Causation, Responsibility, and Typicality

Justin Sytsma[1]

Abstract: There is ample evidence that violations of injunctive norms impact ordinary causal attributions. This has struck some as deeply surprising, taking the ordinary concept of causation to be purely descriptive. Our explanation of the findings—the responsibility view—rejects this: we contend that the concept is in fact normative, being akin to concepts like responsibility and accountability. Based on this account, we predicted a very different pattern of results for causal attributions when an agent violates a statistical norm. And this pattern has been borne out by the data (Sytsma et al. 2012, Livengood et al. 2017, Sytsma under review a). These predictions were based on the responsibility attributions that we would make. In this paper, I extend these previous findings, testing responsibility attributions. The results confirm the basis of our predictions, showing the same pattern of effects previously found for causal attributions for both injunctive norms and statistical norms. In fact, the results for responsibility attributions are not statistically significantly different from those previously found for causal attributions. I argue that this close correspondence lends further credence to the responsibility view over competing explanations of the impact of norms on causal attributions.

Injunctive norms have a notable impact on ordinary causal attributions.[2] To illustrate, consider the Pen Case from 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.

[1] I want to thank Joshua Knobe for very helpful feedback on a previous draft of this paper.

[2] See, for example, 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), and Livengood and Sytsma (forthcoming). By "ordinary causal attributions" I mean the use of language like "X caused Y". Under "injunctive norms" I include both prescriptive norms (what should be done) and proscriptive norms (what should not be done), although "prescriptive norm" is often used to refer to both.

In this case, two agents perform symmetric actions, jointly bringing about a bad outcome. The only difference is that Professor Smith violated an injunctive norm (faculty members aren't supposed to take pens), while the administrative assistant did not (administrative assistants are allowed to take pens). When participants were asked to rate agreement with a causal attribution for each agent, however, ratings were very notably higher for Professor Smith.

How should we explain findings like this? In a series of papers my colleagues and I have argued that injunctive norms matter for ordinary causal attributions concerning agents because the ordinary concept of causation at play in such attributions is a broadly moral concept, akin to the concepts of responsibility and accountability.[3, 4] Our underlying motivation for this *responsibility view* is that human cognition is highly attuned to recognizing applicable injunctive norms (whether implicit or explicit), detecting and responding to violations, and navigating factors that exacerbate or mitigate those motivations.[5] As Joshua Knobe (2010, 328) has put it, "we are moralizing creatures through and through." Taking moral considerations broadly, such that they include norms that fall well short of what we might want to term "moral," our view follows directly from this basic insight: if humans are (broadly) moralizing creatures through and through, then we should not find it surprising that a host of ordinary concepts are responsive to our broadly moral evaluations, including both causation and responsibility. Thus, while some have thought that causation is a purely descriptive matter, such that broadly moral evaluations

[3] 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 a, b, c). Our focus has primarily been on attributions to agents and this will be my focus here. There is some evidence, however, that the impact of injunctive norms extends to what philosophers would consider non-agents (e.g., Hitchcock and Knobe 2009). We believe that such effects can be explained in terms of the application of the same normative concept at play in causal attributions concerning agents. In line with this, research indicates that people have a tendency to take an agentive perspective on nature as a whole, including what philosophers would think of as non-agents (e.g., Bloom 2007, Rose 2015, and Rose and Schaffer 2017).

[4] While discussions of the impact of injunctive norms on causal attributions often describe these in terms of moral considerations, this is arguably too strong. In keeping with Sytsma et al. (2012), I'll instead speak of broadly moral considerations, where this is intended to correspond with the injunctive norms at issue.

[5] That we are highly attuned to norms and their violation is hardly controversial (e.g., Sripada and Stich 2007), even as there is a great deal of disagreement about why we are so attuned and how this came about.

must impact causal attributions *indirectly*, we deny this: we hold that evaluations of injunctive norms are *directly* relevant to causal attributions.

In previous work (Sytsma et al. 2012, Livengood et al. 2017, Sytsma under review a), we have shown what might seem like a surprising pattern of findings concerning the effects of another type of norm on causal attributions—statistical norms (i.e., whether an action or event is typical or atypical). A prominent indirect explanation of the impact of injunctive norms, the *counterfactual view* developed by Knobe and colleagues[6], predicts that statistical norms should also show the same type of effects on causal attributions. But we found a more complex pattern than this, with the results pushing against the predictions of the counterfactual view. In brief, we found that causal attributions for agents were insensitive to one type of statistical norm (what is typical or atypical for members of a relevant population to which the agent belongs); and we found that while causal attributions were sensitive to another type of statistical norm (what is typical or atypical for the agent herself) when the agent knows about the likely outcome of her action, the effect ran in the opposite direction to that predicted by the counterfactual view (causal ratings being higher when the agent acted typically than when she acted atypically).

Our predictions about the impact of statistical norms on causal attributions were derived from thinking about responsibility, and as such we predict that similar effects should be seen for responsibility attributions. In this paper I test these predictions, repeating the studies from Livengood et al. (2017) and Sytsma (under review a), but now asking participants about responsibility instead of causation. In line with our predictions, the complex pattern of effects for injunctive and statistical norms previously found for causal attributions is also found for

[6] See Hitchcock and Knobe (2009), Halpern and Hitchcock (2015), Kominsky et al. (2015), Phillips et al. (2015), Icard et al. (2017), Kominsky and Phillips (2019).

responsibility attributions. In fact, the results for these two types of attributions are not statistically significantly distinguishable across these studies.

Here is how I will proceed. In Section 1, I discuss the previous findings for violations of statistical norms in the Pen Case and extend this to responsibility attributions. In Section 2, I extend the findings for the Lauren Alone Case from Livengood et al. (2017) to responsibility attributions. The implications of these results for explanations of ordinary causal attributions is discussed in Section 3.

1. The Pen Case

Knobe and Fraser (2008) found that causal ratings for the Pen Case were significantly higher for Professor Smith, who violates an injunctive norm, than for the administrative assistant, who does not violate a norm. Each of the main accounts in the literature is able to explain this effect, including the responsibility view and the counterfactual view.[7] The responsibility view does so by arguing that violations of injunctive norms are directly relevant to the correct application of the ordinary concept of causation at play in causal attributions. The counterfactual view, by contrast, offers an indirect explanation. This view holds that norm violations make the counterfactual on which the norm-violation does not occur salient, such that people are more likely to consider it, and on that counterfactual in scenarios like the Pen Case the outcome would not occur, which leads participants to judge that the norm-violating agent is more causal. This account is not specific to injunctive norms, however, but holds that violations of any type of norm should impact causal attributions. Thus, the counterfactual view predicts that the same type

[7] In addition to these views, other prominent accounts in the literature include 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). I'll return to these views briefly in Section 3.

of effect seen for violations of injunctive norms should also be found for violations of statistical norms. This has not been borne out by the data for scenarios involving agents, however.[8]

1.1 Previous Findings

In Sytsma et al. (2012) we tested whether violations of statistical norms have the same effect on causal attributions for the Pen Case as violations of injunctive norms. We distinguished between two types of statistical norms: an agent could either act typically/atypically with regard to her own usual behavior (agent-level typicality/atypicality) or with regard to the usual behavior of a salient group that she belongs to (population-level typicality/atypicality). The impact of these two types of statistical norm were tested across five studies. We found that causal attributions were insensitive to the population-level statistical norm. And, while they were sensitive to the agent-level statistical norm, the impact went in the opposite direction to that predicted by the counterfactual view: agent-level *typicality*, not *atypicality*, corresponded with higher causal ratings for the norm-violating agent (Professor Smith). Further, these findings have recently been replicated in Study 3 in Sytsma (under review a) using slight variations on eight key conditions from the original studies.[9]

While the effects of statistical norms on causal attributions found in these studies pushes against the counterfactual view, we predicted the effects on the basis of our alternative

[8] While the cases focused on in this paper involve agents, the counterfactual and responsibility views are not specific to agents. And there is some evidence that statistical norms impact causal attributions for non-agents (Kominsky et al. 2015, Icard et al. 2017, Morris et al. 2019). Kominsky et al.'s Experiment 4 and Icard et al.'s Experiment 2 did not test causal attributions, however, but "because" statements. It is unclear whether such statements work analogously to causal attributions (Livengood and Machery 2007; Livengood et al. 2017, fn39). Morris et al. have recently reported similar findings testing causal attributions, though. More work is needed here, but it is possible that statistical norms have different effects for causal attributions concerning non-agents than for agents.

