

# The Responsibility Account<sup>1</sup>

Justin Sytsma

**Abstract:** There is now a great deal of evidence that norm violations impact people’s causal judgments. But it remains contentious how best to explain these findings. This includes that the primary explanations on offer differ with regard to how broad they take the phenomenon to be. In this chapter, I detail how the explanations diverge with respect to the expected scope of the contexts in which the effect arises, the types of judgments at issue, and the range of norms involved. In doing so, I briefly summarize the evidence favoring my preferred explanation—the responsibility account. I then add to the evidence, presenting the results of two preregistered studies that employ a novel method: participants were asked to rank order compound statements combining a causal attribution and a normative attribution.

In this chapter, I lay out and defend one type of explanation of recent findings that norms impact people’s causal judgments—the *responsibility account* (e.g., Sytsma et al. 2012, Livengood et al. 2017, Sytsma et al. 2019, Livengood and Sytsma 2020, Sytsma forthcoming-a, Sytsma and Livengood forthcoming). To do so I will critically contrast this view with three other prominent types of explanation that have been put forward in the literature—the *pragmatic account* (e.g., Samland and Waldmann 2016, Samland et al. 2016), the *bias account* (e.g., Alicke 1992, 2000; Alicke et al. 2011; Rose 2017), and *counterfactual accounts* (e.g., Hitchcock and Knobe 2009, Halpern and Hitchcock 2015, Kominsky et al. 2015, Icard et al. 2017, Kominsky and Phillips 2019). The contrast is not straightforward, however, because while each of these views purports to explain the same basic findings, they arguably diverge in how they construe the broader phenomenon to be explained, and this is tied to the type of explanation offered.

---

<sup>1</sup> Forthcoming in *Advances in Experimental Philosophy of Causation*, P. Willemsen and A. Wiegmann (eds.), Bloomsbury. I want to thank Pascale Willemsen and an anonymous reviewer for helpful comments on a previous draft of this chapter and Jonathan Livengood for suggestions in the early stage of this project.

To draw this out, consider the Pen Case first presented by Knobe and Fraser (2008), which has been widely offered as an illustrative example of the phenomenon in the subsequent literature.

Participants were given the following vignette:

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.

In this scenario, two agents perform symmetric actions (both Professor Smith and the administrative assistant take pens) that jointly lead to a problem. The key difference is that one agent violates a norm (Professor Smith is not allowed to take pens) while the other does not (administrative assistants are allowed to take pens). After reading this vignette, participants were then asked to rate their agreement with two statements using a scale ranging from -3 ("not at all") to 3 ("fully"):

Professor Smith caused the problem.

The administrative assistant caused the problem.

Knobe and Fraser found that overall participants agreed far more strongly with the first claim ( $M=2.2$ ) than with the second ( $M=-1.2$ ). Call this the *norm effect*.

Each of the accounts in the literature purports to explain the norm effect. But what exactly is the scope of this effect? Focusing on the Pen Case, we can offer a minimal characterization of the basic phenomenon. First, Knobe and Fraser report the results of an empirical study that limited the "conversation" between experimenter and participant in notable ways, including that the participant's contribution to the conversation was restricted to rating the two statements. Minimally,

the norm effect can be construed as being specific to such contexts, reflecting pragmatic factors owing to the experimental setup. Second, the statements that participants rated used the lemma “cause,” having the form “X caused Y.” Minimally, the norm effect can be construed as being specific to attributions employing this lemma. Third, in the Pen Case vignette the norm that Professor Smith violates is an explicit departmental rule: she takes a pen even though she was not allowed to do so. Minimally, the norm effect can be construed as being specific to norm violations of this type.<sup>2</sup>

The main types of explanation offered in the literature differ along one or more of these three dimensions, and so differ in how broad they ultimately take the norm effect to be. In this chapter, I detail these differences through the lens of my preferred explanation—the responsibility account. I begin by characterizing the phenomenon at issue, noting disagreements about the scope of the norm effect, focusing on how restricted the context for the effect is taken to be (Section 1), the scope of the judgments at issues (Section 2), and the range of norms involved (Section 3). In doing so, I’ll lay out the primary alternative explanations of the norm effect that have been offered and very briefly summarize some of the recent evidence suggesting in favor of the responsibility account. In Section 4, I then add to this, reporting the results of two new, preregistered studies using a method not previously employed in these debates.

---

<sup>2</sup> Further distinctions can be drawn, although I’ll focus on these three. Most notably, focusing on the form “X caused Y,” in the Pen Case the X’s are agents and the Y is a bad outcome. Work has varied both of these factors, however. With regard to X, the effect has also been found for non-agents (e.g., Hitchcock and Knobe 2009, Sytsma forthcoming-b) and for statements involving an agent’s action or decision (e.g., Hitchcock and Knobe 2009, Livengood et al. 2017). And with regard to Y, while the effect has been reported for good outcomes (e.g., Hitchcock and Knobe 2009, Kominsky et al. 2015), it has also been shown that it can be reversed with ratings being higher for the norm-conforming agent than the norm-violating agent (Schwenkler and Sytsma ms).

## 1. Context

Most of the explanations of the impact of norms on causal judgments that have been offered in the literature take this to be a general phenomenon that is revealed by the empirical studies, not an artifact created by the context of the studies themselves. The pragmatic account put forward by Samland and Waldmann challenges this assumption. Samland and Waldmann (2016) claim that the term “cause” can be understood in two different ways, one corresponding with what they refer to as *causality* and the other with *accountability*. While they offer a rather circular definition of “causality” that would require a good bit of unpacking—i.e., “causality in the narrow sense refers to contingent dependency relations between causes and effects that are generated by causal mechanisms” (165)—the key distinction here is that it is taken to be purely descriptive, while accountability is taken to be sensitive to normative considerations. Samland and Waldmann then argue that in studies showing the norm effect, “pragmatic contextual features steer subjects toward an accountability understanding of the causal test question” (165).

The most straightforward interpretation of Samland and Waldmann’s view is that the dominant attributional use of the lemma “cause” expresses a purely descriptive concept (causality), but that in contexts like that found in Knobe and Fraser’s study using the Pen Case, pragmatic factors lead participants to instead interpret the queries as asking for a judgment that is sensitive to normative considerations. Samland and Waldmann write that “the ambiguity of queries about the cause in scenarios demonstrating norm effects is grounded in the presupposition relation between accountability and causation,” which is such that “agents are only held accountable for outcomes they have caused” (165). They then note that in the Pen Case “the causal relations are trivial,” such that “it is unlikely that subjects will think that they are supposed to solely judge this causal relation” (165). What this suggests is not so much that Samland and Waldmann hold that the term

“caused” is polysemous, but that in rather specific contexts they expect people to infer that the experimenters were actually interested in another concept altogether—one that presupposes relevant descriptive relations, but where information about norm violations is also relevant. The upshot is that according to the pragmatic account, studies showing the impact of norms on causal judgments miss their intended mark: they don’t actually tell us anything about ordinary causal judgments.

Our responsibility account is in some ways similar to Samland and Waldmann’s pragmatic account. Both views explain the norm effect in terms of participants taking the test statements to express a normative concept. While the pragmatic account holds that this is due to participants taking the statements to be intending to ask about something other than who caused the outcome, however, the responsibility account denies this. We contend that participants do in fact take the statements to be asking about who caused the outcome, but believe that the dominant use of “caused” in such statements expresses a normative concept. Specifically, we hold that the concept of causation expressed by the dominant attributional use of the lemma “cause,” at least in English, is sensitive to both descriptive and normative considerations, being akin to concepts like responsibility and accountability. In this way, the responsibility account takes the norm effect to be notably broader than the pragmatic account does with regard to context: we hold that this is a general phenomenon, not one that is specific to the experimental contexts.

