Causation, Norms, and Cognitive Bias

Levin Güver & Markus Kneer

Abstract

Extant research has shown that ordinary causal judgments are sensitive to normative factors. For instance, agents who violate a norm are standardly deemed more causal than norm-conforming agents in identical situations. In this paper, we explore two competing explanations for the Norm Effect: the Responsibility View and the Bias View. According to the former, the Norm Effect arises because ordinary causal judgment is intimately intertwined with moral responsibility. According to the alternative view, the Norm Effect is the result of a blame-driven bias. In a series of five preregistered experiments (N = 2688), we present evidence in favour of the Bias View. In particular, and against predictions made by the Responsibility View, we show that participants deem agents who violate nonpertinent or silly norms – norms that do not relate to the outcome at hand – as more causal, and that this effect cannot be explained in terms of participants ascribing foreknowledge, desire, or foreseeability of harm to the norm-violating agent. We close with a discussion of these findings and point to important implications for the just assessment of proximate cause in the law.

Keywords: causation; norms; bias; blame; responsibility; foreseeability; negligence

1. Introduction

1.1 The Impact of Norms on Perceived Causation

A growing body of literature has revealed that ordinary causal judgement is susceptible to the violation of norms: when two agents perform the same action, yet one does so in violation of a norm, the norm-violating agent is taken to be *the* cause of the harmful outcome (Alicke, 1992, 2000; Henne et al., 2021; Henne & O'Neill, 2022; Hitchcock & Knobe, 2009; Icard, Kominsky, & Knobe, 2017; Knobe & Fraser, 2008; Kominsky et al., 2015; Samland & Waldmann, 2016; Olier & Kneer, 2022). This phenomenon has since been called the *Norm Effect*.[1] To illustrate, imagine the following scenario (*Rollerblading*): Mark is rollerblading on a path while Lauren is walking ahead of him. Suddenly, a cat jumps out of the brush, startling her. Lauren sidesteps to the left, directly into the lane of Mark, who is unable to break in time. The two collide. Who caused the accident?

[1] The Norm Effect has been shown to arise not only in the context of prescriptive (or injunctive) norms, i.e. norms concerning what should or should not be done, but also with descriptive (or statistical) norms, which describe what typically happens (Gerstenberg & Icard, 2020; Kirfel & Lagnado, 2018; Knobe & Fraser, 2008; Livengood, Sytsma, & Rose, 2017; Morris et al., 2019; Sytsma, Livengood, & Rose, 2012). In the following, we principally focus on prescriptive norms.

Participants overwhelmingly point to the cat. Now imagine a slight variation of the situation – the addition of a norm prohibiting Mark from rollerblading on the path – and ask again: who caused the accident? Mark's violation of a salient norm leads to a drastic shift in participants' judgements, the majority now considering Mark the cause (Güver & Kneer, 2023). This is an example of the Norm Effect, and several accounts compete to explain the underlying causal mechanisms (*e.g.* Gerstenberg & Icard, 2020; Henne et al., 2021; Hitchcock & Knobe, 2009; Samland & Waldmann, 2016); for a review, see Willemsen & Kirfel, 2019; more generally, see Rose & Danks, 2012; Livengood & Rose, 2016; Henne, 2023; Bebb & Beebee, 2024). In the following, we will focus on but two such explanations: The *Responsibility View* and the *Bias View*.

1.2 Two Explanations

According to the Responsibility View, causal judgements are intimately connected with responsibility judgements. When ordinary people use locutions such as "Mark *caused* the accident", they take themselves to be saying something akin to "Mark is *responsible* for the accident" (Sytsma, 2019a, 2022; Sytsma et al., 2023; Sytsma, Livengood, & Rose, 2012), where "responsible" is understood as a "normative evaluation" (Sytsma, 2019b). Thus, when people use the expression "cause", they take it to refer to a *normative* concept (Sytsma, 2021; Sytsma, Livengood, & Rose, 2012). The Norm Effect, then, is simply the upshot of the folk correctly applying this normative concept of causation (Sytsma, 2021), as schematised in Figure 1.

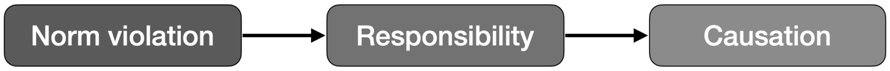

Figure 1: A simple pathway model of the Responsibility View.

According to the Bias View (also called the Culpable Control Model, Alicke, 1992, 2000), however, the concept of causation is, in fact, descriptive, and the Norm Effect constitutes a bias. As Alicke writes, our "desire to praise or denigrate those whose actions we applaud or deride" leads to a performance error, *i.e.* a norm-sensitive attribution of causal contribution to the agent (Alicke, Rose, & Bloom, 2011; Alicke & Rose, 2012; Rose, 2017; see also Rogers et al., 2019). When it comes to norm violations, the culprit are blame judgments: knee-jerk reactions which makes us see the agent in a negative light, thereby tainting our evaluation of her (Alicke, 2008). It is this tainted view of the agent which, in an act of backwards-rationalisation, leads us to exaggerate her causal contribution in bringing about the outcome (Alicke, 1992, 2000; schematised in Figure 2).

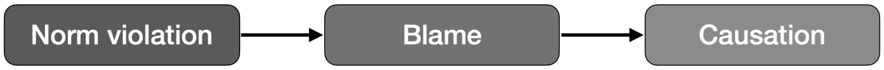

Figure 2: A simple pathway model of the Bias View.

Both the Responsibility View and the Bias View posit that the Norm Effect is driven by a normative judgement, be it one of responsibility or blame. As such, the positions are very similar. How can they be distinguished?

By responsibility judgments, Sytsma means "broadly *moral* evaluations" (Sytsma, 2021). Unlike the Bias View, the Responsibility View requires us to distinguish "features that are *irrelevant* to appropriately assessing responsibility" (Sytsma, 2019b) from those that *legitimately* heighten the agent's responsibility towards the outcome. Features that are irrelevant to the agent's responsibility – such as race, gender, sexual orientation, or general character (Alicke, Rose, & Bloom, 2011) – should, on the Responsibility View, not have an influence on causation, even if they inadequately influence *perceived* responsibility. The Bias View, on the other hand, does not draw a distinction between legitimate and illegitimate drivers of blame. It states that any feature apt to influence *perceived* blameworthiness – be it legitimate or illegitimate – can influence folk causal judgement. To tease apart the two views, one must thus probe whether factors irrelevant to moral responsibility proper influence causal judgement or not.

An early example of this approach can be found in Alicke (1992), whose results seemed to suggest that persons with bad general character – a feature irrelevant to responsibility in the specific situation at hand (a road accident) – were indeed deemed more causal. However, as Sytsma (2019b) has suggested, the participants in Alicke's original study might have implicitly drawn inferences from the agent's bad character to factors that *are* relevant to the agent's responsibility. In several replications, Sytsma illustrates that the difference between drivers speeding home – one to hide a vial of cocaine, the other to hide a present – is not only one of general character but also of perceived driving ability. A difference in driving ability, in turn, is relevant to the assessment of agential responsibility when an accident occurs.

A more sophisticated version of the Responsibility View, such as the one proposed by Sytsma (2019b), accounts for the mediating role of several potentially inferred factors (Figure 3). Returning to our opening example, *Rollerblading,* participants may, for instance, infer that Mark should have foreseen a crash (*foreseeability*) or did foresee it (*foresight*) when he violated a norm in skating on the wrong path. In other scenarios, one might even go as far as inferring a *desire* to cause an accident.[2] Just like ability or skill, the potentially inculpating mental state (*mens rea*) of the agent is relevant to the assessment of moral responsibility, and hence, on Sytsma's account, to the determination of causal responsibility.

[2] This is consistent with the findings of Kirfel and Phillips (2021, 2023) who show that agents who unknowingly violate norms are not judged more causal than their norm-adhering counterparts: agents that unknowingly violate norms neither ought to be held responsible nor are they blameworthy (unless they should have known of the risk of harm, *i.e.* acted negligently).

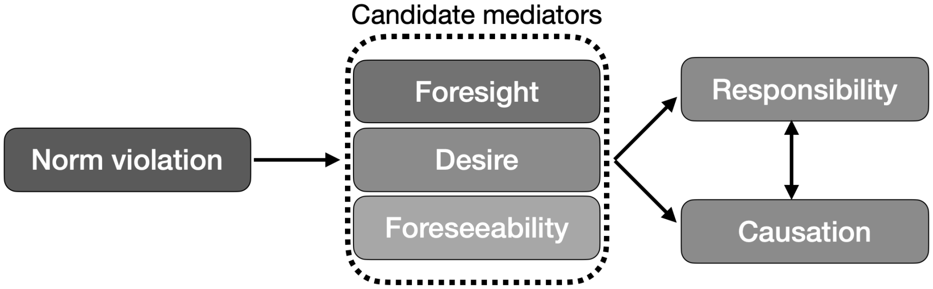

Figure 3: A more complex pathway model of the Responsibility View.

At this point it is helpful to distinguish two variations of the (complex) Responsibility View (see Güver & Kneer, 2023, p. 140ff.). According to a permissive version, *any* factor that impacts folk attributions of blame or responsibility constitutes a legitimate impact on causation attributions (precisely because they impact responsibility). An account of this sort could be called the *Anything-Goes View*, since it is *too* permissive. If folk attributions were, for instance, influenced by gender in misogynistic ways, then a similar influence on causation would be fine on this crude account. After all, any factor that sways *perceived* responsibility would constitute an adequate influence on causation. Presumably, this is not the view proposed by Sytsma and Livengood. Echoing Alicke, Sytsma argues that certain factors are "peripheral" to the assessment of moral responsibility, including "the actor's or victim's race and character" (Alicke, Rose, & Bloom, 2011, p. 674). In contrast to the *Anything-Goes View*, there are thus legitimate and illegitimate factors influencing perceived responsibility, and only the legitimate ones should impact causation. Hence, if adherence to salient norms – despite the fact that they have no clear connection to causation – are considered as a relevant factor for the assessment of moral responsibility, then their impact on causation is justified, too. However – and herein lies the difference to the *Anything-Goes View* – factors that should *not* influence perceived moral responsibility, such as race, gender, or general character, should *not* influence perceived causation either.