[9] I tested just the conditions from Sytsma et al. (2012) where the administrative assistant did not violate a norm, making a slight modification to the conditions where Professor Smith did not violate an injunctive norm: in the original study it was stated that the agents were *able* to take pens; in the replication it was instead stated that they were *allowed* to take pens to more clearly mark permissibility.

responsibility view. Reflecting on the responsibility attributions we would make, we predicted that violations of population-level statistical norms would have little impact on causal attributions for scenarios like the Pen Case. Our reasoning was that excuses like "everyone was doing it" aren't generally taken to notably mitigate responsibility for a bad outcome. In contrast, for agent-level statistical norms, we predicted that causal ratings would be higher when the action is typical for the agent *and* she knows that she is doing something she shouldn't and that a bad outcome might result. The idea is that when the agent could be expected to know that a bad outcome might result from her behavior, people would be more likely to see her as being responsible for the outcome when she typically acts in such a reckless way. This draws out that the responsibility view expects a number of factors beyond just injunctive norms to matter for causal attributions, including what the agent knew and what she desires, because factors like these matter for responsibility attributions (e.g., Cushman 2008, Gailey and Falk 2008, Lagnado and Channon 2008, Malle et al. 2014, Young and Saxe 2011). And there is a growing body of evidence that such factors also matter for causal attributions (Samland and Waldmann 2016, Kirfel and Lagnado 2017, Sytsma under review c).

1.2 Study 1

Our predictions about the impact of statistical norms on causal attributions for the Pen Case were arrived at through reflecting on the responsibility attributions we would make. As such, assuming our responsibility judgments are typical, we would expect to find similar effects if we were to directly test responsibility attributions. Testing this prediction for the Pen Case was the goal of my first study.

Each participant in Study 1 was given one of the eight variations on the Pen Case from Study 3 in Sytsma (under review a). In each condition, the administrative assistant did not violate a norm. In the first four conditions, I varied whether Professor Smith violated an injunctive norm (permissible, impermissible) and whether she violated a population-level statistical norm (population typical, population atypical). In the last four, I again varied whether Professor Smith violated an injunctive norm, but this time also varied whether she violated an agent-level statistical norm (agent typical, agent atypical). After reading the vignette, participants were asked to rate their agreement or disagreement with a responsibility attribution for each agent on a 7-point scale anchored at 1 with "strongly disagree," at 4 with "neutral," and at 7 with "strongly agree."[10] The attributions were presented in random order. Results were collected from 338 participants.[11]

Results are shown in Figure 1 (population-level) and Figure 2 (agent-level) with those from Sytsma (under review a) Study 3 shown for comparison. To look at the effect of term, I began by analyzing the results of Study 1 together with the results from Sytsma (under review a), treating the population-level and agent-level probes separately and running ANOVAs with *term* (responsible, cause), *injunctive norm* (no violation, violation), and *statistical norm* (no violation, violation) as between-participant factors. Term was not significant for ratings of either Professor

[10] See Sytsma (under review a) for vignettes. For the first four probes the attributions were "Professor Smith is responsible for the problem" and "The administrative assistant is responsible for the problem"; for the second four, the administrative assistant was named "John" and the attribution tested was "John is responsible for the problem."
[11] Participants for each study in this paper were recruited through advertising for a free personality test on Google. In addition to answering the questions reported below, participants were asked basic demographic questions and after the philosophical questions were given a 10-item Big Five personality inventory. In line with Sytsma (under review a), participants were restricted to native English-speakers who were 16 years of age or older. Participants for Study 1 were 61.8% women (three non-binary), average age 34.4 years, ranging from 16 to 90. Given the higher percentage of women, the four ANOVAs described below were run with gender as a third between-participants factor. No significant gender effects were found.

Smith or the administrative assistant for either set of probes.[12] For the population-level probes, statistical norm was not significant for ratings of either agent, while injunctive norm was significant for both.[13] For the agent-level probes, both statistical norm and injunctive norm were significant for ratings of Professor Smith, while just statistical norm was significant for the administrative assistant.[14] There were no significant interaction effects in any of the ANOVAs.[15] And results were similar when analyzing just the results of Study 1.[16]

In line with the predictions of the responsibility view, we find that ratings for causal attributions and responsibility attributions for the Pen Case are remarkably similar. Further, violating the injunctive norm had a notable impact on ratings for Professor Smith. In Study 1, ratings were significantly higher for Professor Smith when she violated just the injunctive norm compared to when she violated neither norm for both the population- and agent-level conditions.[17] And ratings were significantly higher for Professor Smith in each pair of conditions when she violated both norms compared to when she violated just the statistical norm.[18] Comparable results were found for causal attributions in Sytsma (under review a). Turning to the statistical norms, in line with our previous results for causal attributions, ratings were insensitive

---

[12] Professor Smith, Population-level: $F(1, 322)=1.34$, $p=0.25$, $\eta^2=0.003$. Administrative Assistant, Population-level: $F(1, 322)=0.18$, $p=0.67$, $\eta^2=0.001$. Professor Smith, Agent-level: $F(1, 327)=0.002$, $p=0.96$, $\eta^2=0.000$. Administrative Assistant, Agent-level: $F(1, 327)=2.82$, $p=0.094$, $\eta^2=0.008$.

[13] Professor Smith: statistical norm, $F(1, 322)=0.23$, $p=0.63$, $\eta^2=0.001$; injunctive norm, $F(1, 322)=68.3$, $p=3.7e^{-15}$, $\eta^2=0.17$. Administrative Assistant: statistical norm, $F(1, 322)=0.19$, $p=0.66$, $\eta^2=0.001$; injunctive norm, $F(1, 322)=4.07$, $p=0.045$, $\eta^2=0.012$.

[14] Professor Smith: statistical norm, $F(1, 327)=37.2$, $p=3.0e^{-9}$, $\eta^2=0.096$; injunctive norm, $F(1, 327)=19.6$, $p=1.3e^{-5}$, $\eta^2=0.051$. Administrative Assistant: statistical norm, $F(1, 327)=13.2$, $p=3.3e^{-4}$, $\eta^2=0.038$; injunctive norm, $F(1, 327)=1.64$, $p=0.20$, $\eta^2=0.005$.

[15] There was a borderline significant three-way interaction, however, for the Administrative Assistant in the agent-level conditions: $F(1, 327)=3.37$, $p=0.067$, $\eta^2=0.010$.

[16] Professor Smith, Population-level: injunctive norm, $F(1, 164)=34.1$, $p=2.8e\text{-}8$, $\eta^2=0.17$. Administrative Assistant, Population-level: no significant effects. Professor Smith, Agent-level: statistical norm, $F(1, 166)=30.5$, $p=1.2e\text{-}7$, $\eta^2=0.14$; injunctive norm, $F(1, 166)=13.8$, $p=2.7e\text{-}4$, $\eta^2=0.066$. Administrative Assistant, Agent-level: statistical norm, $F(1, 166)=4.34$, $p=0.039$, $\eta^2=0.025$; borderline significant interaction, $F(1, 166)=3.73$, $p=0.055$, $\eta^2=0.021$.