The responsibility account predicts that people’s judgments about causal attributions (statements like “X caused Y”) will generally be quite similar to their judgments about normative attributions such as responsibility attributions (statements like “X is responsible for Y”). And there is evidence for this claim both from corpus studies and experimental studies. First, Sytsma et al. (2019) present evidence suggesting that the dominant attributional use of “caused” as found in the Corpus of Contemporary American English is sensitive to normative information, being similar to

the attributional use of “responsible” in this regard. Unlike other terms that plausibly express descriptive “causality” (such as “created,” “induced,” “led to”), “caused” and “responsible” tend to be disproportionately used in the context of negatively valenced outcomes, and a distributional semantic analysis shows that they are remarkably close together in semantic space. Such findings for ordinary English usage suggest against the norm effect being merely a pragmatic effect.

Second, a spate of papers have found a close correspondence between judgments about causal attributions and responsibility attributions, including a number of recent papers testing the norm effect that have shown that it occurs for both types of attribution and found no statistically significant difference between judgments about each. This includes recent studies testing variations on the Pen Case discussed above (Sytsma forthcoming-a), the Email Case tested below (Sytsma forthcoming-a, ms-b; Sytsma and Schwenkler ms), the Machine Case (see Hitchcock and Knobe 2009; Sytsma forthcoming-b), and the Motion Detector Case (see Kominsky et al. 2015; Sytsma ms-a). Further, not only have responses for causal attributions and responsibility attributions been found to be statistically indistinguishable when presenting the attributions between-subjects (Sytsma and Livengood forthcoming; Sytsma forthcoming-a, forthcoming-b, ms-a), but also when presenting them within-subjects, either on alternating pages (Schwenkler and Sytsma ms) or together on the same page (Sytsma forthcoming-b, ms-b). If the norm effect is due to pragmatic factors leading participants to interpret the causal queries as intending to elicit a normative judgment rather than the dominant descriptive use, then the effect should be notably diminished when participants are also given a normative attribution: pragmatic considerations should now promote interpreting the two attributions as intending to elicit different judgments. But that is not what we find, suggesting that the norm effect is a general effect.

## 2. Causal Judgments

The second point of disagreement that requires clarification concerns the scope of the “causal judgments” to be explained. As noted above, key examples of the norm effect in the literature have employed questions using the lemma “cause,” prominently including the statements tested by Knobe and Fraser (2008), which were of the form “X caused Y.”<sup>3</sup> A central question, then, is whether the main accounts of the norm effect in the literature are principally aiming to explain the impact of norms on causal attributions or hold that this is part of a broader phenomenon that calls for explanation.

Danks et al. (2014, 255) interpret a number of authors involved in the debate as making a broad claim, and perhaps as being committed to what they term the *Ubiquity Thesis*: “Normative considerations (broadly construed) influence causal cognition (broadly construed) and are perhaps even constitutive of various cognitive processes involved in aspects of causal cognition.” Here, “causal cognition (broadly construed)” is taken to include both causal learning and causal reasoning, with causal perception and causal inference being counted as part of the former. What ties all of these processes together—what appending the adjective “causal” is meant to signify—and to what extent each of these processes warrant that adjective are complicated issues that Danks et al. do not engage with. Nonetheless, they contend that (at least some subset of) researchers in the debate are committed to the Ubiquity Thesis, and they present evidence that the norm effect isn’t found for certain “causal” inferences.

Textual support for researchers being committed to the Ubiquity Thesis is minimal, however, with Danks et al. offering just a few quotes that concern “causal judgments” or “causal

---

<sup>3</sup> There are a few exceptions to this general rule, however, such as the use of “made” in a study of children’s judgments about the Pen Case by Samland et al. (2016) and the use of “because” in the fourth study by Kominsky et al. (2015). See Livengood and Machery (2007), and Livengood et al. (2017) for evidence that “X caused Y” and “Y because X” statements at least sometimes come apart, generating different judgments.

intuitions.” But these phrases could themselves be interpreted either broadly or narrowly, depending on how one understands the intent of “causal.” More narrowly, “causal” can be understood first and foremost as reflecting the ordinary use of the lemma “cause.” In fact, while Sytsma et al. (2012) are cited as an example of the broad claim, we only ever meant to be making a narrow claim: we focus on “the use of causal language” (814)—on how likely people are to “say that an agent... caused an outcome” (815). And this is made clearer still in subsequent work where we explicitly state that we are concerned with causal attributions (e.g., Livengood et al. 2017, Sytsma et al. 2019, Livengood and Sytsma 2020).

While focusing on the narrower claim about the ordinary use of the lemma “cause” might be taken to limit the interest of the norm effect, as Danks et al. in fact suggest, there are good reasons for this narrow focus. First, the primary aim of the responsibility account is to explain the effect shown in studies like that seen above for the Pen Case, and such studies have overwhelmingly tested causal attributions. Second, we take it to be standard in philosophical discussions to assume that the ordinary concept of causation is *at least* expressed by the dominant attributional use of the lemma “cause.” For example, Paul and Hall (2003, 2) focus on “our folk-theoretical notion of ‘cause’,” while Skow (2019) writes that “the most fundamental causal locution is ‘X caused Y to Z by Ving’” (18), although he notes that the “by Ving” is grammatically optional (138). As such, if our account of the norm effect is correct—if we are correct that the ordinary concept of causation expressed by the dominant attributional use of “cause” is a normative concept—then this has notable implications for the philosophical discussions. Further, it would raise concerns about the sense in which “causal cognition (broadly construed)” is best considered *causal* if much of what is included diverges from the ordinary use of the root term.



It is unclear whether any of the participants in the present debate are committed to the Ubiquity Thesis. But this is itself problematic: it is unclear whether the Ubiquity Thesis applies because authors have often been unclear on what they take the scope of the norm effect to be. Further, even if the accounts at issue fall short of the broad claim given by the Ubiquity Thesis, there are nonetheless differences in how broad different researchers take the norm effect to be, and this corresponds with substantive differences in the types of explanations that have been offered.

The responsibility account is a narrow account in this regard, focusing on causal attributions and offering a linguistic/conceptual explanation of the phenomenon. Our account does not rule out that a similar effect might be found for other judgments, but neither does it make any specific predictions about them: it allows (of course) that the dominant use of some terms is descriptive, while the dominant use of others is normative. That said, we do expect that the range of terms typically used in a normative way might be surprising to many philosophers. And this is in accord with the underlying motivation for our responsibility account, as laid out in Sytsma (forthcoming-a): human cognition is generally quite attuned to recognizing applicable norms, detecting and responding to violations, and navigating factors that might exacerbate or mitigate those violations.

In contrast, the pragmatic, bias, and counterfactual accounts are broad accounts with regard to the judgments at issue for the norm effect. First, as seen above, the pragmatic account holds that the studies on the norm effect induce participants to interpret the queries as intending to ask a normative rather than a descriptive question. But we would expect the same pragmatic pressures to hold for other terms that plausibly express causality (as Samland and Waldmann understand it). Second, the bias and counterfactual accounts both hold that the ordinary concept of causation is normative but that an underlying mechanism leads it to be applied in a normative way, producing the norm effect. And while they differ with regard to the underlying mechanism that they posit, in

each case the mechanism would be expected to reveal itself across a wider range of judgments than just those involving “cause.”