1.3 The Present Experiments

In his interesting and rich paper, *The Character of Causation*, Sytsma (2019b) discusses two potential challenges to the Responsibility View. First, he explores factors which – in specific contexts at least – clearly should not impact responsibility for a particular action, such as the agent's general moral character (or motive, *cf.* Goulette & Verkampt, 2023). The key example in this regard is Alicke's famous experiment, in which a driver speeding home to hide cocaine from his parents is deemed more causally responsible for an accident than another one who wants to hide a present for his parents' wedding anniversary. This, Sytsma acknowledges, constitutes *prima facie* evidence in favour of the Bias View. However, he reports a new experiment which shows that there seems to be a confound: From the fact that one of the drivers wants to hide cocaine, people infer that he is a poor driver. Speeding despite low driving ability is reckless, and hence it does, after all, seem appropriate to attribute more moral responsibility – and on the Responsibility View causal responsibility – to the cokehead.

Sytsma's second set of experiments concerns the Norm Effect, or, more particularly whether the influence of norm violation on causation constitutes evidence in favour of the Bias View or the Responsibility View. He writes:

> Alicke's bias view holds that not only do features of the agent's mental states matter, such as her knowledge and desires concerning the norm and the outcome, but also peripheral [i.e. *prima facie* irrelevant] features of the agent whose impact could only reasonably be explained in terms of bias. In contrast, our responsibility view holds that the impact of norms does not reflect bias, but rather that ordinary causal attributions issue from the appropriate application of a concept with a normative component. As such, we predict that while judgments about the agent's mental states that are relevant to adjudicating responsibility will matter, peripheral features of the agent will only matter insofar as they warrant an inference to other features of the agent that are relevant. (2019b, p. 25)

Sytsma, it appears, agrees with Alicke that norm conformity is "peripheral" to causal attributions, *except* if it triggers justifiable inferences regarding mediators that correlate with moral responsibility. The mediators of interest could be inculpating mental states (intention, knowledge, recklessness etc.) or the above-discussed abilities of the agent. Differently put, Sytsma seems to hold that a *direct* effect of norm violation on causal attributions is evidence in favour of the Bias View, though *indirect* effects via *mens rea* and other "nonperipheral" factors support the Responsibility View. Indirect effects are addressed in a second set of studies which demonstrate that, once mental states such as foreknowledge and desire are explicitly controlled for, the (direct) effect of norm on causality attribution is marginal (see e.g. Sytsma, 2019b, Study 4). This suggests that the effect of norms is not peripheral, as it exerts its influence not directly, but via *mens rea*, i.e. features to which moral responsibility (and, on Sytsma's account, therefore causation) should be sensitive.

In this paper, we would like to present two challenges to the Responsibility View, and to shed further light on the Norm Effect more generally. The first challenge explores violations of norms which are *not pertinent* to the harm caused on the one hand, or outright *silly* on the other. As Sytsma (2019a) explicitly acknowledges, it "is imperative for [the Responsibility View] that the norm-violating action is connected to the outcome" (p. 14). Since this is not the case for either type of norm tested, we take it that all parties to the debate agree that violations of nonpertinent or silly norms are "peripheral" factors, i.e. factors which clearly should not influence attributions of moral responsibility or causal contribution. The second challenge focuses on the effect of *pertinent* norms, i.e. norms which are connected to the outcome of interest. Here we try to show that there are cases where the influence of norm violations is *not* mediated by desire, foresight and foreseeability, and hence (presumably) direct.[3] In sum, we aim to present a contingent objection to the Responsibility View (contingent on the mediators tested, and the precise account of the

[3] The number of potential mediators is unlimited, though the number of those that make explanatory sense is small. We focus on those addressed by advocates of the Responsibility View.

Responsibility View favoured), and a more direct challenge in the form of silly norm violations. After all, silly norm violations should impact neither moral responsibility nor causal attributions, be it directly or indirectly, because there are no reasonable features related to the downstream DVs that should be sensitive to an agent's adherence or violation of nonsense rules.

In Experiment 1, we show that nonpertinent and silly norm violations – i.e. norm violations that do not relate to the outcome at hand and thus ought not, on the Responsibility View, influence causal judgements – *do* have a pronounced effect on participants' causal ascriptions, and that this effect cannot be explained by recourse to Sytsma's proposed mediators of foreknowledge and desire. Experiment 2 builds on these findings and highlights additionally that participants in a within-subjects design, i.e. when given the chance to reflect on the scenarios at hand, do *not* let nonpertinent and silly norm violations influence their causal judgement. Experiment 3 replicates the findings of Experiments 1 and 2 with a novel scenario. In Experiment 4, we pre-empt the criticism according to which foreknowledge and desire are the wrong mediators and test instead whether participants deem the outcome foreseeable, i.e. whether they believe the agent to be acting negligently. While our *ex post* data suggests that participants do judge the accident as more foreseeable, our *ex ante* data reveals that participants fall prey to the hindsight bias, and once the hindsight bias is corrected for, the foreseeability of harm is unable to do the necessary explanatory work. We replicate these findings in Experiment 5, and close by considering their implications, in particular for the law, where attributions of *proximate cause* are of central relevance to just verdicts (see Knobe & Shapiro, 2021).

2. Experiment 1

In our first experiment, we explore both challenges to the Responsibility View empirically. We test whether the Norm Effect is mediated by inferred mental states (foreknowledge and desire), or whether it is direct. We also explore whether a Norm Effect arises in cases where the agent violates a norm unrelated to the outcome at issue (i.e. a nonpertinent norm), or where he violates an silly norm, unrelated to *any* kind of potentially harmful outcome. The vignette, titled *Festival*, is based on a real criminal case.[4] All preregistrations, materials, data, and additional analyses are available under https://osf.io/24uvf/?view_only=ccd04f1940bd468eafd42757a2ea099b.

2.1 Participants

We recruited 305 participants on Amazon Mechanical Turk. Their IP address was restricted to the United States. As preregistered,[5] participants who failed an attention check, spent less than 10 seconds reading the vignette, failed a comprehension question, or were not native English speakers were excluded. 195 participants remained (female: 45%; mean age: 40 years, SD = 12 years, range: 19–72 years).

[4] See https://www.bbc.com/news/world-asia-33300970 (accessed 20 September 2023) and https://www.taipeitimes.com/News/front/archives/2016/04/27/2003644910 (accessed 20 September 2023).

[5] Available under https://aspredicted.org/FKE_PAO.

2.2 Methods and materials

In the *Festival* scenario (full vignette in Appendix Section 4.1), Mark attends a music festival where Lauren is responsible for the special stage effects. During the concert, Lauren launches coloured powder over the dancing crowd which, unbeknownst to both her and the crowd, is combustible. The powder comes into contact with Mark's cigarette, ignites, and injures several festivalgoers.

The study took a between-subjects design and participants were randomly sorted into either the no norm, pertinent norm, nonpertinent norm, or silly norm condition. The no norm condition is silent as to whether smoking is permitted on the festival grounds. In the pertinent norm condition, smoking is explicitly forbidden. In the nonpertinent norm condition, the festival organizers prohibited attendees to be topless. Nevertheless, Mark attends in his underwear only. In the silly norm condition, the festival – in an attempt to break a world record – had asked everyone to wear a green cap. Mark, who had initially agreed to do so, ultimately decides against it, and the festival fails to break the record.

Having read the vignette, participants were asked to rate the extent to which Mark and Lauren causally contributed to injuring the festivalgoers. Follow-up question asked participants to judge the extent to which Mark had foreknowledge and the desire to harm them. Finally, participants were asked to rate Mark's blameworthiness, moral responsibility, and deserved punishment. All responses were recorded on 7-point Likert scales.

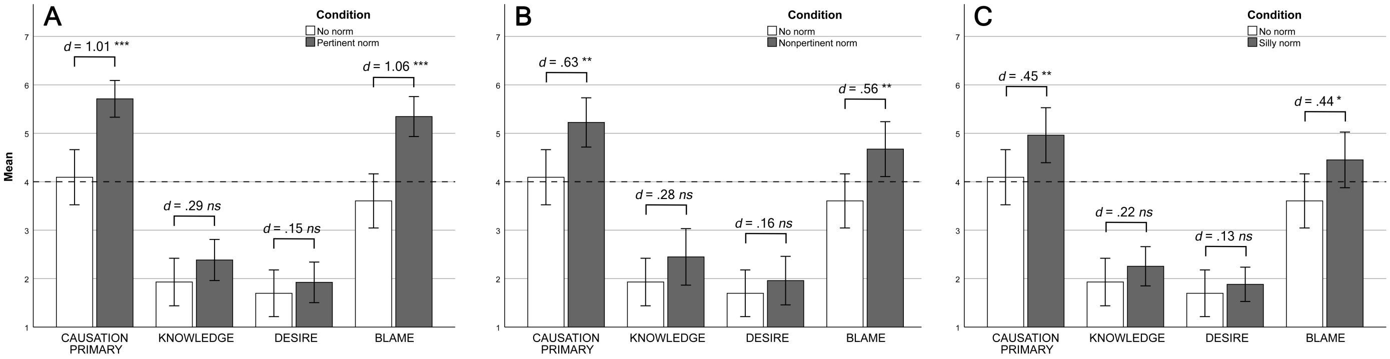

Figure 4: Comparison of means between the no norm and pertinent norm (A), nonpertinent norm (B), and silly norm (C) conditions. Effect sizes are given in terms of Cohen's d, * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. Error bars denote 95% CI.

2.3 Results

ANOVAs. One-way ANOVAs revealed a significant influence of norm status (no norm, pertinent, nonpertinent, and silly norms) on Mark's perceived causal contribution, blame, responsibility, and deserved punishment (all *p*s < .011), with large effect sizes for Mark's causal contribution and blameworthiness (η_p^2s > .156) and moderate effect sizes for responsibility and punishment (responsibility: η_p^2 = .126; punishment: η_p^2 = .115). Importantly, the effect of norm status on knowledge and desire proved nonsignificant (*p*s > .234). We further ran planned comparisons for all dependent variables, contrasting each of the three norm types with the no norm condition.

No norm v. pertinent norm. A comparison of the no norm and pertinent norm conditions revealed that participants in the latter judged Mark significantly and pronouncedly more causal, blameworthy, responsible, and deserving of punishment (*p*s < .001, *d*s > 1.00, large effects). There was no statistically significant difference across conditions for the knowledge and desire variables (*p*s > .159; see Figure 4A and, for this and the following contrasts, Appendix Section 3.1).

No norm v. nonpertinent norm. A comparison of the no norm and nonpertinent norm conditions yielded surprising results: In the nonpertinent norm condition – where all Mark did was violate the dress policy – participants rated him more causal, blameworthy, and responsible for the accident, as well as more deserving of punishment (all *p*s < .009, all *d*s > .55). The difference in knowledge and desire ratings did not reach statistical significance (*p*s > .173; see Figure 4B).