[17] Population-level: $t(83.986)=4.94$, $p=2.0e\text{-}6$, d=1.06, one-tailed; $W=1414.5$, $p=7.3e\text{-}6$. Agent-level: $t(81.84)=2.61$, $p=0.0054$, d=0.57, one-tailed; $W=1166$, $p=0.0047$.

[18] Population-level: $t(79.967)=3.49$, $p=4.0e^{-4}$, d=0.77, one-tailed; $W=1185.5$, $p=5.7e^{-4}$. Agent-level: $t(83.846)=3.16$, $p=0.0011$, d=0.67, one-tailed; $W=1254.5$, $p=0.0012$.

to the population-level statistical norm. And while responsibility attributions were sensitive to the agent-level statistical norm, ratings for Professor Smith were significantly *higher* for Professor Smith when she violated neither norm than when she violated just the agent-level statistical norm[19]; further, ratings were significantly *higher* for Professor Smith when she violated just the injunctive norm than when she violated both norms.[20] Again, comparable results were found for these comparisons for causal attributions in Sytsma (under review a).

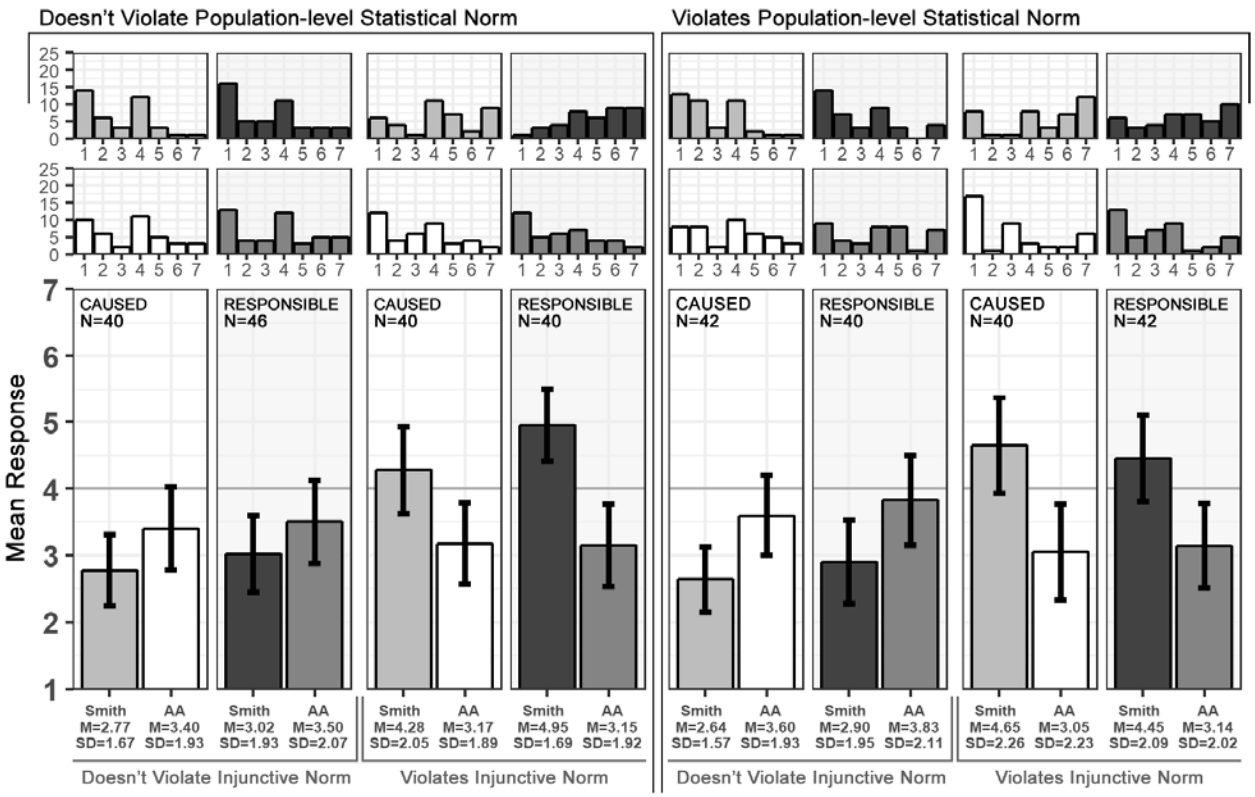

Figure 1: Results of first four conditions (population-level) for Sytsma (under review a) Study 3 and Study 1; showing 95% confidence intervals, histograms shown above.

[19] $t(81.538)=4.23$, $p=3.0e^{-5}$, d=0.92, one-tailed; $W=443.5$, $p=2.8e^{-5}$.
[20] $t(83.404)=3.66$, $p=2.2e^{-4}$, d=0.79, one-tailed; $W=520.5$, $p=2.2e^{-4}$.

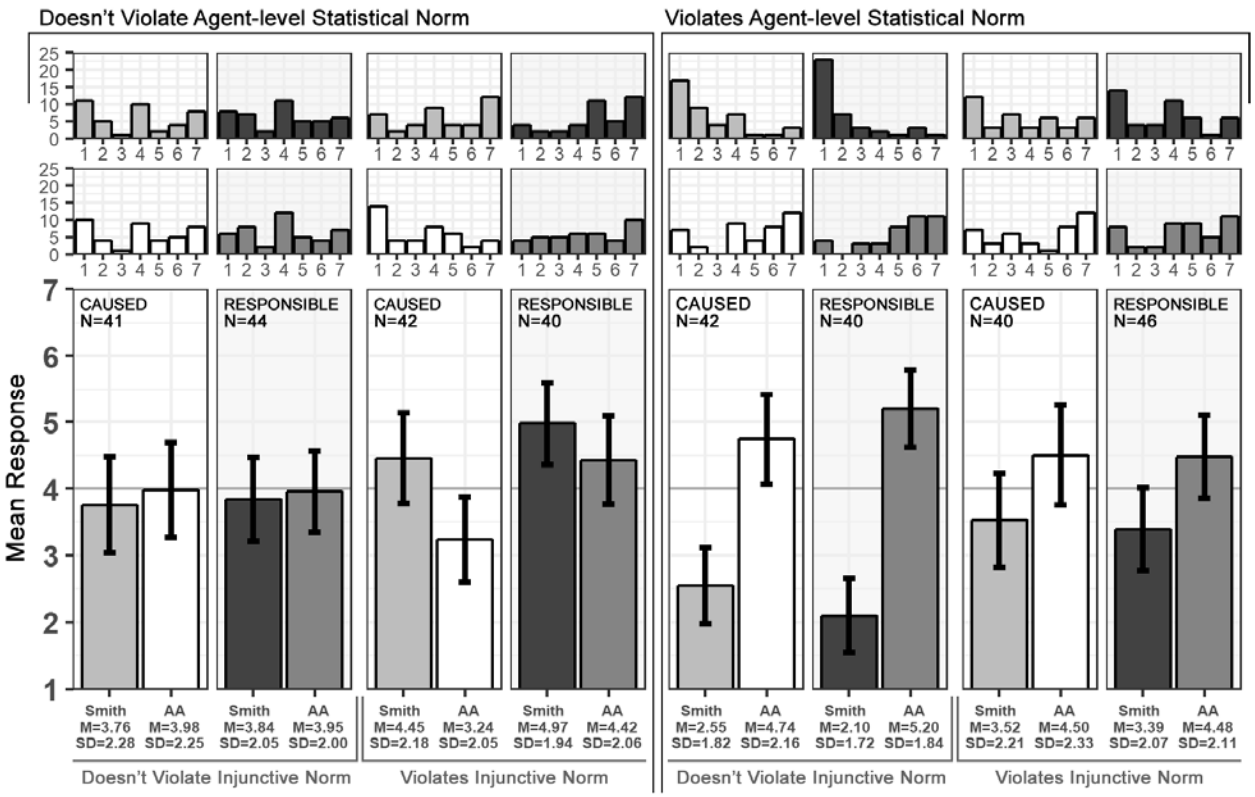

Figure 2: Results of last four conditions (agent-level) for Sytsma (under review a) Study 3 and Study 1; showing 95% confidence intervals, histograms shown above.