The bias account contends that the norm effect comes about because the differential desire to blame (or praise) one of the agents leads people to exaggerate that agent’s contribution to bringing about the outcome, pulling their causal judgments concerning that agent toward their blame (or praise) judgments. In other words, Alicke and colleagues hold that the desire to blame (or praise) biases the application of the descriptive concept of causation, generating judgments that are sensitive to normative considerations. As Rose (2017, 1327) puts it, the norm effect reflects “an *error*, rooted in a motivational bias to blame those who engage in harmful or offensive actions.” Counterfactual accounts point to a different cognitive mechanism to explain the norm effect. Focusing on the account put forward by Hitchcock and Knobe (2009), they hold that norms impact which counterfactuals people find relevant, and hence which counterfactuals they are most likely to consider. They then contend that the process by which people arrive at causal judgments works through the counterfactuals that they consider and, hence, that norms impact causal judgments only indirectly.

While the recent debate has in large part centered on judgments about statements using the lemma “cause,” the explanations offered by the pragmatic, bias, and counterfactual accounts are not specific to such judgments: the mechanisms underpinning these explanations should reveal themselves in a larger class of judgments, presumably including judgments about other phrases that might be taken to express a descriptive causality relation (e.g., “created,” “induced,” “led to,” “linked to,” “because of”), and perhaps also counterfactual constructions (e.g., “if Professor Smith had not taken a pen, the problem would not have occurred”), statements involving lexical causatives (e.g., “Billy *deleted* the e-mails”), or causal inferences. There is

currently a paucity of evidence suggesting that the norm effect consistently occurs across the range of judgments these views might predict, however, and some evidence that it does not. This includes the corpus evidence noted above (Sytsma et al. 2019), mixed results for “because” statements (see Footnote 3), evidence in this volume concerning causatives (Schwenkler and Sievers forthcoming), and studies on causal inference (Danks et al. 2014).

The bias and counterfactual accounts should be expected to make predictions about a broader range of judgments because they explain the norm effect in terms of an underlying mechanism that isn’t specific to judgments involving the lemma “cause.” And these mechanisms generate a number of further predictions that have not been borne out by the present data. First, like the pragmatic account, the bias and counterfactual accounts have difficulty explaining the close correspondence between judgments about causal attributions and judgments about responsibility attributions noted in the previous section (Sytsma forthcoming-a). While the bias account would predict *some* correspondence between these judgments, to explain the extremely close correspondence found across a range of studies would require positing an implausibly strong bias. The basic worry for counterfactual accounts is a bit different: these accounts would seem to hold that there are two different mechanisms at play—the sensitivity of causal judgments to normative considerations being explained *indirectly* via the counterfactuals we consider while the sensitivity of responsibility judgments to normative considerations is explained *directly* in terms of responsibility being a normative concept—but it is implausible that two different mechanisms would produce statistically indistinguishable effects across a range of cases.

Second, the general mechanism posited by the bias account rests on people’s “desire to praise or denigrate those whose actions we applaud or deride” (Alicke et al. 2011, 670). Such desires should be aroused not just by features of agents that are relevant to appropriately

assessing their responsibility for an outcome, however, but also by “peripheral features of the event such as the actor’s or victim’s race or character” (674). And while information pertinent to assessing an agent’s general character does have an impact on causal attributions, recent studies indicate that this effect is not driven by the desire to denigrate, but by an inference to the agent’s knowledge and desires in performing the action, such that when the agent’s knowledge and desires are made clear, peripheral features of the agent no longer show an effect (Sytsma ms-c).

Third, the general mechanism posited by counterfactual accounts hinges on the overall normality of the actions described (as discussed further in the next section), but a pair of recent papers find that causal attributions sometimes diverge from what would be predicted on the basis of norms alone. First, Sytsma and Livengood (forthcoming) argue that sometimes acting in accord with the relevant norms renders the actor responsible for a negative outcome that results. We suggest that the switch version of the trolley problem is one such case: while popular sentiment holds that the agent ought to flip the switch, saving five innocents at the expense of one, it also seems that in doing so the agent shares in the responsibility for the one person’s death since that person was in no danger prior to the agent’s intervention. And what we find is that causal judgments for this case are similar to both people’s normative judgments and their responsibility judgments, with ratings for each being higher when the agent flips the switch than when the agent refrains. This is in line with the responsibility account, but runs counter to counterfactual accounts, which would expect an *inverse* relationship between normative judgments and causal judgments. Second, according to counterfactual accounts, the valence of the outcome should not matter for causal judgments—just the normality of the actions. But Schwenkler and Sytsma (ms) have found that the norm effect is also reversed when the norm-

conforming agent knowingly acts to bring about a good outcome, while the norm-violating agent acts for malicious reasons.

### 3. Norms

The third point of disagreement that requires clarification concerns the scope of the norms at issue for the norm effect. The prototypical examples from the literature, such as the Pen Case, involve agents violating explicit rules, such as when Professor Smith takes a pen in violation of department policy. This is an example of a *proscriptive norm*: a norm specifying what ought not to be done.

The complement of proscriptive norms are *prescriptive norms*: norms that specify what ought to be done. Together these can be referred to as *injunctive norms*.<sup>4</sup>

Injunctive norms are in line with standard discussions of normativity in philosophy and psychology, which have generally focused on our ability to recognize and enforce *oughts*. For instance, Darwall (2001) opens his Routledge Encyclopedia of Philosophy entry on “Normativity” by stating the issue in just this way, writing that “normative judgments concern oughts: what one ought to do, desire, believe, infer, conclude, think, feel and so on.” The implicit or explicit rules corresponding with such oughts are then standardly referred to as “norms.” As Kelly and Setman (2020) summarize in their Stanford Encyclopedia of Philosophy entry on “The Psychology of Normative Cognition”:

Norms are the social rules that mark out what is appropriate, allowed, required, or forbidden in different situations for various community members. These rules are informal in the sense that although they are sometimes represented in formal laws, such as the rule governing which side of the road to drive on, they need not be explicitly codified to effectively influence behavior.

---

<sup>4</sup> Note that the expression “prescriptive norm” is often used in the literature to refer to what I here term injunctive norms, including both prescriptive *and* proscriptive norms.

One thing to note here is that the norms at issue are not restricted to those oughts that we are likely to classify as distinctively moral. Although debate rages concerning how to classify norms (see O’Neill 2017), including how to distinguish between moral and non-moral norms, the key point for present purposes is that while normativity covers moral norms, it is not exhausted by them. And similarly for injunctive norms.

The pragmatic, bias, and responsibility accounts all operate on this standard understanding of norms where they’re tied to oughts. As suggested by Danks et al.’s focus on “normative considerations (broadly construed)” in their statement of the Ubiquity Thesis discussed above, however, some participants in the debate have embraced a more expansive notion of norms. Specifically, counterfactual accounts operate on a broad conception of normality. To illustrate, Hitchcock and Knobe (2009) distinguish between two meanings for the term “norm.” The first corresponds with injunctive norms: these are “norms that actually tell us what ought to happen under certain circumstances” and include “purely moral norms, where violating the norm would be intrinsically wrong” in addition to more general norms including legal norms and “norms arising from policies adopted by social institutions,” as well as norms of proper functioning that “apply to artifacts and biological organisms (and their components)” (597-598). Hitchcock and Knobe’s second meaning for “norm” diverges from the sense laid out above, however, instead concerning what is typical or atypical. They refer to these as *statistical norms* and write that “these norms simply capture information about the relative frequencies of certain events” (597). As such, statistical norms are descriptive, while injunctive norms are normative (in the standard sense).