No norm v. silly norm. Contrasting the no norm and silly norm conditions, we found the previous pattern to persist: participants judged Mark significantly and considerably more causal, blameworthy, responsible, and deserving of punishment (all *p*s < .039, all *d*s > 0.43, moderate effects). There was no statistically significant impact of norm type on knowledge and desire judgements (*p*s > .302; see Figure 4C).

2.4 Discussion

Our experiment replicated previous findings concerning the Norm Effect: Mark was judged more causal in the condition where he violated a pertinent norm vis-à-vis the condition where he did not violate any norm. However, knowledge and desire were unaffected by norm violation, and significantly below the midpoint (and hence unlikely inferred factors). If one were to hold, like Sytsma seems to, that the Responsibility View can *only* accommodate the Norm Effect if there is a further, reasonable mediating factor – like *mens rea* – triggered by the difference in norms, then our findings count as tentative counterevidence to this account (tentative because there might be other, untested factors).

Our experiment also replicated preliminary findings according to which nonpertinent and silly norms exert an effect on blame and causation (Güver & Kneer, 2023). These results directly challenge the Responsibility View. The nonpertinent and silly norm conditions were explicitly designed so that violating them would not make the agent more responsible for the outcome, given the clear lack of connection between the norm violation (*e.g.* failing to adhere to the dress code) and the harm that ensued (the injury of some festivalgoers). As Sytsma (2019b) has stressed, in the absence of a connection between action and outcome, an agent should not be held responsible. Yet this is exactly what we find: Participants judged Mark in the nonpertinent and silly norm conditions as pronouncedly more causal, blameworthy, responsible, and deserving of punishment than in the norm-adhering condition. And this despite there being no significant difference in desire or knowledge ascriptions.[6]

The findings are most naturally interpreted as consistent with the Bias View: Participants view the norm-violating agent in a negative light, irrespective of the kind of norm violated. They want to blame the norm-violating agent, and thus exaggerate his causal contribution, in an attempt to justify the attributed blame.

Note, however, that it could be that the folk *do* view nonpertinent and silly norm violations as relevant to moral responsibility (i.e., there could be a difference between what moral philosophy and folk ethics deem normatively appropriate). If this were the case, *some* version of the Responsibility View – or what we have labelled the *Anything-Goes View* above – could argue that the difference in perceived causality is, after all, justified (although it would clash with normativity proper).

To explore this question in more detail, and to put further pressure on the explanatory adequacy of the Bias View, we ran Experiment 1 as a within-subjects design, where participants were confronted with both conditions (no norm v. pertinent norm/nonpertinent norm/silly norm) side by side. The aim was to see whether this reduces the bias (if adequately characterized as such), given that the single difference across conditions stares participants into the face. People could thus decide whether the difference in norm status, according to their moral outlook, merits a difference in attributed moral responsibility (or blame) and causation. Contrasting designs in this way has been fruitful in other areas of moral psychology: For instance, in studies investigating moral luck, the sizeable between-subjects effect of outcome (neutral v. bad) on blame largely disappears in within-subjects designs, suggesting that it *does* constitute a bias (Kneer & Machery, 2019; Frisch et al. 2021; Kneer & Skoczen, 2023; see also Baron, 2008; Hsee, 1996).

3. Experiment 2

Experiment 2 explores whether a more permissive version of the Responsibility View, which ties folk causality not to moral responsibility proper, but to *perceived* moral responsibility, could explain the results reported in Experiment 1. As briefly sketched above, the extent to which such a view is plausible is another matter (see also Güver & Kneer, 2023).

3.1 Participants

358 participants were recruited online via Amazon Mechanical Turk. Their IP address was restricted to the United States. As preregistered,[7] participants who failed an attention check, spent less than 20 seconds reading the vignette, or were not native English speakers were excluded. 287 participants remained (female: 49.5%; mean age: 44 years, SD = 14 years, range: 21–84 years).

[6] For the role of knowledge – or lack thereof – in ascriptions of causation and responsibility, see *e.g.* Samland et al. (2016); Kirfel & Lagnado (2021c); Engelmann (2022); Kirfel et al. (2023).

[7] Available under https://aspredicted.org/3T4_JWR.

3.2 Methods and materials

The study, building on the *Festival* vignette introduced above, took a mixed-factorial design (within-subjects factor – norm status: no norm v. norm; between-subjects factor – norm type: pertinent v. nonpertinent v. silly). It was identical to Experiment 1 in all respects, except that participants were presented with *two* vignettes on the same page and were subsequently asked to judge all measures with respect to *both* vignettes, *i.e.* the causal contributions of the primary agents (Mark and John) and secondary agents (Lauren and Mary), the primary agents' foreknowledge and desire, as well as their blameworthiness, moral responsibility, and deserved punishment. As in Experiment 1, all responses were recorded on 7-point Likert scales.

3.3 Results

3.3.1 General Results

ANOVAs. Repeated-measures ANOVAs revealed a significant effect of norm status on the causal contribution of the primary agent (p < .001, η_p^2 = .070) and the moral variables (ps < .001, η_p^2s < .110). The effect on knowledge and desire, however, was very small (η_p^2s < .020) and reached significance only for knowledge (p = .020) but not desire (p = .062).

We ran planned contrasts for a more detailed breakdown of the impact of norm type on the dependent variables. Table 1 contrasts a summary of the key findings with the between-subjects findings from Experiment 1 (for full tables and analyses, see Appendix Section 3.2).

		Between-subjects					Within-subjects				
Contrast	*Variable*	*df*	*t*	*p*	Cohen's *d*	95% CI	*df*	*t*	*p*	Cohen's *d*	95% CI
NN v. PN	Causation Primary	76	–4.77	<.001	1.01	[–1.44;–.58]	91	–6.72	<.001	0.70	[–.93;–.47]
	Knowledge	93	–1.42	0.160	0.29	[–.70;.12]	91	–2.68	0.009	0.28	[–.49;–.07]
	Desire	93	–.71	0.478	0.15	[–.55;.26]	91	–1.80	0.075	0.19	[–.39;.02]
	Blame	93	–5.15	<.001	1.06	[–1.49;–.63]	91	–8.42	<.001	0.88	[–1.12;–.64]
NN v. NP	Causation Primary	90	–3.00	0.004	0.63	[–1.04;–.21]	93	–.84	0.403	0.09	[–.29;.12]
	Knowledge	89	–1.37	0.174	0.28	[–.69;.13]	93	–.75	0.455	0.08	[–.28;.13]
	Desire	90	–.75	0.454	0.16	[–.57;.25]	93	–.80	0.428	0.08	[–.28;.12]
	Blame	90	–2.69	0.008	0.56	[–.98;–.14]	93	–.53	0.597	0.06	[–.26;.15]
NN v. SN	Causation Primary	92	–2.16	0.034	0.45	[–.86;–.03]	100	0.42	0.675	0.04	[–.15;.24]
	Knowledge	92	–1.04	0.303	0.22	[–.62;.19]	100	–.46	0.650	0.05	[–.24;.15]
	Desire	92	–.63	0.528	0.13	[–.54;.28]	100	–.67	0.503	0.07	[–.26;.13]
	Blame	92	–2.11	0.038	0.44	[–.85;–.02]	100	–.18	0.855	0.02	[–.21;.18]

Table 1: Comparison of effect sizes for the no norm v. pertinent norm (NN v. PN), nonpertinent norm (NN v. NP), and silly norm (NN v. SN) conditions across ascriptions of causation, mental states, and blame. 95% Confidence Intervals (CIs) for the reported *d*-values.

3.3.2 Planned comparisons

No norm v. pertinent norm. Participants judged the primary agent in the pertinent norm condition, John, as more causal than his no norm counterpart, Mark (p < .001, d = .70) (see Figure 5a). Norm status also had a significant effect on the moral variables (ps < .001, ds > .77). The effects on knowledge and desire were small (ds < .29) and reached significance only for knowledge (p = .009). Only about one in three participants rated the causal contribution of Mark, as well as blame and responsibility identically across scenarios. This suggests that a significant majority, when viewing the two scenarios side-by-side, considered the pertinent norm a *legitimate* influence on causation, blame and responsibility.

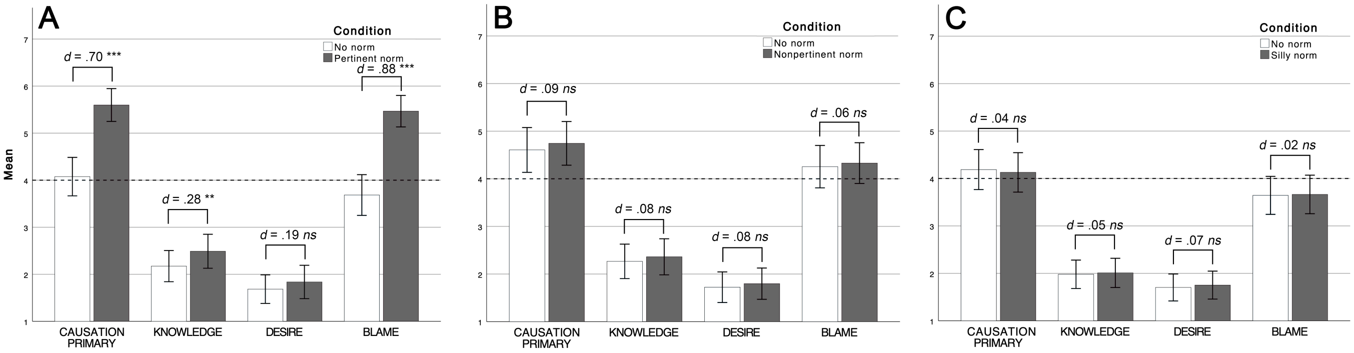

Figure 5: Comparison of means between the no norm and pertinent norm (A), nonpertinent norm (B), and silly norm (C) conditions. Effect sizes are given in terms of Cohen's d, * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. Error bars denote 95% CI.

No norm v. nonpertinent norm. In comparing the no norm and nonpertinent norm conditions (see Figure 5b), we did not find any statistically significant differences in participants' assessments of the dependent variables (all *p*s > .173), with very small effect sizes throughout (all *d*s < .15). By contrast to the pertinent norm comparison, more than 70% of the participants rated Mark's causal contribution, deserved blame and his moral responsibility identically across conditions. This suggests that most people considered the nonpertinent norm as *irrelevant* for the assessment of these dependent variables.

No norm v. silly norm. A comparison of the no norm and silly norm conditions, too, did not yield any statistically significant differences for the dependent variables (all *p*s > .123), with tiny effect sizes throughout (all *d*s < .08, except for the secondary agent's causal contribution at *d* = .15) (see Figure 5c). More than 80% of the participants rated Mark's causal contribution, deserved blame and his moral responsibility identically across conditions, again suggesting that they considered it as irrelevant for the latter's assessment.