2. Lauren Alone Case

The standard scenarios that have been studied in the recent empirical literature on ordinary causal attributions involve two agents performing actions that are symmetric outside of one agent violating a norm. In Study 17 in Livengood et al. (2017), we diverged from this formula, testing a scenario that featured just a single agent. In this section, I extend these findings to responsibility attributions.

2.1 Previous Findings

In the Lauren Alone Case a single agent is noted and her action is described as being sufficient to bring about a bad outcome. In these cases, Lauren works for a company that uses a mainframe

computer. Unbeknownst to the company, the mainframe has recently become unstable, such that the system will crash if anyone logs into it. One day, Lauren logs into the mainframe and the system crashes. Participants were asked to rate whether they agreed or disagreed that Lauren caused the system to crash using the same 7-point scale as in Study 1. On a second page, participants were then told that the company learned about the problem with the mainframe and because of this implemented a policy prohibiting employees from logging in. Although Lauren knew about the policy, she once again logged into the mainframe and once again the system crashed. Participants were asked to rate the same causal attribution as on the first page.

We ran five variations on this scenario. In the first, no typicality information was provided.[21] In the remaining four, typicality information was provided on the first page, either describing Lauren's behavior as population typical, population atypical, agent typical, or agent atypical. We found that in each condition participants tended to deny that Lauren caused the system to crash on the first page and to affirm that she caused the system to crash on the second page. Overall, whether Lauren acted typically or atypically relative to either the population or her own behavior had no notable effect on participants' responses.

With regard to the Pen Case, in Sytsma et al. (2012) we predicted that agent typicality would have an impact on causal attributions. As noted above, however, the reasoning here involved Professor Smith knowing about the injunctive norm and the potential result of her behavior. But this is not the case for Lauren on the first page of the probes from Livengood et al. (2017): no injunctive norm is specified and there is little reason to suspect that Lauren knows

[21] The results for this condition have recently been replicated (Sytsma under review c), including that the same effects were seen using a between-participants design with participants either receiving the probe from the first page or a slightly modified stand-alone version of the probe from the second page.

about the problem with the mainframe.[22] As such, the responsibility view makes a different prediction about this case. In the absence of an injunctive norm and knowledge of the likely result of her action, we predict that causal attributions should also be insensitive to agent-level typicality. Again, our reasoning here was based on reflecting on the responsibility attributions we would make. As such, assuming our responsibility judgments are typical we would expect to find similar results when testing responsibility attributions for the Lauren Alone Case.

2.2 Study 2

Each participant in Study 2 was given one of the five variations on the Lauren Alone Case from Livengood et al. (2017). After reading the vignette on a given page, participants were asked to rate their agreement or disagreement with the responsibility attribution "Lauren is responsible for the system crashing" using the same 7-point scale as in Study 1.[23] Participants were not able to go back to first page. Results were collected from 259 participants.[24]

Results are shown in Figure 3 with those from Livengood et al. (2017) Study 17 for comparison. To look at the effect of term, I began by analyzing the results of Study 2 together with the results from Livengood et al., running ANOVAs for responses on each page with *term* (responsible, cause) and *condition* (typicality not specified, agent-level typical, agent-level atypical, population-level typical, population-level atypical) as between-participant factors. No significant effects were seen for responses on either page.[25] Corresponding one-way ANOVAs

[22] That participants don't infer that Lauren knows about the problem with the mainframe is confirmed by Studies 3 and 4 in Sytsma (under review c).
[23] See Livengood et al. (2017) for vignettes.
[24] 69.1% women (six non-binary), average age 39.5 years, ranging from 16 to 86. The ANOVAs described below were run with gender as a second between-participants factor. No significant gender effects were found.
[25] Page 1: Term, $F(1, 518)=1.25$, $p=0.26$, $\eta^2=0.002$; Condition, $F(4, 518)=1.10$, $p=0.36$, $\eta^2=0.008$. Interaction, $F(4, 518)=0.44$, $p=0.78$, $\eta^2=0.003$. Page 2: Term, $F(1, 518)=0.34$, $p=0.56$, $\eta^2=0.001$; Term*Condition, $F(4, 518)=1.79$, $p=0.13$, $\eta^2=0.013$; Term*Condition, $F(4, 518)=1.43$, $p=0.22$, $\eta^2=0.011$.

for just the results of Study 2 showed no significant effect for condition for responses on the first page, but interestingly showed an effect for responses on the second page.[26]

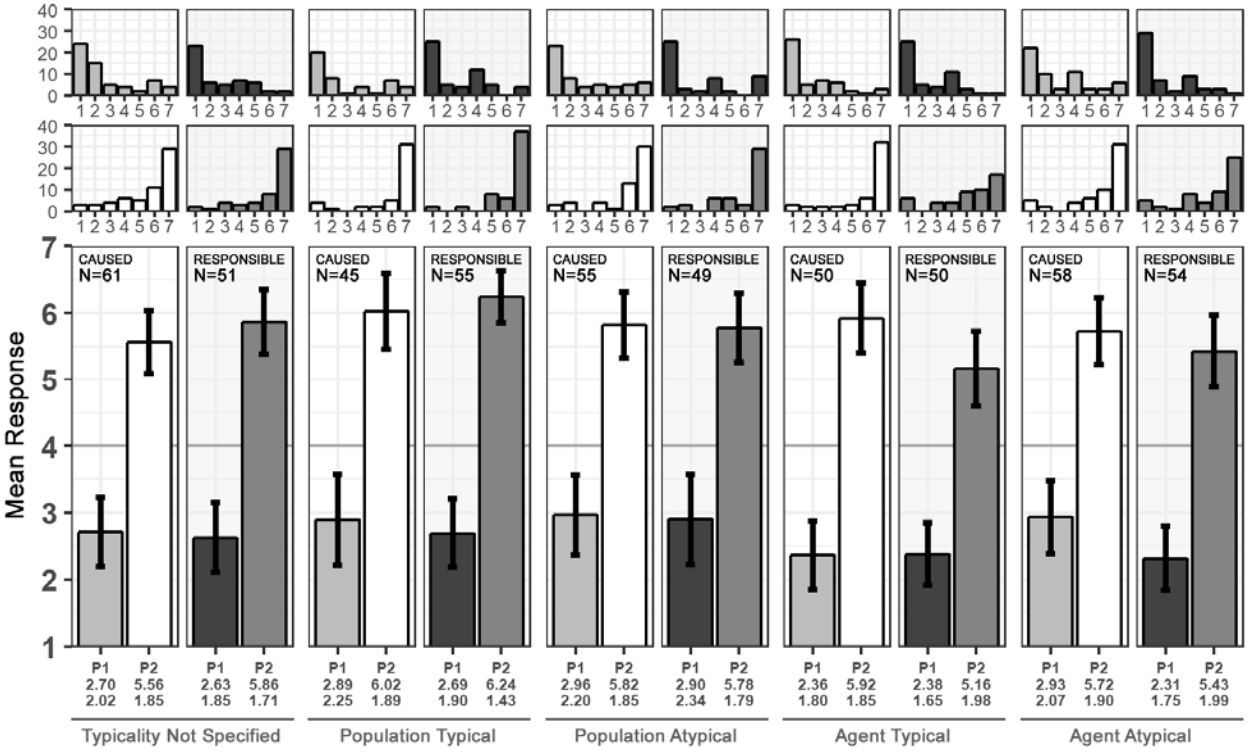

Figure 3: Results for Livengood et al. (2017) Study 17 and Study 3 with means followed by standard deviations below the bar graphs; showing 95% confidence intervals, histograms shown above.