Counterfactual accounts like that given by Hitchcock and Knobe hold that the type of norm doesn’t matter—that causal judgments are sensitive to both violations of injunctive norms and violations of statistical norms. In contrast, when I say that according to our responsibility account

the dominant use of causal attributions employs a *normative* concept, this is specific to injunctive norms: we hold that causal attributions are *directly* sensitive to injunctive norms. That said, our account predicts that statistical norms will sometimes also impact causal attributions, although we hold that they do so indirectly, with statistical norms sometimes impacting our judgments about injunctive norms (Sytsma et al. 2012, Livengood et al. 2017).

We expect that impact of statistical norms on causal attributions will work somewhat differently for attributions involving agents and attributions involving non-agents (see Sytsma forthcoming-a). Focusing on agents, we have presented a body of evidence indicating that violations of statistical norms do not impact causal attributions in the same way that violations of injunctive norms do for scenarios like the Pen Case (Sytsma et al. 2012, Livengood et al. 2017, Sytsma forthcoming-a). Instead we've shown a striking pattern of findings that run counter to the predictions of counterfactual accounts but were predicted on the basis of our responsibility account. First, we found that causal attributions are insensitive to one type of statistical norm—*population-level statistical norms*, which concern what is typical or atypical for members of a relevant population to which the agent belongs. Second, we found that causal attributions are sensitive to another type of statistical norm—*agent-level statistical norms*, which concern what is typical or atypical for the agent herself—but only when the norm-violating agent knows about the likely outcome of her action; and, in such cases we found that the “norm effect” runs in the opposite direction to that observed for violations of injunctive norms, with causal ratings being higher when the agent acts typically than when she acts atypically.

Our predictions about these effects were based on thinking about responsibility judgments. Reflecting on the responsibility attributions that we would make, we expected that violations of population-level statistical norms would have little to no impact on causal

attributions for scenarios like the Pen Case. We reasoned that excuses like “everyone was doing it” aren’t generally taken to be good excuses and, as such, would not mitigate an agent’s responsibility for a bad outcome. By contrast, for agent-level statistical norms we expected that causal ratings would be higher when the action is typical of the agent *and* when she knows that she is doing something she shouldn’t and that a bad outcome might result. We reasoned that when the agent can be expected to know that a bad outcome might result from her behavior, people will be more likely to see her as being responsible for the outcome when she typically acts in this reckless way. Here it would just seem to be a matter of time before the agent brings about the outcome, blocking consideration of potential mitigating factors. While these predictions were based on the responsibility judgments that we would make, I’ve recently shown that the same complex pattern of effects found for causal attributions is also found for responsibility attributions in a broader sample (Sytsma forthcoming-a). In fact, as noted above, ratings for the two types of attributions were not statistically significantly distinguishable.

Summing up, while each of the main accounts on offer in the recent literature propose to explain the norm effect, they diverge with regard to how broad they expect this phenomenon to be, and do so along at least three dimensions—diverging with regard to the contexts that will elicit the effect, the judgments involved in the effect, and the norms that will give rise to it. These differences are summarized in Table 1. But, as sketched above, these assumptions about the extent of the norm effect are not all equally supported by the present data, and the details of the accounts shaping their treatment of the scope of the effect give rise to a number of further predictions that have not been borne out by subsequent studies. Perhaps most notably, while the close correspondence between judgments about causal attributions and responsibility attributions found in recent studies is directly predicted by our responsibility account, it is tough to



convincingly reconcile with the competing accounts in the literature. To conclude this chapter, I expand on these findings, presenting the results of two new studies that further demonstrate that people tend to treat causal attributions and normative attributions as going together.

	<b>Context</b>	<b>Judgments</b>	<b>Norms</b>
<b>Pragmatic</b>	Narrow	Broad	Narrow
<b>Responsibility</b>	Broad	Narrow	Narrow
<b>Bias</b>	Broad	Broad	Narrow
<b>Counterfactual</b>	Broad	Broad	Broad

**Table 1:** Summary of the scope of the norm effect for each view along the dimensions of context, judgments, and norms.

#### 4. Rank Ordering Compound Statements

In this section I extend the previous findings showing a close correspondence between judgments about causal attributions and responsibility attributions, presenting the results of two preregistered studies ([osf.io/fpwq3/](https://osf.io/fpwq3/)) using a novel design: rather than asking participants about separate attributions, I asked them to rank order compound statements involving both a causal attribution and a normative attribution—either responsibility (Study 1a) or blame (Study 1b).

##### 4.1 Method

Each participant read the following vignette in which two agents perform actions that are symmetric, outside of one violating an injunctive norm (Billy) while the other does not (Suzy), and in so doing jointly bringing about a bad outcome:

Billy and Suzy work for a company that has a central computer.

In order to make sure that one person is always available to answer incoming phone calls, the company issued the following official policy: Billy is the only one permitted to log into the central computer in the afternoons, whereas Suzy is the only one permitted to log into the central computer in the mornings. Billy is never permitted to log into the central computer in the morning.

Unfortunately, a problem has recently developed with the computer system: if two people log into the computer in the morning, some important work e-mails will be immediately deleted.

This morning, Billy and Suzy both log into the central computer at the same time. Immediately, some important work e-mails are deleted.<sup>5</sup>

After reading the vignette, participants were given two rank-ordering questions in random order, one about Billy and one about Suzy, followed by a comprehension check (“How many people need to log into the central computer in the morning for the e-mails to be deleted?” with options of 0, 1, 2, 3, and 4 or more). All questions were presented on the same page.

The rank-ordering questions each included four compound statements involving a causal attribution and a normative attribution, with the order of the two attributions being varied in the statements. To illustrate, the first ordering of the question about Billy in Study 1a read as follows (letters are added for convenience and were not presented to participants):

Please rank the following four claims about Billy in order of how much you agree with them, with (1) being the claim you most strongly agree with and (4) being the claim you most strongly disagree with:

**[A]** Billy did **not** cause the e-mails to be deleted *and* Billy is **not** responsible for the e-mails being deleted.

**[B]** Billy did **not** cause the e-mails to be deleted *but* Billy is responsible for the e-mails being deleted.

**[C]** Billy caused the e-mails to be deleted *but* Billy is **not** responsible for the e-mails being deleted.

**[D]** Billy caused the e-mails to be deleted *and* Billy is responsible for the e-mails being deleted.

In the second ordering, the normative attributions preceded the causal attributions. For example, the second ordering for **[D]** would now read: “Billy is responsible for the e-mails being deleted

---

<sup>5</sup> The vignette is based on the computer case presented by Knobe (2006) and discussed in Sytsma et al. (2012), and is similar to the variations tested by Reuter et al. (2014), Kominsky et al. (2015), and Livengood et al. (2017), showing the expected norm effect with Likert-scale ratings of causal attributions. Further variations have been tested by Icard et al. (2017), Sytsma (forthcoming), Sytsma and Livengood (forthcoming), and Schwenkler and Sytsma (ms), among others.

and Billy caused the e-mails to be deleted.” The questions about Suzy were identical, except for “Suzy” replacing each instance of “Billy.” Similarly, the questions in Study 1b were identical to those in Study 1a, except for “to blame” replacing each instance of “responsible.” For each question, the order of the four options was randomized.

Participants for each study were recruited through advertising for a free personality test on Google with the ads being displayed in North America.<sup>6</sup> Prior to the test questions, participants answered basic demographic questions. After the test questions they took a 10-item Big Five personality inventory. Participants were restricted to native English speakers, 16 years of age or older, who indicated that they had not previously taken the survey, and who completed the three test questions. Results were collected from 200 participants who met the restrictions and passed the comprehension check (100 per study and 50 per condition).<sup>7</sup> These participants were 73.0% women (three non-binary) with an average age of 46.7 years and ranging in age from 16 to 80.