3.4 Discussion

With respect to the nonpertinent and silly norm conditions, the results paint a clear picture: Whereas in the between-subjects comparisons (Experiment 1) there were significant and medium-to-large effects for causation and the moral variables, the effects vanished in the within-subjects comparisons (Experiment 2), see Table 1 and Figure 5. Upon reflection, more than two thirds of the participants did not judge nonpertinent and silly norm-violating agents differently from norm-conforming ones (see Appendix Section 3.2, Table 10). This suggests that, according to lay participants themselves, violations of nonpertinent and silly norms – at least if these are evident as the only difference across cases – should *not* impact causation judgments, and hence the effects that arise in between-subjects designs must be considered a bias. These findings for causation track the results of between- and within-subjects contrasts on *mens rea* attributions reported by Kneer & Machery (2019) and Kneer & Skoczen (2023). Here too, pronounced between-subjects effects of outcome on negligence and blame disappear once people see both cases side-by-side, suggesting an outcome bias.

The situation is more complex in the case of pertinent norms and allows interesting insights into the *Norm Effect* as discussed in the literature more generally. Here, too, we find a reduction in effect size across all variables. The effect on causation, for instance, is reduced from large (d = 1.01) to medium-sized (d = .70). Additionally, one third of the participants gave identical ratings to the causation and blame questions across the no norm and pertinent norm conditions (Appendix Section 3.2, Table 9). Inciting reflective judgement via a within-subjects design thus *weakens* the Norm Effect.

Nevertheless, a residual – and considerable – effect persists. When it comes to pertinent norms, participants judge the norm-violating agent as more causal and blameworthy, even in direct comparison to a norm-adhering agent. Furthermore, as the Responsibility View predicts, there is a strong correlation between perceived causation and moral responsibility, both in the no norm condition (r = .73), and the pertinent norm condition (r = .55) (see Appendix Section 3.2).

4. Experiment 3

In order to explore whether our findings up to this point replicate, we ran another experiment with a different scenario. More precisely, we aimed to increase the external validity of the results so far, namely (i) the curious – and pronounced – effects for nonpertinent and silly norm infractions on causality and blame attributions in between-subjects designs (Exp. 1), (ii) their independence from *mens rea* mediators invoked by proponents of the Responsibility View (Exp. 1), and (iii) the substantial decrease in effect size in within-subjects designs and hence their interpretation as evident biases (Exp. 2). Since the experiment – or rather, two experiments – are very similar to Experiments 1 and 2, we will be relatively concise.

4.1 Participants

We recruited participants for the two sub-experiments separately on Amazon Mechanical Turk, restricting the IP address to the United States. For the between-subjects design, 283 participants were recruited. As preregistered,[8] we excluded participants who failed an attention check, spent less than 15 seconds reading the vignette, gave a wrong answer to the comprehension question, or were not native English speakers. 212 participants remained (female: 47%; mean age: 43 years, SD = 13 years, range: 22–75 years). For the within-subjects design, 396 participants were recruited. In line with our preregistration criteria,[9] we excluded participants who failed an attention

[8] Available under https://aspredicted.org/DIZ_UZQ.

[9] Available under https://aspredicted.org/KJQ_FNW.

check, spent less than 25 seconds reading the vignette (which was longer than in the between-subjects design), or were not native English speakers. 354 participants remained (female: 50%; mean age: 44 years, SD = 13 years, range: 20–76 years).

4.2 Methods and materials

Participants received a short vignette based on a German Imperial Court of Justice (*Reichsgericht*) case.[10] In the scenario Mark places several trash bags outside his apartment building. Nearby, construction workers are cutting concrete with a buzz saw. The sparks light the trash bags ablaze and the apartment building caught fire, injuring several tenants. (The complete *Trash bag* scenario can be found in the Appendix, Section 4.2).

In the no norm condition, Mark was free to store his trash bags at the building's entrance. In the pertinent norm condition, city regulations prohibited the storing of objects near building entrances. In the nonpertinent norm condition, although the city required its citizens to use blue trash bags, Mark continued to use grey ones, which were identical in all properties but colour. In the silly norm condition, due to the abundance of sailors living in Mark's apartment building, all tenants were required to tie their trash bags with a special sailor's knot, which Mark did not do.

Participants were, as in the previous experiments, asked to rate the causal contributions of the primary agent, Mark, and the secondary agents (the workers). They were further asked to judge Mark's state of mind (knowledge and desire), as well as the moral variables of blame, responsibility, and punishment. All items were presented on 7-point Likert scales as in Experiments 1 and 2.

In the *between-subjects* design, participants were randomly assigned to one of the four norm conditions (no norm, pertinent norm, nonpertinent norm, silly norm). In the *within-subjects* design, participants were randomly assigned to pairs of scenarios contrasting the no norm condition with one of the three norm conditions on the same screen (as in Experiment 2). They were asked to rate the causal contributions, mental states as well as blame, responsibility and punishment for each of the agents of the two scenarios.

4.3 Results

4.3.1 General Results

Between-subjects design. One-way ANOVAs investigating the influence of norm type (no norm v. pertinent norm v. nonpertinent norm v. silly norm) revealed a significant main effect on the causal contributions of Mark and the moral variables of blame, responsibility, and punishment (all ps < .001), with large effect sizes throughout (η_p^2s > .242). The effect of norm type on knowledge and desire proved nonsignificant, though knowledge was close (p = .058).

[10] RGSt 61, 318.

Within-subjects design. We ran repeated-measures ANOVAs to explore the influence of the three types of norms (pertinent v. nonpertinent v. silly) on the dependent variables. Aggregating across the three norm type conditions, we found participants' causal ascriptions to differ significantly with respect to the primary agent, Mark ($p < .001$, $\eta_p^2 = .163$, a large effect). The difference in mental state ascriptions was small (η_p^2s < .048) and reached significance only for knowledge ($p < .001$). The moral variables, on the other hand, all differed significantly and pronouncedly across conditions (all *p*s < .001, all η_p^2s > .132).

4.3.2 Planned Comparisons

For each design, we ran planned comparisons for a more detailed breakdown of the impact of norm type on the dependent variables, see Table 2 (complete tables in Appendix Section 3.3).

No norm v. nonpertinent norm. In the between-subjects design, the pertinent norm significantly increased attributions of causality, blame, responsibility and deserved punishment (all *p*s < .001, all *d*s > 1.73). In the within-subjects design, we also found a significant and pronounced norm effect for these variables (all *p*s <. 001). The effect size – though it remained substantial (all *ds* > .78) – was reduced to about half for said DVs. There was no significant norm effect on desire in either design (*p*s > .057, *ds* < .18), whereas there was a small-to-medium sized effect on knowledge in both (*p*s < .016, *ds* < .49). In the within-subjects design, the proportion of identical responses for causation was 42%, for blame 31%.

		Between-subjects					Within-subjects				
Contrast	*Variable*	*df*	*t*	*p*	Cohen's *d*	95% CI	*df*	*t*	*p*	Cohen's *d*	95% CI
NN v. PN	Causation Primary	104	–9.52	<.001	1.85	[–2.30;–1.3	123	–8.79	<.001	0.79	[–.99;–.59]
	Knowledge	94	–2.49	0.015	0.48	[–.87;–.10]	123	–4.77	<.001	0.43	[–.61;–.24]
	Desire	104	0.21	0.834	0.04	[–.34;.42]	123	–1.91	0.058	0.17	[–.39;–.01]
	Blame	104	–11.99	<.001	2.33	[–2.81;–1.8	123	–10.33	<.001	0.93	[–1.14;–.76]
NN v. NP	Causation Primary	92	–5.45	<.001	1.08	[–1.50;–.67]	115	–4.33	<.001	0.40	[–.59;–.21]
	Knowledge	101	–.64	0.525	0.13	[–.51;.26]	115	–.63	0.529	0.06	[–.24;.12]
	Desire	101	0.71	0.478	0.14	[–.25;.53]	115	–.58	0.566	0.05	[–.24;.13]
	Blame	86	–6.29	<.001	1.25	[–1.67;–.83]	115	–4.05	<.001	0.38	[–.56;–.19]
NN v. SN	Causation Primary	107	–4.64	<.001	0.89	[–1.28;–.49]	113	–1.34	0.184	0.13	[–.31;.06]
	Knowledge	106	–1.70	0.092	0.33	[–.70;.05]	113	0.00	1.000	0.00	[–.18;.18]
	Desire	107	–.63	0.531	0.12	[–.50;.26]	113	–.20	0.842	0.02	[–.20;.17]
	Blame	107	–5.59	<.001	1.07	[–1.47;–.67]	113	–1.93	0.056	0.18	[–.37;.004]

Table 2: Comparison of effect sizes for the no norm v. pertinent norm (NN v. PN), nonpertinent norm (NN v. NP), and silly norm (NN v. SN) conditions across ascriptions of causation, mental states, and blame. 95% Confidence Intervals (CIs) for the reported *d*-values.

No norm v. nonpertinent norm. In the between-subjects design, the nonpertinent norm significantly increased attributions of causality, blame, responsibility and deserved punishment (all *p*s < .001, all *d*s > 1.07, large effects). In the within-subjects design, we also found a significant, yet much smaller norm effect for these variables (all *p*s <. 001, *ds* > .34); the effect size was reduced to at most half of the between-subjects effect. We could find no significant effect on desire or knowledge in either design (*ps* >. 477). In the within-subjects design, the proportion of identical responses for causation was 74%, for blame 78%.

No norm v. silly norm. In the between-subjects design, the silly norm significantly increased attributions of causality, blame, responsibility and deserved punishment (all *p*s < .001, all *d*s > .88, large effects). In the within-subjects design, none of the effects reached significance (all *ps* > .055, all *ds* < .19). We could find no significant effect on desire or knowledge in either design (*ps* > .091). In the within-subjects design, the proportion of identical responses for causation and blame were 78%.

4.3.3 Meta-Analysis of Effects across Designs

In the within-subjects designs, the vast majority of participants did *not* perceive a difference in causality due to the violation of nonpertinent or silly norms as compared to the no norm condition. In order to provide statistical support for the claim that the effect sizes in the within-subjects design were significantly smaller than for the between-subjects design, we ran meta-analyses contrasting the results of the two design types from Experiments 1–3. Figure 6 presents the mean effects of norm status on all DVs for all three contrasting pairs, estimated with the restricted maximum-likelihood method based on a random effects model (see Viechtbauer, 2010). As shown, for the nonpertinent and silly norm, but not the pertinent norm contrast with the no norm condition, the effect of norm status on causation and blame was significantly and substantially reduced in the within-subjects design. These findings are consistent with the difference in proportions of identical responses for the pertinent norm/no norm contrast (low proportion) and the other two contrasts (high proportion) reported in the previous section.