Planned comparisons, however, indicate that the effect of condition for responses on the second page do not reflect the impact of either type of statistical norm; further, the overall pattern of effects matches that found by Livengood et al. for causal attributions. First, the mean response on the first page was significantly below the neutral point for each condition in Study 2, while the mean response on the second page was significantly above the neutral point for each

[26] Page 1: $F(1, 254)=0.79$, $p=0.53$, $\eta^2=0.012$. Page 2: $F(1, 254)=2.82$, $p=0.026$, $\eta^2=0.043$.

condition.[27] Second, comparing responses for the second and third conditions (agent-level typical versus agent-level atypical) for Study 2 showed no significant difference for either page, and similarly comparing responses for the fourth and fifth conditions (population-level typical versus population-level atypical).[28] And the same held for causal attributions in Livengood et al.'s study.[29] These findings indicate that neither statistical norm has a notable effect on participants' responses on either the first page or the second page. Finally, I analyzed the combined results for the first three conditions, with the agent-level statistical norm as an ordered factor (atypical, unspecified, typical), and similarly for population-level typicality using the first, fourth, and fifth conditions. Again, no significant effects were seen for responses on either page.[30] And corresponding one-way ANOVAs for just the results of Study 2 showed no significant effect for either agent-level typicality or population-level typicality.[31]

---

[27] Page 1, typicality not specified: $t(50)=5.29$, $p=1.4e^{-6}$, d=0.74, one-tailed; $V=131$, $p=6.8e^{-6}$. Page 2, typicality not specified: $t(50)=7.78$, $p=1.8e^{-10}$, d=1.09, one-tailed; $V=1079$, $p=9.4e^{-8}$. Page 1, agent-level typical: $t(49)=6.93$, $p=4.2e^{-9}$, d=0.98, one-tailed; $V=49$, $p=4.0e^{-7}$. Page 2, agent-level typical: $t(49)=4.14$, $p=6.9e^{-5}$, d=0.58, one-tailed; $V=843$, $p=3.8e^{-4}$. Page 1, agent-level atypical: $t(53)=7.09$, $p=1.6e^{-9}$, d=0.96, one-tailed; $V=71$, $p=8.3e^{-8}$. Page 2, agent-level atypical: $t(53)=5.27$, $p=1.3e^{-6}$, d=0.72, one-tailed; $V=898.5$, $p=2.5e^{-5}$. Page 1, population-level typical: $t(54)=5.10$, $p=2.3e^{-6}$, d=0.69, one-tailed; $V=141$, $p=1.6e^{-5}$. Page 2, population-level typical: $t(54)=11.6$, $p<2.2e^{-16}$, d=1.57, one-tailed; $V=1457$, $p=8.5e^{-10}$. Page 1, population-level atypical: $t(48)=3.30$, $p=9.2e^{-4}$, d=0.47, one-tailed; $V=225.5$, $p=0.0022$. Page 2, population-level atypical: $t(48)=6.93$, $p=4.8e^{-9}$, d=0.99, one-tailed; $V=861.5$, $p=4.4e^{-7}$.
[28] Page 1, agent-level typical vs. agent-level atypical: $t(101.95)=0.20$, $p=0.85$, d=0.038, two-tailed; $W=1391$, $p=0.78$. Page 2, agent-level typical vs. agent-level atypical: $t(101.42)=0.68$, $p=0.50$, d=0.13, two-tailed; $W=1205.5$, $p=0.33$. Page 1, population-level typical vs. population-level atypical: $t(92.697)=0.49$, $p=0.62$, d=0.098, two-tailed; $W=1342$, $p=0.97$. Page 2, population-level typical vs. population-level atypical: $t(91.502)=1.44$, $p=0.15$, d=0.29, two-tailed; $W=1511.5$, $p=0.22$.
[29] Page 1, agent-level typical vs. agent-level atypical: $t(105.98)=1.53$, $p=0.13$, d=0.29, two-tailed; $W=1209$, $p=0.12$. Page 2, agent-level typical vs. agent-level atypical: $t(104.4)=0.54$, $p=0.59$, d=0.10, two-tailed; $W=1585.5$, $p=0.35$. Page 1, population-level typical vs. population-level atypical: $t(93.331)=0.17$, $p=0.87$, d=0.034, two-tailed; $W=1205.6$, $p=0.81$. Page 2, population-level typical vs. population-level atypical: $t(93.261)=0.54$, $p=0.59$, d=0.11, two-tailed; $W=1384.5$, $p=0.25$.
[30] Agent-level, Page 1: Term, $F(1, 318)=1.35$, $p=0.25$, $\eta^2=0.004$; Norm, $F(2, 318)=0.75$, $p=0.47$, $\eta^2=0.005$; Term*Norm, $F(2, 318)=0.91$, $p=0.40$, $\eta^2=0.006$. Agent-level, Page 2: Term, $F(1, 318)=1.29$, $p=0.26$, $\eta^2=0.004$; Norm, $F(2, 318)=0.18$, $p=0.84$, $\eta^2=0.001$; Term*Norm, $F(2, 318)=2.13$, $p=0.12$, $\eta^2=0.013$. Population-level, Page 1: Term, $F(1, 310)=0.21$, $p=0.64$, $\eta^2=0.001$; Norm, $F(2, 310)=0.43$, $p=0.65$, $\eta^2=0.003$; Term*Norm, $F(2, 310)=0.031$, $p=0.97$, $\eta^2=0.000$. Population-level, Page 2: Term, $F(1, 310)=0.94$, $p=0.33$, $\eta^2=0.003$; Norm, $F(2, 310)=1.68$, $p=0.19$, $\eta^2=0.011$; Term*Norm, $F(2, 310)=0.28$, $p=0.76$, $\eta^2=0.002$.
[31] Agent-level, Page 1: $F(2, 152)=0.46$, $p=0.63$, $\eta^2=0.006$. Agent-level, Page 2: $F(2, 152)=1.77$, $p=0.17$, $\eta^2=0.023$. Population-level, Page 1: $F(2, 152)=0.24$, $p=0.79$, $\eta^2=0.003$. Population-level, Page 2: $F(2, 152)=1.17$, $p=0.31$, $\eta^2=0.015$.

3. General Discussion

The results from Sytsma et al. (2012), Livengood et al. (2017), and Sytsma (under review a) suggest that while causal attributions are sensitive to violations of injunctive norms, they are insensitive to violations of population-level statistical norms, and they are insensitive to violations of agent-level statistical norms when the agent isn't prohibited from performing the action and doesn't know the likely outcome of her behavior. When the agent violates an injunctive norm and knows the likely result of her action, however, ordinary causal attributions are now sensitive to agent-level statistical norms, with causal ratings being *higher* when the agent acts *typically* than when she acts *atypically*. These results for statistical norms run counter to the predictions of the counterfactual view but were predicted on the basis of our alternative responsibility view. In the previous two sections, I tested the basis for our predictions, looking at responsibility attributions for the Pen Case and the Lauren Alone Case. As predicted, the complex pattern of results previously found for causal attributions was also found for responsibility attributions. In fact, responsibility attributions were not statistically significantly distinguishable from causal attributions. This adds to a growing body of evidence showing a striking similarity between causal attributions and responsibility attributions (Sytsma and Livengood under review; Sytsma under review a, b).