#### *4.2 Predictions*

The responsibility account predicts that there will still be a pronounced norm effect for the causal attributions in these studies, despite participants now being able to clearly register their normative judgments. As such, I predict that for the norm-violating agent, Billy, participants will tend to rank one of the two statements affirming causation (**[C]**, **[D]**) in first place and to rank

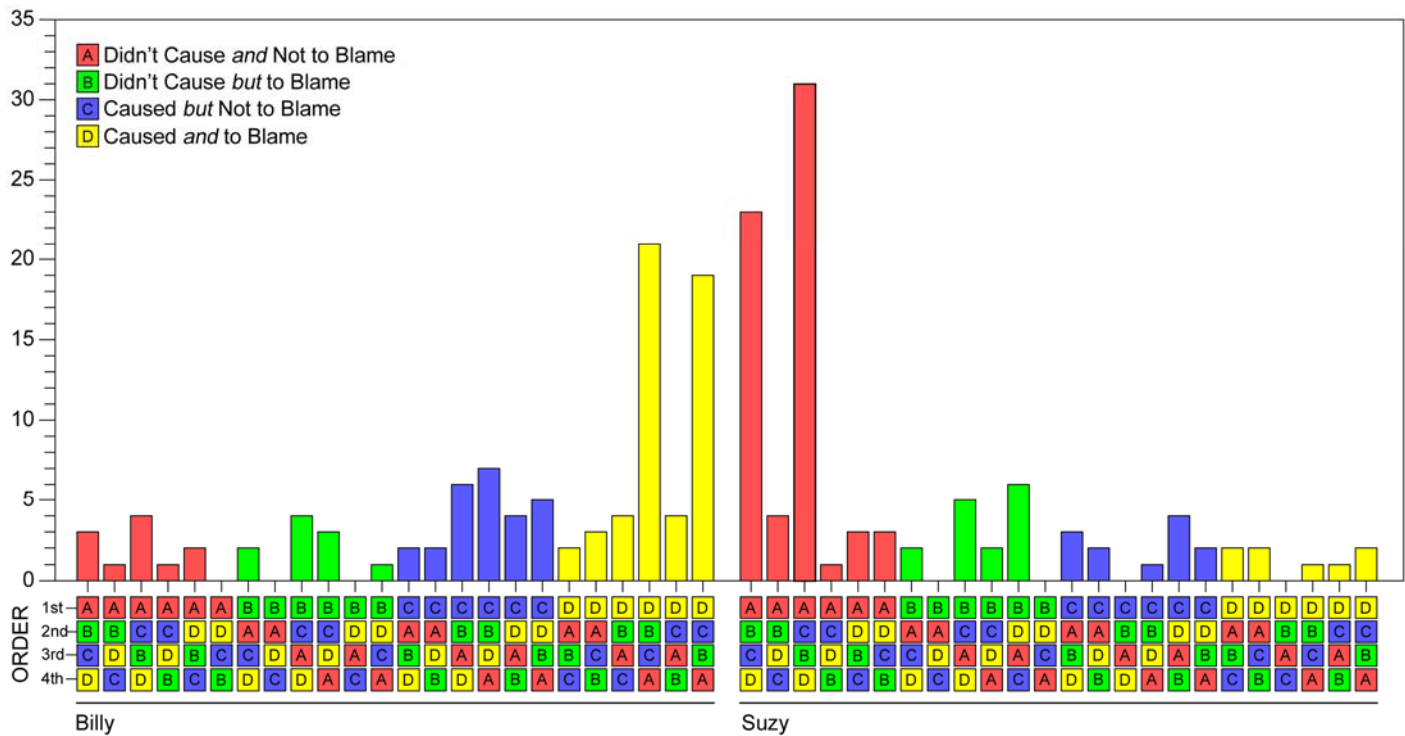
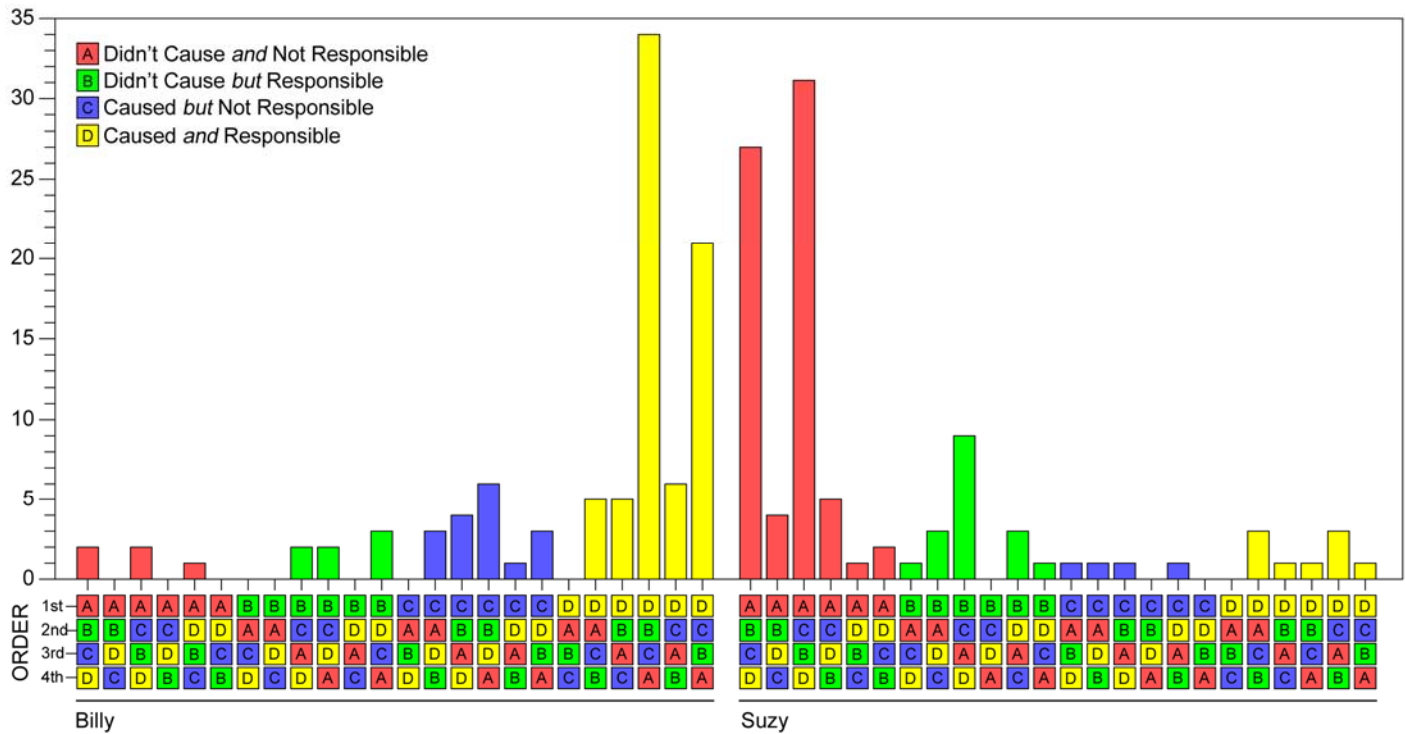
---

<sup>6</sup> One notable benefit of using a “push strategy” like this one (i.e., recruiting participants who were not directly looking to participate in research) is that participants are more likely to be “experimentally naïve” and less likely to be motivated to provide the responses that they think the experimenters are looking for (Haug 2018). Samples collected using the recruitment strategy employed here have been previously compared against samples collected with other methods in replication studies. And the present strategy has been consistently found to generate a diverse sample in terms of geography, socio-economic status, religiosity, political orientation, age, and education. Studies using this strategy have been previously reported in publications including, e.g., Livengood et al. (2010), Feltz and Cokely (2011), Sytsma and Machery (2012), Murray et al. (2013), Machery et al. (2015), Kim et al. (2016), Livengood and Rose (2016), Sytsma and Reuter (2017), Sytsma and Ozdemir (2019), Reuter and Sytsma (2020), Fischer et al. (forthcoming), Fischer and Sytsma (forthcoming).

