar points might be made following other views from the literature, including Alicke's blame view (Alicke 1992, 2000; Alicke et al. 2011; Rose 2017) and Samland and Waldmann's pragmatic view (Samland and Waldmann 2016, Samland et al. 2016).

shifted, however, when the varied agent violates the norm. All else being equal, we are typically inclined to blame rule-breakers for the harms that result, which suggests that the varied agent deserves a greater share of the responsibility. The second consideration starts by noting that all else is seldom equal and that in assigning responsibility we typical think about whether there are mitigating factors. In disjunctive cases, one potential mitigating factor is the other agent and the fact that her action would have brought about the outcome in any case. Recognizing this, when the varied agent violates the norm, we might reasonably think that his misdeed was not the root problem, since the outcome would have occurred anyway. And, absent information in the scenario suggesting otherwise, we might then reason that the varied agent did not know that his action would bring about the outcome, and hence that he presumably did not desire to bring about the outcome.[18] Alternatively, we might suspect that the varied agent knew that the other agent would be performing her action, and as such had good reason to believe that the outcome would occur regardless of his action. Either way, this would seem to mitigate the varied agent's responsibility, as he would not have reason to believe that his action would alter the outcome at issue.

Thinking about responses to disjunctive cases in terms of these opposing considerations arguably fits well with the distributions of responses seen in Figure 1. As noted above, when the varied agent does not violate the norm, participants tend to treat the two agents equivalently. When the varied agent violates the norm, however, we see responses shift away from this. In the Motion Detector Case, the most notable shift is away from treating the two agents as both neutral, with a corresponding increase in participants showing complete agreement with the causal attribution for the varied agent and complete disagreement with the causal attribution for the fixed agent or vice versa. This can be explained in terms of different participants tending to notice or to emphasize one

[18] There is a good bit of empirical work indicating that factors related to an agent's intentions and the foreseeability of the outcome matter for responsibility judgments (see, e.g., Cushman 2008, Gailey and Falk 2008, Lagnado and Channon 2008, Malle et al. 2014, Young and Saxe 2011).

or the other of the two considerations laid out above. And a similar pattern is seen for the Email Case, although now the responses tend to shift toward complete agreement with the causal attribution for the varied agent.

The singling-out account expands on this basic framework for explaining causal attributions in disjunctive cases. The idea is that using single evaluations produces a pragmatic effect: asking about just one person in a complicated situation involving two people raises the question of why the same question wasn't asked about the second person, and this suggests that the first person is being singled out. This is likely to strike some participants as being unfair and might lead them to temper their responses. When just one agent is singled out, participants might worry that any agreement that the agent caused the outcome would be to suggest that that agent was solely responsible for the outcome, when the situation is in fact more complicated than that. And, as such, they might be reluctant to show agreement, giving a lower rating than they might otherwise have done.

The singling-out account can readily explain the general effect: taking causal attributions to be akin to responsibility attributions, when asked about just one agent people will have a tendency to think about the role of the other agent and so temper their judgment. In contrast, when asked about both agents this tendency won't arise (at least not with regard to the other agent). As such, we would predict that causal ratings will be lower in the single evaluation conditions than in the corresponding joint evaluation conditions. With regard to the specific effect, the same type of explanation can be called on. When comparatively assessing the two agents in the norm violation conditions, participants will be likely to recognize that the only relevant difference between them is that one violated a norm while the other did not, and this might prompt them to assign a higher degree of responsibility to the varied agent. Taken together, we would then expect to find the specific effect.

4. Conclusion

In this paper I've provided evidence that the failure of the studies in Sytsma (under review) to replicate two effects for causal attributions predicted by the counterfactual view can be in part explained by a pair of effects seen when varying whether participants are asked about both agents in a scenario or just one. And I've offered four potential explanations of these evaluation effects. While the first two are inconsistent with the data from the present experiment, the last two remain open possibilities, although it is beyond the scope of this paper to test them. And, of course, it might be that some further explanation all together could be given.

I close by noting that explaining these evaluation effects is important for the debates on the effect of norms on ordinary causal attributions, since some key findings at play in these debates appear to rest, in part, on whether the studies employ joint or single evaluations. And whether we should focus on one type of study or the other would seem to depend, at least in part, on how we explain these effects. If the explanation derived from the responsibility view is accurate, then we should focus on studies employing joint evaluation, with the result that the present findings would strongly favor the responsibility view over the counterfactual view. The reason is that on this explanation, the dip in causal ratings for single evaluations would reflect pragmatic concerns about unfairly singling out a given agent for blame. If instead the variation on the counterfactual view is accurate, then we should instead focus on studies employing single evaluations, with the result that the present results by themselves would not favor the responsibility view over the counterfactual view. The reason is that on this explanation, the dip in causal ratings for single evaluations would reflect that such evaluations make the corresponding counterfactual more salient, amplifying the relevant effect.

That said, even if we should prefer single evaluations over joint evaluations for studying causal attributions (and it is at best unclear that we should), the present results are still problematic

for the counterfactual view since the predicted effects are reversed in the Email Case even when using single evaluations. Further, these effects are notably larger than those found for the Motion Detector Case when using single evaluations. Absent an account that explains away these results, the evidence more strongly disconfirms than confirms the predictions of the counterfactual view.

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.

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

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

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

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. 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.

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

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.

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.

Sytsma, J. (under review). “Structure and Norms: Investigating the Pattern of Effects for Causal Attributions.” http://philsci-archive.pitt.edu/16626/

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.” http://philsci-archive.pitt.edu/16200/

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.