The close correspondence found between causal attributions and responsibility attributions calls out for explanation. While it could be that two separate mechanisms are at play, one generating the pattern of effects for causal attributions and a separate one generating the pattern of effects for responsibility attributions, such a coincidence does not seem overly plausible, and absent evidence to the contrary we should presume that there is a deeper connection. That is, we should favor a common explanation of these effects. In fact, Knobe has

argued for just such a conclusion with regard to the impact of broadly moral evaluations on causal attributions and other judgments, such as free action and intentionality (e.g., Knobe 2010, forthcoming; Phillips et al. 2015). For instance, in a recent paper he notes that research on people's judgments about these issues "has revealed something surprising," that "people's moral judgments appear to influence their judgments about all of these seemingly non-moral questions" (forthcoming, 1). Knobe then argues that while the effect of moral judgments in each of these areas might be due to unrelated processes, "given the obvious similarities between them, it is certainly tempting to seek a unified account" (6-7). I agree. And this reasoning holds especially strongly for the results for causation and responsibility, where we don't simply get similar effects, but statistically indistinguishable results for the exact same scenarios.

It might be argued that there is a difference between the findings for responsibility, on the one hand, and causation, free action, and intentionality on the other; specifically, that while the impact of broadly moral evaluations on the former is hardly surprising, the impact on the latter is quite surprising. That responsibility attributions are sensitive to injunctive norms is not surprising, nor following the reasoning laid out above, are the findings with regard to statistical norms: we would expect responsibility judgments to be sensitive to considerations concerning what people should or shouldn't do.[32] In contrast, that injunctive norms have a corresponding impact on causal attributions has been taken to be surprising. For instance, after noting that it is not surprising that people's causal judgments would have an impact on their moral judgments, Knobe and Fraser (2008, 441, italics in original) argue that it is "more surprising… that the relationship can sometimes go *in the opposite direction*" with our moral judgments sometimes impacting our causal judgments.

[32] While philosophers often distinguish between moral responsibility and causal responsibility (e.g., Talbert 2019), the impact of injunctive norms on responsibility attributions in the scenarios tested would seem to indicate that participants were calling on a concept more like the former.

Taking the results for responsibility to be unsurprising, and the results for causation, free action, and intentionality to be surprising, it might be thought that it is just the latter that call out for a common explanation. It is unclear why our surprise as theorists would warrant such a conclusion, however. If similarity of effects recommends seeking a common account, it would seem to do so even if we find some of those effects unsurprising. In fact, in such cases a strategy recommends itself—we often attempt to extend our explanation of the unsurprising result, which we would seem to have a better grip on, to explaining the surprising results. This is the type of strategy employed by the responsibility view: we hold that causal attributions are sensitive to injunctive norms for the same general reason that responsibility attributions are—i.e., that the ordinary concepts at play in these attributions are normative concepts.

While we have not (yet) explicitly extended the responsibility view to judgments about free action or intentionality, it seems that a similar strategy could be applied. It might be that the concept of intentionality, for instance, is also best thought of as being a normative concept. In fact, some accounts in the literature on the effect of broadly moral considerations on judgments such as intentionality are similar to our view for causal attributions, such as Hindriks's (2008, 2014) normative reasons explanation.[33] While the same basic type of strategy applies here, it is likely that the relationship between these concepts is more complex. For example, Hindriks (2014, 56) plausibly suggests that intentionality judgments play a role in our responsibility

[33] Hindriks explains the effect for intentionality judgments in terms of a misalignment between what should motivate the agent and the agent's actual reasons for acting, and he argues that the same holds for responsibility attributions. He then takes his explanation to provide reason to reject the call for a common explanation of the effects for intentionality and other folk-psychological concepts, on the one hand, and causal attributions on the other. The reason Hindriks offers is that "causing is a more objective notion deciding and acting intentionally," such that "it would be implausible if the attitudes of the acting agent bore systematically on what she caused" (2014, 67). Interestingly, Sauer (2014) arrives at the opposite conclusion for basically the same reason. Accepting the call for a common explanation, he holds that we should reject accounts that would not apply to causal attributions, assuming for instance that "whether or not an agent had a certain type of causal impact on the unfolding events does not depend on the described agents' beliefs" (491). There is evidence, however, that whether or not the agent knows that the outcome will occur if she acts impacts causal attributions (e.g., Sytsma under review c).

attributions, hypothesizing that intentionality judgments facilitate attributions of responsibility. Further, there is evidence that intentionality judgments emerge quickly and impact the subsequent formation of blame judgments (Monroe and Malle 2017).

Insofar as the broad type of account we offer can be extended to related effects in the literature, offering a common explanation does not favor the counterfactual view over the responsibility view. Knobe (forthcoming, 1) writes:

> Why might people's judgments about these apparently non-moral questions be influenced by moral considerations? A variety of different hypotheses immediately suggest themselves. Perhaps the effect can be explained in terms of people's emotional reactions, or perhaps it can be explained in terms of motivated cognition, or in terms of conversational pragmatics.

Extending the responsibility view, our proposed explanation is more straightforward than these and more straightforward than Knobe's preferred explanation in terms of thinking about alternative possibilities: the basic explanation is that appearances are sometimes deceiving and that these questions are not in fact generally non-moral, even if many philosophers have assumed that they are. And such an explanation would seem applicable not just to causation, but also to free action and intentionality.

3.1 Explanations

Focusing on causal attributions and responsibility attributions, I contend that we should favor a common explanation for both, and that for such an explanation we should favor one that extends the unsurprising findings to the surprising findings. The responsibility view offers such an explanation. Other options are available, however. One option would be to explain the impact of norms on both causal attributions and responsibility attributions indirectly in terms of another factor instead of directly in terms of broadly moral evaluations. Focusing on the counterfactual view, advocates

might extend their account to responsibility attributions, arguing that broadly moral evaluations impact each via their effect on the alternative possibilities we consider. Alternatively, one might reject the call for a common explanation here, arguing that different mechanisms explain the impact of broadly moral evaluations on causal attributions and responsibility attributions. For instance, advocates of the counterfactual view might argue that broadly moral considerations impact the alternative possibilities we consider, which then impact our causal attributions, and that our causal attributions in turn impact our responsibility attributions. Finally, it might be argued that broadly moral evaluations directly impact responsibility attributions, while indirectly impacting causal attributions via alternative possibilities. These four simple models are pictured in Figure 4.