<sup>7</sup> 53/253 (20.9%) of participants failed the check question.

one of the two statements denying causation ([A], [B]) in last place; and I predict that the opposite pattern will be found for the norm-conforming agent, Suzy. In addition, the responsibility account predicts that participants will treat the causal attributions and the normative attributions similarly, showing a stronger preference/dispreference for the “and” statements compared to the “but” statements. More specifically, I predict that for Billy participants will tend to show the strongest preference for [D] and the strongest dispreference for [A], while for Suzy they will tend to show the strongest preference for [A] and the strongest dispreference for [D].

The pragmatic account makes contrasting predictions. According to this view, when asking about compound attributions we would expect pragmatic considerations to lead participants to read “caused” and “responsible”/“blame” in contrasting senses since this would make the claims more informative, whereas reading them in the same sense would render them redundant. As such, advocates of the pragmatic account should expect the norm effect for causal attributions to largely disappear, with participants interpreting “caused” in the (supposedly) dominant descriptive sense. While Samland and Waldmann (2016, 165) claim that “the causal relations are trivial” for cases like this, as noted above, it is in fact unclear whether we would expect people to hold that *both* agents caused the outcome or that *neither* agent caused the outcome on a purely descriptive reading. If the former, then their pragmatic account would predict that people will show the strongest preference for [D] for Billy and the strongest preference for [C] for Suzy, judging that while both caused the outcome, Billy but not Suzy is responsible for (to blame for) the outcome. If the latter, then they would predict that people will show the strongest preference for [B] for Billy and the strongest preference for [A] for Suzy, judging that while neither caused the outcome, Billy but not Suzy is responsible for (to blame for) the outcome.



**Figure 1:** Histograms for each rank ordering for the questions in Studies 1a (top) and 1b (bottom).

### 4.3 Results

Rank orderings are shown in Figure 1. To analyze the results, I used the **pmr** package in R (Lee and Yu 2013). I began by comparing the marginal frequencies between Studies 1a and 1b for each of the two rank-ordering questions. That is, for each of the two agents, I compared the number of participants ranking each statement (**[A]**-**[D]**) in each position (1<sup>st</sup>-4<sup>th</sup>) between the two studies. The marginal frequencies were not significantly different for either Billy ( $\chi^2=19.58$ ,  $df=15$ ,  $p=.19$ ) or Suzy ( $\chi^2=14.17$ ,  $df=15$ ,  $p=.51$ ). In other words, the results were comparable whether combining causal attributions with responsibility attributions or with blame attributions. As such, I will collapse the two studies in the subsequent analysis. The same procedure was used to check for order effects, comparing marginal frequencies for each agent between the condition where the causal attribution was given first in each compound statement and the condition where it was given second. A significant order effect was found for the rankings for both Billy ( $\chi^2=44.76$ ,  $df=15$ ,  $p<.001$ ) and Suzy ( $\chi^2=25.79$ ,  $df=15$ ,  $p=.040$ ). As such, differences between the two orderings will be noted below. In brief, what we find is that the ordering impacted the relative preference for statements **[B]** and **[C]**, but not **[A]** and **[D]**.

To test whether the norm effect previously found for variations on the Email Case using Likert items replicates for the rank ordering questions, I began by comparing the marginal frequencies between the rank orderings for the two agents. There was a significant difference ( $\chi^2=368.02$   $df=15$ ,  $p<.001$ ). More importantly, a significant majority of participants ranked one of the two items stating that the norm-violating agent, Billy, caused the outcome (**[C]**, **[D]**) in first place (82.5%;  $\chi^2=83.20$ ,  $df=1$ ,  $p<.001$ ) and a significant majority ranked one of the two items denying that Billy caused the outcome (**[A]**, **[B]**) in last place (72%;  $\chi^2=37.84$ ,  $df=1$ ,  $p<.001$ ). And the opposite was found for the norm-conforming agent, Suzy: a significant

majority of participants ranked one of the two items denying that Suzy caused the outcome in first place (77.5%;  $\chi^2=59.40$ ,  $df=1$ ,  $p<.001$ ) and a significant majority of participants ranked one of the two items affirming that Suzy caused the outcome in last place (78.5%;  $\chi^2=63.84$ ,  $df=1$ ,  $p<.001$ ).<sup>8</sup> Thus, despite having options to register that Billy was responsible (to blame) for the outcome and that Suzy was not responsible (to blame) for the outcome, the norm effect was still in clear evidence, with participants preferring to say that Billy caused the outcome and participants preferring to say that Suzy did not cause the outcome. This runs strongly counter to what we would expect on the basis of the pragmatic account.

Looking more closely at the rankings for Billy, the mean ranks again suggest a clear preference for **[D]** and dispreference for **[A]**, while **[C]** and **[B]** are similar: **[A]** 3.38, **[B]** 2.44, **[C]** 2.42, **[D]** 1.76.<sup>9</sup> This was confirmed by looking at the pairwise frequencies. A significant majority of participants ranked **[D]** *higher* than **[A]** (82%;  $\chi^2=80.64$ ,  $df=1$ ,  $p<.001$ ), *higher* than **[B]** (72%;  $\chi^2=37.84$ ,  $df=1$ ,  $p<.001$ ), and *higher* than **[C]** (70%;  $\chi^2=31.20$ ,  $df=1$ ,  $p<.001$ ). Similarly, a significant majority of participants ranked **[A]** *lower* than **[B]** (78%;  $\chi^2=61.60$ ,  $df=1$ ,  $p<.001$ ) and *lower* than **[C]** (77.5%;  $\chi^2=59.40$ ,  $df=1$ ,  $p<.001$ ). Finally, a very slight majority of participants ranked **[C]** *higher* than **[B]** (50.5%;  $\chi^2=.005$ ,  $df=1$ ,  $p=.94$ ). Further, a significant majority of participants ranked **[D]** first (62.5%;  $\chi^2=12.00$ ,  $df=1$ ,  $p<.001$ ) and a significant

---

<sup>8</sup> Similar results held for the first ordering (caused first), with a significant majority ranking **[C]** or **[D]** in first place for Billy (82%;  $\chi^2=39.69$ ,  $df=1$ ,  $p<.001$ ), a significant majority ranking **[A]** or **[B]** in last place for Billy (74%;  $\chi^2=22.09$ ,  $df=1$ ,  $p<.001$ ), a significant majority ranking **[A]** or **[B]** in first place for Suzy (74%;  $\chi^2=22.09$ ,  $df=1$ ,  $p<.001$ ), and a significant majority ranking **[C]** or **[D]** in last place for Suzy (77%;  $\chi^2=28.09$ ,  $df=1$ ,  $p<.001$ ). And likewise for the second ordering (caused second), with a significant majority ranking **[C]** or **[D]** in first place for Billy (83%;  $\chi^2=42.25$ ,  $df=1$ ,  $p<.001$ ), a significant majority ranking **[A]** or **[B]** in last place for Billy (70%;  $\chi^2=15.21$ ,  $df=1$ ,  $p<.001$ ), a significant majority ranking **[A]** or **[B]** in first place for Suzy (81%;  $\chi^2=37.21$ ,  $df=1$ ,  $p<.001$ ), and a significant majority ranking **[C]** or **[D]** in last place for Suzy (80%;  $\chi^2=34.81$ ,  $df=1$ ,  $p<.001$ ).