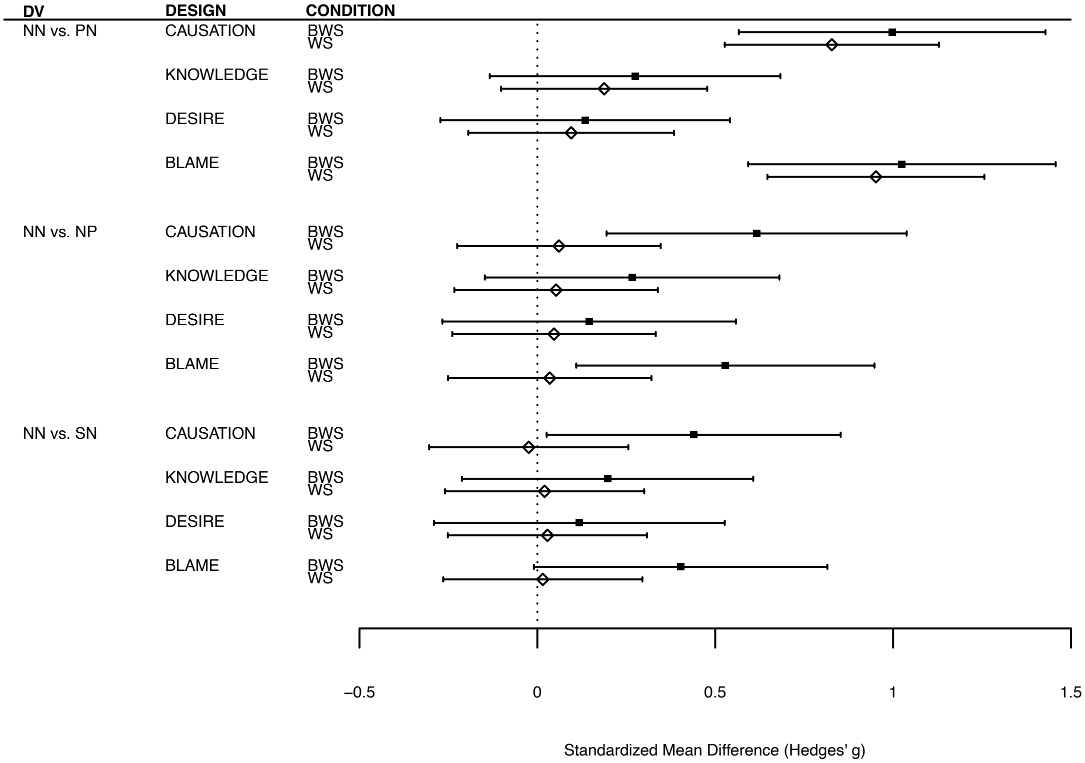

Figure 6: Effects of norm status on the dependent variables across designs in a random effects model in terms of Hedges' *g* (Hedges, 1981, Hedges & Olkin, 1985). Error bars denote 95% confidence intervals.

4.4 Discussion

All our results from Experiments 1 and 2 replicated. Despite their normative absurdity, nonpertinent and silly norms had a large effect on perceived causality, blame and the other moral DVs. This, we take it, constitutes a serious problem for the Responsibility View. Pointing to an indirect effect via the mediators knowledge and desire is not an option as the difference *vis-à-vis* the no norm condition was either nonsignificant or very limited in size. Importantly, the effect of the silly norm disappears entirely in the within-subjects design, and, for the nonpertinent norm is drastically reduced from a large effect (d = 1.08) in the between-subjects design to a small one in the within-subjects design (d = .40), and driven by a minority of participants (about 20%, the rest judge the two cases identically). This shows that, in conditions that encourage reflective judgment (having the two cases side-by-side), people do *not* view such norms as relevant to causal judgment (and the same, by and large, holds for blame and responsibility). As regards the pertinent norm: The very large effects measured in the between-subjects design on causation (d = 1.85) and blame (d = 2.33) are reduced to about half in the within-subjects designs (d = .79, d = .93), but remain borderline large. Hence, their extraordinary between-subjects size is presumably at least partially driven by bias. However, the within-subjects effects and the fact that about 60% of participants rate causation differently in the no norm v. pertinent norm contrast suggest that the majority of people *do* think that pertinent norms *are* relevant to the assessment of causation (and the same holds for blame).

Proponents of the Responsibility View might argue that, despite following Sytsma (2019b), the mediators tested might not have been the most appropriate ones. Since our vignettes involve accidents, it comes as no surprise that participants do not ascribe knowledge or desire to the agent. And indeed, mean knowledge and desire attribution are extremely low in both our experiments using the *Festival* vignette (Studies 1 and 2, all Ms < 2.50, significantly below the midpoint of the scale, all *ps* < .001), and those using the *Trash Bag* vignette (Study 3, all Ms < 2.24, significantly below the midpoint, all *ps* < .001). Instead of the inculpating mental states of knowledge and desire to harm, one might want to test the carelessness or negligence of the agent, which is determined in relation to how reasonably foreseeable the accident was (Engelmann & Waldmann, 2021, 2022; Kirfel & Lagnado, 2021a, 2021b; Lagnado & Channon, 2008; Kneer & Machery, 2019; Kneer, 2022; Nobes & Martin, 2022; Sarin & Cushman, 2022, Murray et al. 2023). It could turn out that participants judge agents that violate norms – even nonpertinent or silly ones – as acting more negligently than their norm-adhering counterparts and thus rightfully consider them more responsible. Deducing the scope of an agent's causal reach from what can reasonably be foreseen is, furthermore, common practice in the law, which holds that an agent can only be held liable for a harmful outcome if said agent could have reasonably foreseen it (Dressler, 2015; Goldberg & Zipursky, 2010; Owen, 2009). In Experiments 4 and 5 we explore whether the Responsibility View can be saved by recourse to foreseeability as an alternative mediator.

5. Experiment 4

Experiment 4 investigates whether the findings of the previous experiments can be explained by recourse to a different potential mediator, namely the foreseeability of an accident. Quite independently of the narrower concerns of the academic debate, this question is of central relevance to the law, as the reasonable foreseeability of an accident is *the* key desideratum in negligence and recklessness attribution. Evidently, from the legal perspective nonpertinent and silly norm violations should not impact foreseeability (for discussion, see e.g. Brown, 2023; Margoni & Brown, 2023; Green, 1961; VerSteeg, 2011). If they did, this would clearly constitute a bias. And if biased foreseeability would – in line with the predictions of the Responsibility View – influence causal responsibility, then the biasing norm-effect would distort the adequate assessment of *both mens rea* (the "guilty mind") and *actus reus* (the "guilty act").

In order to account for the hindsight bias which frequently besets foreseeability judgements (Kamin & Rachlinski, 1995; Margoni & Surian, 2022; Kneer & Skoczen, 2023; Margoni & Brown, 2023; Rachlinski, 1998, 2000; for a review, see Roese & Vohs, 2012), we presented participants with both *ex ante* (outcome information not yet available) and *ex post* (outcome information available) conditions of the *Trash bag* vignette which was introduced in Experiment 3.

5.1 Participants

We recruited 1014 participants on Prolific. Their IP address was restricted to the United States. In line with our preregistration criteria,[11] we excluded participants who failed a general attention check, spent less than 15 seconds reading the vignette, or were not native English speakers. 960 participants remained (female: 46%; mean age: 42 years, SD = 13 years, range = 19–94 years).

5.2 Methods and materials

The study took a 4 (norm type: no norm v. pertinent norm v. nonpertinent norm v. silly norm) × 2 (presentation of foreseeability question: *ex ante* v. *ex post*) between-subjects design. Participants were randomly assigned to one of the four conditions of the *Trash bag* vignette from Exp. 3.

Participants in the *ex post* conditions were given the vignette in full (*i.e.* including the building catching fire), and asked the exact same questions as in Experiment 3, except that we replaced the foreknowledge and desire questions with a single foreseeability question. It read:

> **Foreseeability:** To what extent do you agree or disagree with the following statement: "Mark could have reasonably foreseen the coming about of injuries." (1 = completely disagree, 7 = completely agree)

Participants in the *ex ante* conditions were given the vignette only up to the mention of Mark placing his trash bags outside and were asked to make an initial evaluation as to the foreseeability of an accident. Afterwards, participants were told about the accident and asked to rate the causal contributions of Mark and the workers and assess the moral variables.

5.3 Results

ANOVAs. We ran a series of 4 (norm type) × 2 (presentation order) between-subjects ANOVAs for all dependent variables. As regards foreseeability, we found a significant and moderately-sized main effect of norm type ($p < .001$, $\eta_p^2 = .079$) as well as a large effect of presentation order ($p < .001$, $\eta_p^2 = .162$). The interaction was nonsignificant ($p = .170$, $\eta_p^2 = .005$).

For causation, we found a significant and large effect of norm type ($p < .001$, $\eta_p^2 = .165$), though neither presentation order ($p = .105$, $\eta_p^2 = .003$), nor the interaction ($p = .400$, $\eta_p^2 = .003$) seemed to have influenced participants' judgements. For our moral variables of blame, responsibility, and punishment, we found a large effect of norm type throughout (blame: $p < .001$, $\eta_p^2 = .177$; responsibility: $p < .001$, $\eta_p^2 = .166$; punishment: $p < .001$, $\eta_p^2 = .199$, while both the presentation order and interaction remained nonsignificant (all *ps* > .174) (see Appendix Section 3.5 for full results).

[11] Available under https://aspredicted.org/BPB_3YS.

		Ex ante					Ex post				
Contrast	*Variable*	*df*	*t*	*p*	Cohen's *d*	95% CI	*df*	*t*	*p*	Cohen's *d*	95% CI
NN v. PN	Causation Primary	242	–9.58	<.001	1.25	[–1.52;–.97]	182	–11.53	<.001	1.58	[–1.89;–1.27]
	Foreseeability	242	–6.05	<.001	0.79	[–1.05;–.52]	210	–7.27	<.001	1.00	[–1.28;–.71]
	Blame	229	–10.57	<.001	1.35	[–1.63;–1.07]	196	–11.52	<.001	1.58	[–1.89;–1.27]
NN v. NP	Causation Primary	285	–6.04	<.001	0.71	[–.95;–.47]	216	–4.33	<.001	0.59	[–.86;–.32]
	Foreseeability	285	–.24	0.815	0.03	[–.26;.20]	216	–2.54	0.012	0.34	[–.61;–.08]
	Blame	285	–6.74	<.001	0.80	[–1.04;–.56]	216	–4.42	<.001	0.60	[–.87;–.33]
NN v. SN	Causation Primary	281	–6.44	<.001	0.77	[–1.01;–.52]	216	–4.34	<.001	0.59	[–.86;–.32]
	Foreseeability	281	–.51	0.612	0.06	[–.29;.17]	216	–2.03	0.044	0.28	[–.54;–.01]
	Blame	281	–6.30	<.001	0.75	[–.99;–.51]	216	–4.66	<.001	0.63	[–.90;–.36]

Table 3: Comparison of effect sizes for the no norm v. pertinent norm (NN v. PN), nonpertinent norm (NN v. NP), and silly norm (NN v. SN) conditions across ascriptions of causation, mental states, and blame. 95% Confidence Intervals (CIs) for the reported *d*-values.