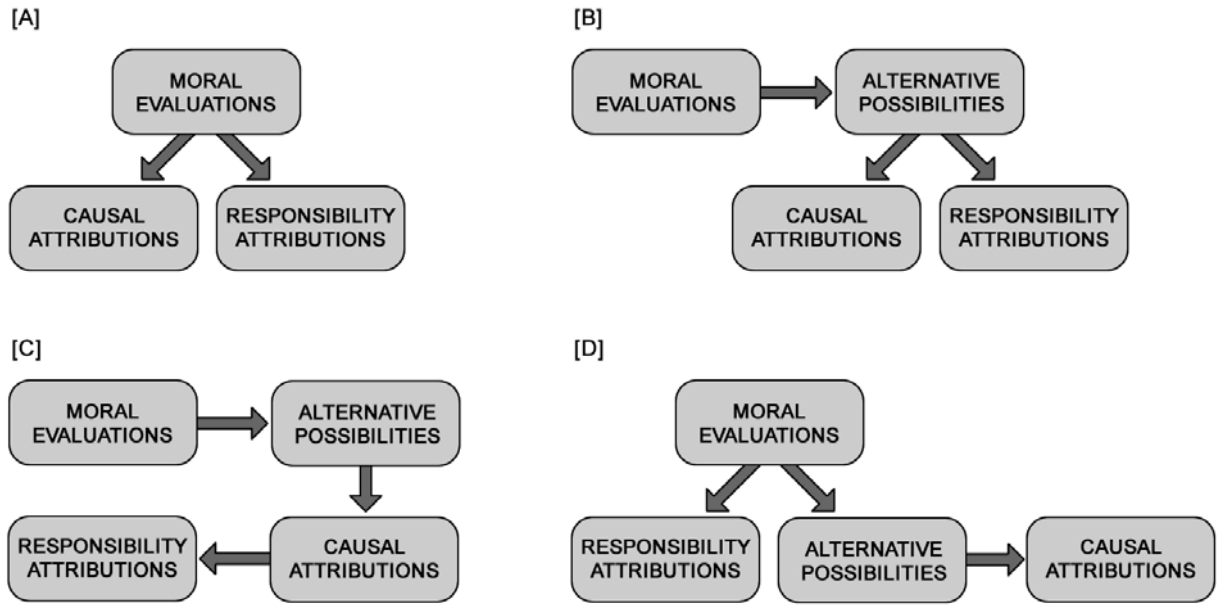

Figure 4: Four simple models for the impact of broadly moral evaluations on causal attributions and responsibility attributions.

3.1 Common Explanation

The first two models in Figure 4 offer a common explanation of the effect of broadly moral evaluations on causal attributions and responsibility attributions. Model **[A]** corresponds with the responsibility view: we explain the close correspondence between causal attributions and responsibility attributions in terms of both being influenced by evaluations of broadly moral evaluations. The idea is that people tend to morally evaluate a situation, and such evaluations are then called on in making a range of attributions, including both causal attributions and responsibility attributions. As discussed in Sytsma (under review a), the responsibility view does not deny that counterfactual thinking plays a role in people's judgments but does deny that the impact of broadly moral evaluations on causal attributions (and responsibility attributions) works primarily through counterfactual salience.

One tempting possibility is that the alternative possibilities we consider play a role in our broadly moral evaluations and that such evaluations then play a role in both causal attributions and responsibility attributions. In these discussions, that people arrive directly at the intended broadly moral judgments is often taken for granted, with the task then being to explain "the surprising ways that people's moral judgments shape their non-moral cognition" (Phillips et al. 2015, 39). For instance, in a fascinating series of studies conducted by Phillips et al. (2015), they tested both relevance judgments for alternative possibilities and the effect of explicitly asking participants to consider alternative possibilities for a range of scenarios, including the Pen Case. While they note that "at the core of [their] account is a claim about the impact of moral judgments on intuitions about the relevance of alternative possibilities" (40), they do not actually test people's moral judgments, however, but assume that they follow directly from the materials presented. Thus, it is an open question whether their results reflect the impact of people's

broadly moral evaluations on the relevance of alternative possibilities or whether the relevance of alternative possibilities instead impacts their broadly moral evaluations.

In fact, in a recent paper, Phillips et al. (2019, 1026) argue that our capacity for representing and reasoning about unrealized possible actions is "central to the most impressive of human abilities: causal reasoning, planning, linguistic communication, moral judgment, etc.," with moral judgment including responsibility attributions. They note, for instance, that "to assign responsibility for a tragedy, we consider what might have been done to prevent it" (1026). And there is a good deal of evidence that "imagined alternatives to reality have widespread effects on moral judgment" (Byrne and Timmons 2018, 90). One option here is that thinking about alternative possibilities plays an important role in our broadly moral evaluations of a situation, and that these evaluations then impact both the causal attributions and the responsibility attributions that we tend to make. Phillips et al. suggest a different model, however, holding that the relevance of alternative possibilities separately influences causal judgments and moral judgments (in line with Model **[B]**).

Model **[B]** offers a common explanation of the impact of broadly moral evaluations on causal attributions and responsibility attributions that is compatible with the counterfactual view. In this model, the impact of broadly moral evaluations works primarily through the counterfactuals that we tend to consider, with counterfactuals on which something more normal happens being more salient. This model suffers from the general issues raised for the counterfactual view noted above—the counterfactual view's predictions about the effects of statistical norms on causal attributions for agents runs counter to the evidence provided by Sytsma et al. (2012), Livengood et al. (2017), and Sytsma (under review a).[34] But now the

[34] Sytsma (under review a), Sytsma (under review b), and Sytsma and Livengood (under review) raise further issue for the counterfactual view.

counterfactual view would have corresponding difficulties explaining the effects of statistical norms on responsibility attributions for agents. Further, the model faces the difficulty suggested above: it seems strange to extend the proposed explanation for the surprising results to the non-surprising results. Here this seems especially problematic as it amounts to arguing that broadly moral evaluations don't primarily impact responsibility attributions directly but work through the more general mechanism of counterfactual salience. The concern is that the direct explanation seems more natural for a concept that advocates of the counterfactual view generally take to be normative, such that to instead call on an indirect mechanism appears unmotivated.

3.2 Separate Explanations

The second two models in Figure 3 offer different explanations for the effect of broadly moral evaluations on causal attributions versus responsibility attributions. Model **[C]** does this by holding that responsibility attributions are impacted by causal attributions. It is often held by philosophers that responsibility in part depends on causation—that you can't be responsible for something that you didn't cause.[35] And this could be extended to moral judgments more generally. In fact, Knobe and Fraser (2008, 441) open their chapter with such a claim. Further, they in part respond to Driver (2008), who opens by noting that "most people think that someone is morally responsible for an event only when that person has caused the event" (423). She goes on to note that the reverse is not the case, however—that causing an event is not generally taken to entail moral responsibility for that event.

Accepting this claim, it raises a worry for the present model, since causation being a necessary but not sufficient condition for responsibility would not produce a general close

[35] But see Sosa (1993), Leslie (1991), and Sartorio (2004), among others, for dissenting views; see Driver (2008) for a limited defense.

correspondence between the two. It could, of course, be argued that the cases examined so far are such that responsibility will be ascribed if causation is—that whatever further conditions are needed for responsibility happen to hold in these cases, at least in those conditions where causation is ascribed. But what would those further conditions be if they don't consist in the broadly moral considerations that we are presently concerned with? And model **[C]** posits that the impact of broadly moral considerations on responsibility attributions works through causal judgments, making it unclear why any divergences would be expected. This draws out the same worry as seen for model **[B]**: it seems that intuitively we want to draw a direct connection between broadly moral evaluations and responsibility, not have this work through an intermediary like counterfactual salience. One possible solution is to make the model more complicated, adding that broadly moral evaluations impact responsibility attributions directly, in addition to impacting them indirectly through causal attributions. It is unclear what would distinguish this from model **[D]**, however, at least insofar as the broadly moral considerations have generally the same impact on responsibility attributions as causal attributions.

Model **[D]** posits that broadly moral evaluations directly impact responsibility attributions and that they indirectly impact causal attributions through thinking about alternative possibilities. This model is able to uphold the intuition that broadly moral evaluations are directly relevant to assessing responsibility, while explaining the correspondence between responsibility attributions and causal attributions through their common root cause. To make this model materially different from **[B]** it would need to posit that responsibility and causal attributions will sometimes pull apart and do so in ways that can be explained by broadly moral evaluations having a different impact when working through the alternative possibilities we consider. It is currently unclear what such situations might be, however, and in the absence of

evidence that this occurs the present evidence would seem to favor the simpler common explanation.