<sup>9</sup> Mean ranks are similar for the first ordering (**[A]** 3.27, **[B]** 2.66, **[C]** 2.40, **[D]** 1.67) and the second ordering (**[A]** 3.48, **[B]** 2.23, **[C]** 2.44, **[D]** 1.85), although the relative order of **[B]** and **[C]** is shifted.

majority ranked [A] last (62%;  $\chi^2=11.04$ ,  $df=1$ ,  $p<.001$ ). Thus, the rankings for Billy confirm the prediction of the responsibility account.

The rankings for Billy also confirm the prediction of pragmatic accounts, *on the assumption* that for a purely descriptive reading of “caused” participants will judge that *both* agents caused the outcome. As such, the pragmatic account would then make a different prediction from the responsibility account for Suzy (i.e., pragmatic accounts would predict that participants will show the strongest preference for [C] and the strongest dispreference for [B]). If the pragmatic account instead predicted that participants would judge that *neither* agent caused the outcome, then the results for Billy run strongly counter to the prediction for this type of view: only a small minority ranked [B] first (9.5%;  $\chi^2=129.6$ ,  $df=1$ ,  $p<.001$ ) and only a small minority ranked [C] last (12.5%;  $\chi^2=111$ ,  $df=1$ ,  $p<.001$ ).

To further test for the preferred ordering for Billy, I used a weighted distance-based model with Spearman’s footrule as the distance measure (Lee and Yu 2012). The modal ranking given is [D] < [B] < [C] < [A], with parameter estimates of 0.46, 0.59, 0.13, and 0.55 (loglikelihood=532.17). The model was a good fit for the data as indicated by the sum of squares Pearson residuals ( $\chi^2=91.34$ ,  $df=24$ ,  $p<.001$ ). The same modal ranking was given for the Luce model and distance-based model, although the  $\phi$ -component model reversed the order of [B] and [C].<sup>10</sup> As noted above, the relative order of preference for these two statements differed between the orderings: when the causal attribution was first in the statements, all four models showed a modal ranking of [D] < [C] < [B] < [A], with participants showing a relative preference for the

---

<sup>10</sup> A pilot study (N=70) using the same design as Study 1a, but with the check question given on a second page with additional questions, showed this alternative ordering for Billy ([D] > [C] > [B] > [A]) for the mean ranks ([A] 3.56, [B] 2.60, [C] 2.34, [D] 1.50) and all four models. And it showed the reverse ordering for Suzy ([A] > [B] > [C] > [D]) for the mean ranks ([A] 1.71, [B] 2.41, [C] 2.56, [D] 3.31) and all four models. Outside of the ordering of [B] and [C] the results were quite similar to those reported here; in fact, the marginal frequencies were not significantly different between the pilot and Study 1a for either Billy ( $\chi^2=17.83$ ,  $df=15$ ,  $p=.27$ ) or Suzy ( $\chi^2=22.12$ ,  $df=15$ ,  $p=.10$ ).



“but” statement affirming the causal attribution ([C]) over the “but” statement denying the causal attribution ([B]); but when the normative attribution was first, all four models showed a modal ranking where these were reversed ([D] > [B] > [C] > [A]). Overall, the rankings suggest that participants did not have a strong preference between the “but” items, which is in keeping with the strong preference shown for affirming both attributions and the strong dispreference shown for denying both attributions.

Turning to the question about the norm-conforming agent, Suzy, we find a quite different pattern of results, with the mean ranks suggesting the opposite order of preference from that seen for Billy: [A] 1.58, [B] 2.62, [C] 2.39, [D] 3.42.<sup>11</sup> This was again confirmed by looking at the pairwise frequencies. A significant majority of participants ranked [A] *higher* than [B] (84%;  $\chi^2=91.12$ ,  $df=1$ ,  $p<.001$ ), *higher* than [C] (74%;  $\chi^2=45.12$ ,  $df=1$ ,  $p<.001$ ), and *higher* than [D] (84.5%;  $\chi^2=93.84$ ,  $df=1$ ,  $p<.001$ ). Similarly, a significant majority of participants ranked [D] *lower* than [B] (78%;  $\chi^2=61.60$ ,  $df=1$ ,  $p<.001$ ) and *lower* than [C] (79.5%;  $\chi^2=68.44$ ,  $df=1$ ,  $p<.001$ ). Finally, a slight majority of participants ranked [C] *higher* than [B] (55.5%;  $\chi^2=2.20$ ,  $df=1$ ,  $p=.14$ , two-tailed). Further, a significant majority of participants ranked [A] first (67.5%;  $\chi^2=23.80$ ,  $df=1$ ,  $p<.001$ ) and a significant majority ranked [D] last (67%;  $\chi^2=22.45$ ,  $df=1$ ,  $p<.001$ ). Thus, the rankings for Suzy confirm the prediction of the responsibility account.

The rankings for Suzy also confirm the prediction of the pragmatic account, *on the assumption* that for a purely descriptive reading of “caused” participants will judge that *neither* agent caused the outcome. But this assumption would then generate the opposite prediction to the responsibility account for Billy, which runs counter to the findings detailed above. If the pragmatic account instead predicts that participants will judge that *both* agents caused the

---

<sup>11</sup> Mean ranks are similar for the first ordering ([A] 1.63, [B] 2.54, [C] 2.52, [D] 3.30) and the second ordering ([A] 1.51, [B] 2.69, [C] 2.26, [D] 3.54).

outcome, then the results for Suzy run strongly counter to the prediction of this type of view: only a small minority ranked [C] first (13%;  $\chi^2=108.04$ ,  $df=1$ ,  $p<.001$ ) and only a small minority ranked [B] last (17%;  $\chi^2=85.80$ ,  $df=1$ ,  $p<.001$ ). In other words, whichever prediction the pragmatic account makes about the supposedly “trivial” descriptive relations, the rankings for one of the two agents will run strongly counter to the prediction.

To further test for the preferred ordering for Suzy, I again used a weighted distance-based model with Spearman’s footrule as the distance measure. The modal ranking given corresponds with the mean rank order—[A] > [C] > [B] > [D]—with parameter estimates of 0.58, 0.25, 0.51, and 0.63 (loglikelihood=504.30). The model was a good fit for the data as indicated by the sum of squares Pearson residuals ( $\chi^2=91.34$ ,  $df=24$ ,  $p<.001$ ). The same modal ranking was given for the Luce model, distance-based model,  $\phi$ -component model. As with the question about Billy, however, the relative order of preference for [B] and [C] differed between the orderings: when the causal attribution was given first in the statements, all four models showed a modal ranking of [A] > [B] > [C] > [D]; but when the normative attribution was instead given first, all four models showed a modal ranking where these were reversed ([A] > [C] > [B] > [D]).