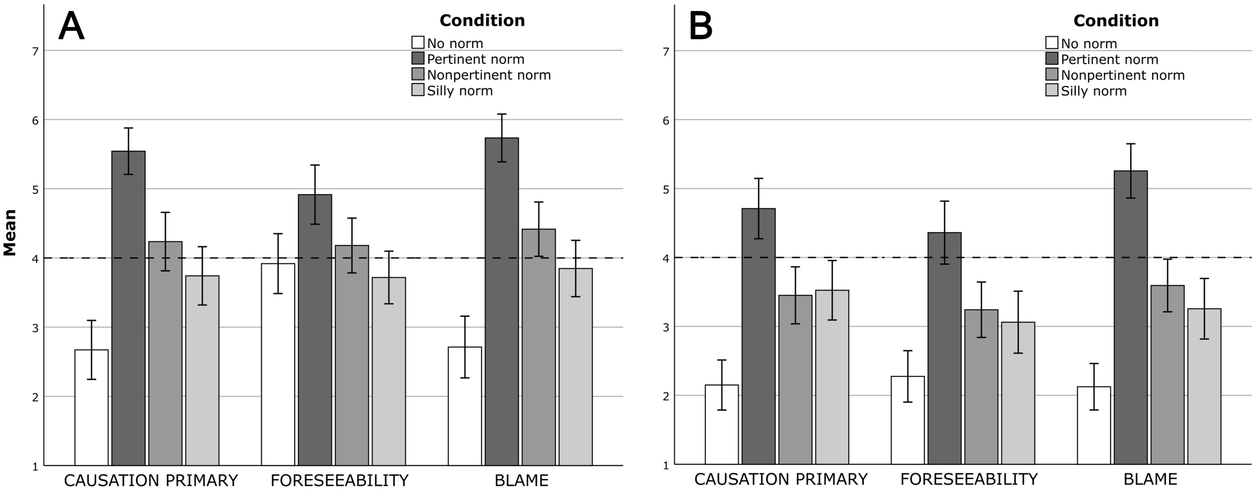

Figure 7: Comparison of means across *ex ante* (A) and *ex post* (B) conditions. Error bars denote 95% CI.

No norm v. pertinent norm. In comparing the no norm and pertinent norm conditions, we found significant differences in foreseeability, Mark's causal contribution, as well as the moral variables across both *ex ante* and *ex post* presentation order (all *p*s < .001). The effect sizes in the *ex post* condition were considerably larger (all *ds* > 1.00) than in the *ex ante* condition (all *ds* > .78), consistent with recent work on the hindsight bias afflicting *mens rea* attribution and related variables (see Kneer & Skoczen, 2023).

No norm v. nonpertinent norm. In the *ex post* condition, participants judged Mark's causal contribution as well as the moral variables to differ significantly across norm type (all *p*s < .001, all *d*s > .56). The effect of norm type on foreseeability was significant (*p* = .012, *d* = .34) – and one might thus consider it as a mediator, which renders the norm effect on causation plausible according to the logic of the Responsibility View. However, this might be too quick. In the *ex ante* conditions, norm type also significantly impacts causation and the moral DVs (all *p*s < .001, all *d*s > .69). Crucially, however, we did *not* find a difference in foreseeability (*p* = .815) suggesting –

as long as one avoids distortion due to hindsight bias – that recourse to foreseeability (just as the other *mentes reae* tested in Studies 1–3), cannot rehabilitate the Responsibility View.

No norm v. silly norm. The results for the no norm/silly norm contrast replicate the pattern just reported in the previous section. Whereas participants in the *ex ante* conditions perceived a difference in Mark's causal contribution as well as the moral variables (all *p*s < .001, all *d*s > .69) but *not* in foreseeability (*p* = .612), participants in the *ex post* conditions judged differently all aforementioned dependent variables (all *p*s < .001, all *d*s > .58), including, though just about, the foreseeability of an accident (*p* = .044, *d* = .28).

5.4 Discussion

Consistent with the results of Experiments 1–3, the effects of the *pertinent* norm on causation and the moral variables were significant and very pronounced (all *ds* > 1.18). In this condition, the norm effect on foreseeability, too, even when assessed *ex ante*, was significant and close-to-large (*d* = .79). This squares well with Sytsma's proposal, according to which the influence of *prima facie* irrelevant factors such as norm violation on causal responsibility can be explained by aid of a mediator such as foreseeability: norm violations, one might argue, *should* impact foreseeability, and thereby moral responsibility and blame. Given the tight connection between moral and causal responsibility, and the fact that norm violations, after all, are not "peripheral" to the former, their impact on perceived causal responsibility is explained.

For the Responsibility View, things are considerably more problematic as regards nonpertinent and silly norms. Replicating the findings from Experiments 1–3, we again found a significant and considerable impact on perceived causation and the moral variables (all *ds* > .58). When the reasonable foreseeability of an accident was assessed *ex ante*, it proved insensitive to nonpertinent and silly norm violations. For nonpertinent and silly norms, then, attempts to rehabilitate the Responsibility View by aid of plausible mediators such as foreseeability, knowledge and/or desire have thus far all failed.

Going beyond the debate on causation (which constitutes our primary concern), the fact that foreseeability, when assessed *ex post*, is responsive to norm effects is an interesting, novel and worrisome finding: in the law, juries are to judge foreseeability with respect to the agent's circumstances and epistemic situation (i.e. in an *ex ante* fashion). As our results show, the hindsight bias might make this difficult, just as it distorts a whole range of other variables relevant to negligence attribution (see Kneer & Machery, 2019; Kneer & Skoczen, 2023).

6. Experiment 5

The fact that nonpertinent and silly norms impact causation, and the by now third possible mediator – foreseeability – is of no help to explain this effect is problematic for the Responsibility View. To explore whether this finding generalizes beyond the *Trash Bag* scenario which we have used, we ran the experiment with the novel *Shooting range* vignette (full scenario in Appendix Section 4.3).

6.1 Participants

1034 participants were recruited online on Amazon Mechanical Turk. Their IP address was restricted to the United States. As preregistered,[12] we excluded participants who failed a general attention check, spent less than 10 seconds on the page presenting the vignette, or were not native English speakers. 680 participants remained (female: 49%; mean age: 42 years, SD = 13 years, range = 20–94 years).

6.2 Methods and materials

Just like Experiment 4, the study took a 4 (norm type: no norm v. pertinent norm v. nonpertinent norm v. silly norm) × 2 (presentation of foreseeability question: *ex ante* v. *ex post*) between-subjects design. Participants were randomly assigned to one of the eight conditions of the *Shooting range* vignette. The story has Mark shooting at an outdoor shooting range while Lauren is hiking in the nearby forest. The sudden appearance of a wild boar frightens Lauren, who tumbles down a hill and comes to halt right in front of the bullet Mark shot moments earlier. The bullet lodges itself in her leg and Lauren has to be taken to the hospital.

The no norm condition mentions a shooting range in regular operation. In the pertinent norm condition, Mark practices at the shooting range although it's closed. In the nonpertinent norm condition, it is prohibited to use the shooting range unless one wears protective gloves and glasses, and Mark does not wear any. In the silly norm condition, it is forbidden to bring any type of food or drinks to the shooting range, and Mark sneaks in a bag of potato chips and a soft drink.

Participants in the *ex post* conditions were given the vignette in full (i.e. including the injury of Lauren). Participants in the *ex ante* conditions were given the vignette only up to the mention of Lauren hiking and were asked to make an initial evaluation as to the foreseeability of an accident. Afterwards, participants were told about the accident and asked to rate the causal contributions of Mark and the boar and assess the moral variables.

In the *ex post* conditions, participants – having read the full vignette – were asked to rate the causal contributions of Mark and the boar, before turning to an *ex post* assessment of the foreseeability of the accident, followed by the three moral variables. The questions were phrased as in the experiments above, and responses were recorded on 7-point Likert scales.

6.3 Results

ANOVAs. A 4 (norm type) × 2 (presentation order) between-subjects ANOVA revealed a significant main effect of both order and norm type on foreseeability (both *p*s < .001), with a small-

[12] Available under https://aspredicted.org/B7H_QXS.

to-moderate effect size for order ($\eta_p^2 = .057$) and a moderate effect size for norm type ($\eta_p^2 = .084$), see Figure 5. The interaction was close to significant ($p = .053$). The main effect of norm type on Mark's perceived causal contribution was significant and large ($p < .001$, $\eta_p^2 = .204$) and was accompanied by a significant yet small effect of presentation order ($p < .001$, $\eta_p^2 = .024$). We further found significant and large main effects of norm type on all moral variables (all *p*s < .001, all η_p^2s > .256), and small main effects for presentation order (all *p*s < .007, all η_p^2s < .029).

		Ex ante					Ex post				
Contrast	*Variable*	*df*	*t*	*p*	Cohen's *d*	95% CI	*df*	*t*	*p*	Cohen's *d*	95% CI
NN v. PN	Causation Primary	165	–10.69	<.001	1.67	[–2.02;–1.31]	160	–8.86	<.001	1.38	[–1.72;–1.04]
	Foreseeability	165	–3.21	0.002	0.50	[–.81;–.19]	160	–7.03	<.001	1.08	[–1.41;–.76]
	Blame	145	–10.67	<.001	1.69	[–2.04;–1.33]	164	–11.94	<.001	1.86	[–2.22;–1.49]
NN v. NP	Causation Primary	160	–5.14	<.001	0.81	[–1.13;–.49]	168	–4.70	<.001	0.71	[–1.02;–.40]
	Foreseeability	160	–.88	0.376	0.14	[–.45;.17]	169	–3.48	0.001	0.53	[–.84;–.23]
	Blame	160	–5.72	<.001	0.90	[–1.23;–.58]	168	–5.74	<.001	0.87	[–1.18;–.55]
NN v. SN	Causation Primary	156	–3.54	0.001	0.57	[–.88;–.25]	156	–4.85	<.001	0.76	[–1.08;–.44]
	Foreseeability	156	0.70	0.488	0.11	[–.20;.42]	155	–2.68	0.008	0.42	[–.73;–.11]
	Blame	156	–3.75	<.001	0.60	[–.92;–.28]	151	–4.07	<.001	0.64	[–.95;–.32]

Table 4: Comparison of effect sizes for the no norm v. pertinent norm (NN v. PN), nonpertinent norm (NN v. NP), and silly norm (NN v. SN) conditions across ascriptions of causation, mental states, and blame. 95% Confidence Intervals (CIs) for the reported *d*-values.