3.3 Bias and Pragmatics

While I have focused on the counterfactual and responsibility views in this paper, these are not the only explanations of the impact of broadly moral considerations on causal attributions in the literature. Other notable accounts include Alicke's *bias view* and Samland and Waldmann's *pragmatic view*. Both views hold that the impact of injunctive norms on causal attributions reflects that participants are committing an error. The bias view holds that participants' judgments tend to be biased by their desire to blame (or praise) the agent, while the pragmatic view holds that participants tend to interpret the causal attributions in the experiments at issue as instead asking about responsibility or accountability.

Both of these views would predict some correspondence between causal attributions and responsibility attributions, taking the bias or pragmatic factors to lead participants to rate causal attributions in a similar way to how they would rate responsibility attributions. For the cases at issue, however, we don't just see *some* correspondence between causal attributions and responsibility attributions, but that they are statistically indistinguishable. Explaining this *close* correspondence on either the bias or pragmatic views would require positing an extremely strong and general error. In effect, to explain the present results these views would need to contend that basically all we are seeing here is error. But we would want very strong evidence indeed to accept such a conclusion, and currently there is no such evidence that I am aware of favoring one or the other of these views over the more charitable responsibility view. In fact, the present

evidence otherwise supports the responsibility view over the bias view (Sytsma under review c) and the pragmatic view (Sytsma et al. 2019).

4. Conclusion

Knobe (2010, 320) notes that "people's ordinary application of a variety of different concepts can be influenced by moral considerations," including their causal attributions. To this list we should add, perhaps unsurprisingly, people's responsibility attributions. The evidence presented in this paper indicates that not only are responsibility attributions impacted by broadly moral considerations, but they are strikingly similar to causal attributions for the scenarios tested. This includes that they show the same pattern of effects for statistical norms previously found for causal attributions—effects that run counter to the counterfactual view developed by Knobe and colleagues. Thus, insofar as we have reason to posit a common explanation for the effects of broadly moral considerations on the application of concepts like causation and intentionality, as Knobe contends, we have reason to extend this to responsibility as well.

In attempting to explain the effect of broadly moral considerations on our judgments, Knobe notes two basic types of approaches we might take:

> One approach would be to suggest that moral considerations actually figure in the competencies people use to understand the world. The other would be to adopt what I will call an *alternative explanation*. That is, one could suggest that moral considerations play no role at all in the relevant competencies, but that certain additional factors are somehow 'biasing' or 'distorting' people's cognitive processes and thereby allowing their intuitions to be affected by moral judgments. (320)

Our responsibility view offers an explanation of the first type, holding that broadly moral evaluations are directly relevant to the application of the ordinary concepts of causation and responsibility. And while we have focused on these two types of judgments, our account might be extended to others that are sensitive to broadly moral considerations. The underlying

motivation for the responsibility view is in line with the reasons Knobe offers for preferring the first type of approach—that we are (broadly) moralizing creatures through and through. Our account follows directly from this, holding that if this is the case then it should not be at all surprising that a host of ordinary concepts in part track our broadly moral concerns.

In contrast, the common explanation that Knobe has developed might be thought to fall short of this initial motivation. He has argued that a range of judgments, including causal judgments, are impacted by the alternative possibilities that we consider, with those possibilities in turn being impacted by normative considerations. Thus, Knobe distances the relevant judgments from the broadly moral considerations in two ways. First, he posits that it is norms generally, not just injunctive norms, that matter. Second, he posits that norms primarily influence the relevant judgments indirectly. We have seen that there are problems for the first posit with regard to causal attributions. Focusing on injunctive norms, however, the second posit is plausible insofar as we take the impact of broadly moral considerations on the relevant judgments to be surprising (as has been common with regard to causal attributions), but it is less plausible when broadly moral considerations seem directly relevant to our judgments (as they do for responsibility attributions). As such, accepting that a common explanation is to be preferred, I have argued that we should favor extending the natural explanation of the unsurprising results to the surprising ones. We should set aside our theoreticians' surprise that ordinary causal attributions would be used in a way that marks our broadly moral evaluations and apply the most direct explanation—that causal attributions are akin to responsibility attributions in having a normative component.

References

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

Alicke, M. (2000). "Culpable Control and the Psychology of Blame." *Psychological Bulletin*, 126(4): 556–574.

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

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

Byrne, R. and S. Timmons (2018). "Moral Hindsight for Good Actions and the Effects of Imagined Alternatives to Reality." Cognition, 178: 82–91.

Cushman, F. (2008). "Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment." *Cognition*, 108: 353–380.

Driver, J. (2008). "Attributions of Causation and Moral Responsibility." In W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality*, pp. 423–439, Cambridge: MIT Press.

Gailey, J. and R. 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.

Hindriks, F. (2008). "Intentional Action and the Praise-Blame Asymmetry." *Philosophical Quarterly*, 58(233): 640–641.

Hindriks, F. (2014). "Normativity in Action: How to Explain the Knobe Effect and its Relatives." *Mind & Language*, 29(1): 51–72.

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.

Kirfel, L. and D. Lagnado (2017). "'Oops, I did it Again': The Impact of Frequency on Causal Judgements." *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Knobe, J. (2010). “Person as Scientist, Person as Moralist.” Behavioral and Brain Sciences, 33: 315–365.

Knobe, J. (forthcoming). “Morality and Possibility.” In J. Doris and M. Vargas (eds.), *The Oxford Handbook of Moral Psychology*. Oxford: Oxford University Press.

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

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.

Lagnado, D. and S. Channon (2008). “Judgments of Cause and Blame: The Effects of Intentionality and Foreseeability.” *Cognition*, 108: 754–770.

Leslie, J. (1991). “Ensuring Two Bird Deaths with One Throw.” *Mind*, 100: 73–86.

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., 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.

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

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

Monroe, A. and B. Malle (2017). “Two Paths to Blame: Intentionality Directs Moral Information Processing Along Two Distinct Tracks.” *Journal of Experimental Psychology: General*, 146(1): 123–133.

Morris, A., J. Phillips, T. Gerstenberg, and F. Cushman (2019). “Quantitative Causal Selection Patterns in Token Causation.” *PLOS One*, 14(8): e0219704.

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.

Phillips, J., A. Morris, and F. Cushman (2019). “How We Know What Not To Think.” *Trends in Cognitive Sciences*, 23(12): 1026–1040.

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., 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.

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

Sartorio, C. (2004). "How to be Responsible for Something without Causing It." *Philosophical Perspectives*, 18: 315–336.
Sauer, H. (2014). "It's the Knobe Effect, Stupid! How (and How Not) to Explain the Side-Effect Effect." *Review of Philosophy and Psychology*, 5:x485–503.

Sosa, D. (1993). "Consequences of Consequentialism." *Mind*, 102: 101–122.

Sripada, C. and S. Stich (2007). "A Framework for the Psychology of Norms." In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind Vol 2: Culture and Cognition*, pp. 280–301, New York: Oxford University Press.

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

Sytsma, J. (under review b). "The Effects of Single versus Joint Evaluations on Causal Attributions." http://philsci-archive.pitt.edu/16678/

Sytsma, J. (under review c). "The Character of Causation: Investigating the Impact of Character, Knowledge, and Desire on Causal Attributions." http://philsci-archive.pitt.edu/16739/

Sytsma, J. and J. Livengood (under review). "Causal Attributions and the Trolley Problem." http://philsci-archive.pitt.edu/16200/

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., 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.

Talbert, M. (2019). “Moral Responsibility.” In E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition).

Young, L. and R. Saxe (2011). “When Ignorance is No Excuse: Different Roles for Intent Across Moral Domains.” *Cognition*, 120: 202–214.