#### *4.4 Discussion*

As noted in the previous sections, a key piece of evidence in favor of the responsibility account, and against competing accounts, is the close correspondence between causal attributions and responsibility attributions. Some of these studies involve judgments solicited between-subjects, but some use a within-subjects design. While each is problematic for competing accounts, the latter is especially problematic for the pragmatic account. The close correspondence between causal attributions and responsibility attributions in the between-subjects studies *might* be

explained in terms of pragmatic factors leading people to rate causal attributions similarly to how they would rate responsibility attributions, although the degree of correspondence is an issue for this type of view. When asking about the attributions together, however, we would expect the proposed pragmatic effect to be notably mitigated. If, for instance, the close correspondence in between-subjects studies is due to people interpreting the causal attribution as really asking about responsibility or accountability for pragmatic reasons, then we would expect participants to give divergent responses when asked about both attributions together: this should tend to promote interpreting the causal attribution as being intended in the dominant descriptive sense, according to Samland and Waldmann's reasoning.

This reasoning is even more clear for the present studies. In the studies reported in this section, participants did not simply rate a causal attribution and a normative attribution at the same time, but assessed joint statements combining both types of attributions. That is, assessing the statements required assessing both attributions together. And if it is the case that the dominant sense of one is purely descriptive, while the other is partly normative, then pragmatic considerations should *if anything* reinforce reading each in the dominant sense since this would be maximally informative. But, despite this, we saw that participants showed a marked preference for affirming both for the norm-violating agent and denying both for the norm-conforming agent. This provides strong evidence against the pragmatic view. It appears that the norm effect is a general effect, not a pragmatic effect.

That the close correspondence between causal attributions and normative attributions remains despite pragmatic pressures that would seem to promote interpreting them as making different claims is equally strong evidence in favor of the responsibility account: the most straightforward explanation is that there is a notable commonality in how people understand these

attributions, with the dominant use of each being partly normative. In contrast, to explain these new findings advocates of the bias account would need to contend that the motivational bias it posits is so strong that even when directly comparing the two types of attributions people fail to keep them separate, with their desire to blame still infecting their causal judgments. Likewise, to explain these new findings advocates of counterfactual accounts would need to contend that the effect of norms on the counterfactuals that people consider is so strong that it overrides the pragmatic pressures to consider the difference between the two types of attributions, which would seem to emphasize the descriptive symmetry between the two agents' actions.

## **5. Conclusion**

At this point, it is beyond dispute that the norm effect occurs for causal judgments: a wealth of findings shows that information about norms impacts people's causal judgments. But the extent of this phenomenon and, relatedly, how best to explain has been heavily contested. Despite this, I believe that answers to these questions are becoming clear. We are not merely dealing with a pragmatic effect or an effect that otherwise owes to a general underlying mechanism that leads us to apply a purely descriptive concept in a way that appears normative; we are simply dealing with the judgments that result from the correct application of a normative concept akin to responsibility or accountability. In the face of the present evidence, this is both the simplest and the most charitable explanation, requiring neither the posit of further mechanisms or general mistakes. The explanation of the norm effect is simply that we ordinarily use the lemma "cause" in a normative way.

## 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., Rose, D., and Bloom, D. (2011). "Causation, Norm Violation and Culpable Control." *Journal of Philosophy*, 108: 670-696.
- Danks, D., D. Rose, and E. Machery (2014). "Demoralizing Causation." *Philosophical Studies*, 171: 251–277.
- Darwall, S. (2001). "Normativity." In *The Routledge Encyclopedia of Philosophy*, Taylor and Francis. <https://www.rep.routledge.com/articles/thematic/normativity/v-1>
- Fischer, E., P. Engelhardt, and J. Sytsma (forthcoming). "Inappropriate stereotypical inferences? An adversarial collaboration in experimental ordinary language philosophy." *Synthese*.
- Fischer, E. and J. Sytsma (forthcoming). "Zombie Intuitions." *Cognition*.
- Feltz, A. and E. Cokely (2011). "Individual differences in theory-of-mind judgments: Order effects and side effects." *Philosophical Psychology*, 24(3): 343-355.
- Halpern, J. and C. Hitchcock (2015). "Graded Causation and Defaults." *British Journal for the Philosophy of Science*, 66: 413–457.
- Haug, Matthew (2018). "Fast, Cheap, and Unethical? The Interplay of Morality and Methodology in Crowdsourced Survey Research." *Review of Philosophy and Psychology*, 9(2): 363-379
- 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.
- Kelly, D. and S. Setman (2020). "The Psychology of Normative Cognition." In E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). <https://plato.stanford.edu/archives/fall2020/entries/psychology-normative-cognition/>
- Kim, H., N. Poth, K. Reuter, and J. Sytsma (2016). "Where is your pain? A Cross-cultural Comparison of the Concept of Pain in Americans and South Koreans." *Studia Philosophica Estonica*, 9(1): 136-169.
- 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.

Lee, P. and P. Yu (2013). “An R package for analyzing and modeling ranking data.” *BMC Medical Research Methodology*, 13: 65.

Livengood, J. and D. Rose (2016). “Experimental Philosophy and Causal Attribution.” In J. Sytsma and W. Buckwalter (eds.), *A Companion to Experimental Philosophy*, Wiley Blackwell, 434-449.

Livengood, J. and J. Sytsma (2020). “Actual causation and compositionality.” *Philosophy of Science*, 87(1): 43-69.

Livengood, J., J. Sytsma, A. Feltz, R. Scheines, and E. Machery (2010). “Philosophical Temperament.” *Philosophical Psychology*, 23(3): 313-330.

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.

Machery, E., J. Sytsma, and M. Deutsch (2015). “Speaker’s Reference and Cross-cultural Semantics.” In A. Bianchi (ed.), *On Reference*, Oxford University Press, 62-76.

Murray, D., J. Sytsma, and J. Livengood (2013). “God Knows (But does God Believe?)” *Philosophical Studies*, 166: 83-107.

O’Neill, E. (2017). “Kinds of Norms.” *Philosophy Compass*, 12(5): e12416.

Paul, L. A. and N. Hall (2003). *Causation: A User’s Guide*. Oxford: OUP.

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

Reuter, K. and J. Sytsma (2020). “Unfelt Pain.” *Synthese*, 197: 1777-1801.

Rose, D. (2017). “Folk Intuitions of Actual Causation: A Two-pronged Debunking Explanation.” *Philosophical Studies*, 174(5): 1323–1361.

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.

Schwenkler, J., and E. T. Sievers (forthcoming). "Cause, 'Cause', and Norm." To appear in P. Willemsen and A. Wiegmann (eds.), *Advances in Experimental Philosophy of Causation*, London: Bloomsbury Press.

Schwenkler, J. and J. Sytsma (ms). "Reversing the Norm Effect on Causal Attributions." <http://philsci-archive.pitt.edu/18220/>

Skow, B. (2019). *Causation, Explanation, and the Metaphysics of Aspect*. Oxford: OUP.

Sytsma, J. (forthcoming-a). "Causation, Responsibility, and Typicality." *Review of Philosophy and Psychology*.

Sytsma, J. (forthcoming-b). "Crossed Wires: Blaming Artifacts for Bad Outcomes." *The Journal of Philosophy*.

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

Sytsma, J. (ms-b). "Resituating the Influence of Relevant Alternative on Attributions." <http://philsci-archive.pitt.edu/16957/>

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

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.

Sytsma, J. and J. Livengood (forthcoming). "Causal Attributions and the Trolley Problem." *Philosophical Psychology*.

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 E. Machery (2012). "On the Relevance of Folk Intuitions: A Reply to Talbot." *Consciousness and Cognition*, 21: 654-660.

Sytsma, J. and E. Ozdemir (2019). "No Problem: Evidence that the Concept of Phenomenal Consciousness is Not Widespread." *Journal of Consciousness Studies*, 26(9-10): 241-256.

Sytsma, J. and K. Reuter (2017). "Experimental Philosophy of Pain." *Journal of Indian Council of Philosophical Research*, 34(3): 611-628.