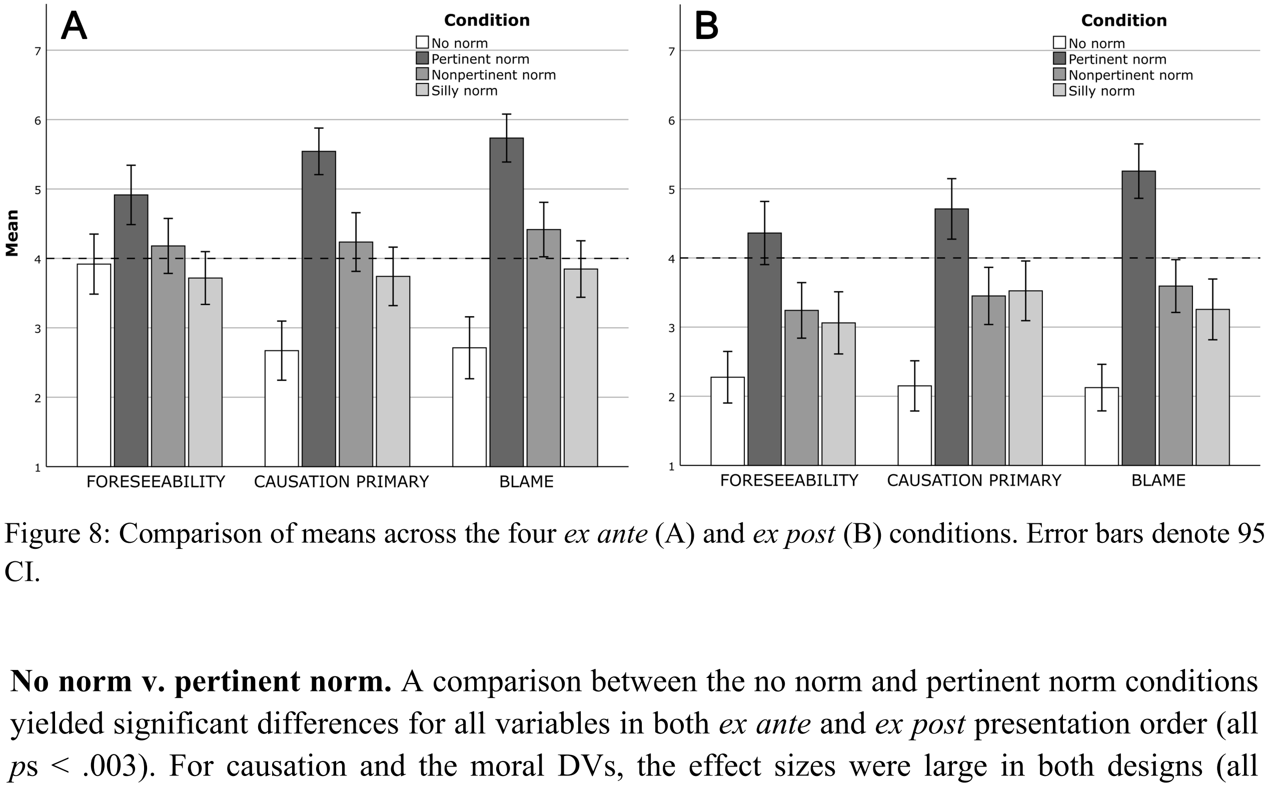

Figure 8: Comparison of means across the four *ex ante* (A) and *ex post* (B) conditions. Error bars denote 95% CI.

No norm v. pertinent norm. A comparison between the no norm and pertinent norm conditions yielded significant differences for all variables in both *ex ante* and *ex post* presentation order (all *p*s < .003). For causation and the moral DVs, the effect sizes were large in both designs (all *ds* > 1.37). For foreseeability, the effect size was large in the *ex post* condition (*d* = 1.08), though significantly smaller in the *ex ante* condition (*d* = .50). It is doubtful that the very large effect of norm status on causation (*d* = 1.67) in this design can exhaustively be explained by the only mid-sized effect on foreseeability.

No norm v. nonpertinent norm. In the *ex post* conditions, there was a significant norm effect on causation, the moral variables and foreseeability (all *ps* < .001, all *ds* > .52). In the *ex ante* conditions, we also found significant (all *ps* < .001) and sizeable effects on causation (*d* = .81) and the moral DVs (all *ds* > .84). Importantly, however, there was no significant effect of norm status on foreseeability (*p* = .376), suggesting that the significant and large norm effect on causation is not mediated by foreseeability once the hindsight bias is corrected for.

No norm v. silly norm. In the *ex post* condition, we again found a significant norm effect on causation, foreseeability and the moral variables (all *ps* < .001, all *ds* > .41). Correcting for the hindsight bias once again rendered foreseeability nonsignificant in the *ex ante* condition (*p* = .488), which suggests that it is of no use to explain the significant norm effect on causation (*p* < .001, *d* = .57) or the moral variables (all *ps* < .01, all *ds* > .41).

6.4 Discussion

Experiment 5, beyond replicating all key findings of Experiment 4, puts a little more pressure on the Responsibility View in the no norm v. pertinent norm contrast. The large norm effect on causation (*d* = 1.67) in the *ex ante* condition cannot be exhaustively explained by the significant, though only mid-sized impact of norm status on foreseeability (*d* = .50). Hence, a considerable direct effect of norm status on perceived causation is unaccounted for. Once again, the nonpertinent and silly norm effects on perceived causation in the *ex ante* conditions cannot be explained by aid of an inferred difference in the foreseeability of an accident, because the latter was nonsignificant.

Finally, the drastic difference in foreseeability judgements *ex post* vis-à-vis *ex ante* (up to a Cohen's *d* of a .50 difference) points to the hindsight bias: the tendency for an event to be deemed more predicable or probable after one has learned that the event has in fact occurred. Thus, while foreseeability could mediate responsibility – and by extension causation – in the *ex post* conditions, the foreseeability judgements themselves are due to hindsight bias and thus can do little to render the Responsibility View more plausible.

7. General Discussion

7.1 Responsibility View v. Bias View

According to the Responsibility View, the folk concept of causation is strongly intertwined with moral responsibility. On this account, factors which legitimately increase the attribution of moral responsibility, such as the foreknowledge of harm, or the agent's desire to harm, can be viewed as legitimately increasing perceived causation. The Bias View, by contrast, takes the concept of causation to be nonnormative. Cases where moral factors increase perceived causation testify to a performance error of human judgment: people are inclined to blame an agent who causes harm more than one who doesn't and, in an attempt of post-hoc rationalization, exaggerate her causal contribution.

Advocates of the Responsibility View acknowledge that not just any factor that *could* influence perceived moral responsibility *should* influence perceived causality. Only factors that are legitimately connected to moral responsibility proper are viewed as exerting a warranted impact on perceived causality. This excludes, for instance, the agent's race, gender, or moral character. It also excludes the violations of norms, which are unrelated to a specific action's outcome. Importantly, there are exceptions: if certain features that are *prima facie* irrelevant to moral responsibility, such as general moral character, engender reasonable inferences to factors which *are* relevant (such as e.g. *mens rea*), advocates of the Responsibility View argue, this should not be considered as evidence against the account.

We have explored two challenges to the Responsibility View. First, we have shown that the violation of nonpertinent and silly norms unconnected to the resulting harm have a significant and considerable impact on perceived causality, with medium to large effect sizes. According to the view of all parties involved, they should *not* impact moral responsibility or blame. However, they do, and – in line with Alicke's Bias View – presumably thereby influence perceived causation. Given that potential, reasonable mediators (foreknowledge, desire to harm, foreseeability) of interest proved nonsignificant, it is difficult for the Responsibility View to tell a convincing story here. What is more, in a within-subjects design (Studies 2 and 3) we show that participants *themselves* hold that nonpertinent and silly norms should *not* influence causality attributions: the vast majority of them rated the causal impact (and blame) of the norm-abiding and norm-violating identically. Grist to the mill of the Bias View.

Replicating extant findings, we also found a powerful effect of norm-violations pertinent to the action. Sytsma seems to hold that pertinent norms should *only* exert an influence on perceived causation if it were mediated by reasonable inferences regarding legitimate influences on moral responsibility. However, and this constitutes our second challenge, the large effects (Cohen's *ds* > 1.00) we found for causation in between-subjects designs cannot *exhaustively* explained through inferences regarding *mens rea* (foresight, foreseeability, desire). That said, as Studies 4 and 5 demonstrate, at least foreseeability can account for some of the effect.

We agree with Sytsma's warning that "researchers need to carefully consider and control for the inferences that participants might draw concerning the agents' mental states and motivations" (2019b, p. 25). Our Experiment 5 underlines this requirement further. Scholars who suggest that moral responsibility and causation are driven by a particular inference to *mens rea*, such as negligence, must be careful to distinguish when such an inference is warranted, and when it is not. As our results show, the large effect of pertinent norm violations on negligence *ex post* (*d* = 1.08) which seems to explain the large effect on causation (*d* = 1.38), shrinks to less than half (d = .50) once the hindsight bias is controlled for, and can no longer fully explain the very large effect on perceived causation (*d* = 1.67).

7.2 The Norm Effect

Although the data reported favours the Bias View, in particular as regards the effects exerted by nonpertinent and silly norms, this does not yet mean that the (pertinent) Norm Effect on causality attributions itself constitutes a bias. After all, one might formulate a weaker version of the Responsibility View, according to which the violation of pertinent norms, *even if unmediated by other factors such as mens rea*, exerts a *legitimate* influence on moral responsibility and (therefore) attributed causality. Note that an account of this sort need not collapse into the unattractive *Anything-Goes View*, as long as moral-philosophical reasons are provided why norm-infractions are relevant to the responsibility of the agent – and such reasons do not seem that hard to come by. One interesting data point in favour of a more permissive Responsibility View of Causation is provided by the within-subjects results: in contrast to the nonpertinent and silly norm comparisons, about two-thirds of our participants *do* consider the violation of pertinent norms relevant to the assessment of blame/responsibility and causation.

7.3 Implications

Whether or not the pertinent Norm Effect is considered a bias or not, our results demonstrate that attributions of causality are easily influenced by factors that clearly should not play any role. An agent who fails to adhere to some silly norm that happens to be in place should not be judged more causally responsible than one who does. This is not only the view of any reasonable philosopher or moral psychologist, but consistent with the folk view, as the within-subjects data shows. One area where these findings are of great importance is the law: both in torts and criminal law, the *actus reus* (the "guilty act") is one of the two key determinants of liability besides *mens rea* (the "guilty mind"), and in common law jurisdictions (such as the UK and the US), the *actus reus* – or, simply put: causation – is determined by lay juries (Knobe & Shapiro, 2021; Lagnado & Gerstenberg, 2017; Lagnado, 2021). As Güver & Kneer (2023) have elaborated, legal practitioners tend to hold that the legal notion of causation *corresponds* to the folk notion. Lord Wright, for instance, has stated in a landmark English case that "[c]ausation is to be understood as the man in the street, and not as the scientist or the metaphysician, would understand it."[13] Similarly, Lord Salmon proclaimed that "[w]hat or who caused an event to occur is essentially a practical question of fact which can best be answered by ordinary common sense rather than abstract metaphysical theory". [14] So too the US Supreme Court, which, in the much-cited *Burrage v. United States*, argued that courts should rely on "the common understanding of causation" and explicate causal relations with reference to what it "is natural to say."[15]

If folk attributions of causality are easily influenced by bias – as the nonpertinent/silly norm data across between-subjects and within-subjects design demonstrate – this is problematic for the law:

[13] *Yorkshire Dale Steamship Co Ltd v Minister of War Transport* [1942] AC 691 (HL) 706.

[14] *Alphacell Ltd v Woodward* [1972] A.C. 824, 847.

[15] *Burrage v. United States*, 571 US 204 (2014). For further experimental papers concerning causation from a legal perspective, see *e.g.* Solan & Darley (2001, pp. 271–272); Macleod (2019, pp. 982–985); Tobia (2021, pp. 91–92); Summers (2018, pp. 3–5), Prochownik (2022).

the folk, or "blame amateurs", as Alicke calls them, might simply not be capable of keeping morally irrelevant factors at bay, and exaggerate the causal contribution of those whom they are unwarrantedly inclined to blame for harmful outcomes.[16] This could result in serious overcriminalization of defendants whose behavior was in some morally or legally irrelevant sense objectionable. Note that the problem is not necessarily limited to common law countries, but might extend to civil law countries, where legal decisions are taken by professional judges. Recent research has shown that legal experts fall prey to the same biases as laypeople, for instance when it comes to outcome bias in *mens rea* attribution (Kneer & Bourgeois-Gironde, 2017; Kneer et al. 2023), confirmation bias (Lidén et al. 2019), sympathy bias (Spamann & Klöhn, 2016; Liu & Li, 2019) or hindsight bias (Strohmaier et al. 2021).

7.4 Limitations and Future Research

Our studies are limited to three scenarios, three potential mediators, and all our participants are US Americans. For improved external validity, future work should explore a broader range of vignettes as well as other mediators of interest. Moreover, similar experiments should be run across different cultures and languages, in particular non-WEIRD countries (cf. Henrich et al. 2010; Henrich, 2020), so as to explore whether the findings constitute a general human disposition of judging causality or not. Some of the cross-cultural work in experimental jurisprudence (see e.g. Hannikainen et al. 2021, Hannikainen et al. 2022) and experimental philosophy (see e.g. Knobe, 2023 for a review) has revealed surprising convergence. However, others have documented extensive differences (for a review, see Stich & Machery, 2023).

Given the important legal dimension of our findings, it should be examined whether professional judges are as susceptible to bias in the determination of proximate cause as our lay samples (in particular in civil law jurisdictions, where experts decide the matter).

Finally, scholars working in experimental jurisprudence should investigate whether folk judgments of the *actus reus* (also with respect to a number of other problematic effects) could be debiased, and suggest concrete and practicable strategies that common law courts could implement.

8. Conclusion

The Responsibility View and the Bias View of causation come apart in their treatment of factors peripheral to moral responsibility: The former, unlike the latter, predicts that such factors will not influence folk causality judgments. In five experiments, we have shown that peripheral factors such as nonpertinent and silly norm violations *do* have a pronounced impact on perceived causation, and that these effects cannot be explained by recourse to potentially legitimate

[16] Our conclusion here is less radical than the one put forth by Rose (2017, 1352), who argues that the "discussion over actual causation should be liberated from any demanded conformity with the folk intuitions" and that "in the dispute over actual causation, folk intuitions deserve to be rejected." For Rose, the folk notion is too unstable and confused to contribute to *any* reasonable account of causation. While we are not unsympathetic to this view, we presently only want to suggest that folk *attributions* of causation are easily and uncontroversially influenced by biasing factors, and that the law must be alert to this fact.

responsibility-enhancing factors such as desire, foreknowledge, or foreseeability. Our results provide strong evidence in favour of the Bias View, and they call into question the Responsibility View of causation.

The status of the (pertinent) *Norm Effect*, as it is standardly explored in the literature, requires further examination. According to Sytsma's formulation of the Responsibility View, the (pertinent) Norm Effect can be regarded "peripheral" if it is not mediated by a nonperipheral factors such as negligence or foreseeability. But given that norm-adherence is quite tightly connected to moral responsibility, this stringent criterion could be dropped without the Responsibility View loosing much of its plausibility. (It will still have problems with nonpertinent and silly norms). Naturally, the Bias View, as well as other recent accounts of the Norm Effect, also have plausible explanations of the effect on offer. It thus seems that the debate concerning the status of the Norm Effect might, by now, be a predominantly theoretical one which depends strongly on the plausibility of the assumptions invoked, and less on what can be elucidated by further experimental inquiry.

Acknowledgments

We would like to thank [REDACTED]. Work on this project was supported by [REDACTED].

References

Alicke, M. D. (1992). Culpable Causation. *Journal of Personality and Social Psychology*, *63*(3), 368–378.

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

Alicke, M. D. (2008). Blaming Badly. *Journal of Cognition and Culture*, *8*(1–2), 179–186.

Alicke, M. D., & Rose, D. (2012). Culpable Control and Deviant Causal Chains. *Personality and Social Psychology Compass*, *6*(10), 723–735.

Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, Norm Violation, and Culpable Control. *The Journal of Philosophy*, *108*(12), 670–696.

Baron, J. (2008). *Thinking and Deciding* (4th ed.). Cambridge University Press.

Bebb, J., & Beebee, H. (2024). Causal selection and egalitarianism. In *Oxford Studies in Experimental Philosophy* (Vol. 5). Oxford University Press.

Brown, T. R. (2023). Minding Accidents. *University of Colorado Law Review*, *94*(1), 89–148.

Dressler, J. (2015). *Understanding Criminal Law* (7th ed.). LexisNexis.

Engelmann, N. (2022). The role of causal representations in moral judgement [Dissertation]. University of Göttingen. Retrieved from https://ediss.uni-goettingen.de/handle/11858/14231?locale-attribute=de.

Engelmann, N., & Waldmann, M. R. (2021). A Causal Proximity Effect in Moral Judgment. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. Retrieved from https://escholarship.org/uc/item/9hp8q72s.

Engelmann, N., & Waldmann, M. R. (2022). How causal structure, causal strength, and foreseeability affect moral judgments. *Cognition*, *226*, 105167.

Frisch, L. K., Kneer, M., Krueger, J. I., & Ullrich, J. (2021). The effect of outcome severity on moral judgement and interpersonal goals of perpetrators, victims, and bystanders. *European Journal of Social Psychology*, *51*(7), 1158-1171.

Gerstenberg, T., & Icard, T. (2020). Expectations affect physical causation judgments. *Journal of Experimental Psychology: General*, *149*(3), 599–607.

Goldberg, J. C. P., & Zipursky, B. C. (2010). *Torts*. Oxford University Press.

Goulette, V., & Verkampt, F. (2023). Blame-validation: Beyond rationality? Effect of causal link on the relationship between evaluation and causal judgment. *Philosophical Psychology*, online first, 1–20.

Green, L. (1961). Foreseeability in Negligence Law. *Columbia Law Review*, *61*(8), 1401–1424.

Güver, L., & Kneer, M. (2023). Causation and the Silly Norm Effect. In S. Magen & K. Prochownik (Eds.), *Advances in Experimental Philosophy of Law* (pp. 133–168). Bloomsbury Publishing.

Hannikainen, I. R., Tobia, K. P., De Almeida, G. D. F., Donelson, R., Dranseika, V., Kneer, M., ... & Struchiner, N. (2021). Are there cross-cultural legal principles? Modal reasoning uncovers procedural constraints on law. *Cognitive Science*, *45*(8), e13024.

Hannikainen, I. R., Tobia, K. P., de Almeida, G. D. F., Struchiner, N., Kneer, M., Bystranowski, P., ... & Żuradzki, T. (2022). Coordination and expertise foster legal textualism. *Proceedings of the National Academy of Sciences*, *119*(44), e2206531119.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Henne, P. (2023). Experimental Metaphysics: Causation. In A. M. Bauer & S. Kornmesser (Eds.), *The Compact Compendium of Experimental Philosophy*. De Gruyter.

Henne, P., & O'Neill, K. (2022). Double Prevention, Causal Judgments, and Counterfactuals. *Cognitive Science*, *46*(5), e13127.

Henne, P., Kulesza, A., Perez, K., & Houcek, A. (2021). Counterfactual thinking and recency effects in causal judgment. *Cognition*, *212*, 104708.

Henne, P., O'Neill, K., Bello, P., Khemlani, S., & De Brigard, F. (2021). Norms Affect Prospective Causal Judgments. *Cognitive Science*, *45*(1), e12931.

Henrich, J. (2020). *The WEIRDest people in the world: How the West became psychologically peculiar and particularly prosperous*. Penguin UK.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29-29.

Hitchcock, C., & Knobe, J. (2009). Cause and Norm. *The Journal of Philosophy*, *106*(11), 587–612.

Hitchcock, C., & Knobe, J. (2009). Cause and Norm. *The Journal of Philosophy*, *106*(11), 587–612.

Hsee, C. K. (1996). The Evaluability Hypothesis: An Explanation for Preference Reversals between Joint and Separate Evaluations of Alternatives. *Organizational Behavior and Human Decision Processes*, *67*(3), 247–257.

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and Actual Causal Strength. *Cognition*, *161*, 80–93.

Kamin, K. A., & Rachlinski, J. J. (1995). *Ex post ≠ ex ante*: Determining liability in hindsight. *Law and Human Behavior*, *19*(1), 89–104.

Kirfel, L., & Lagnado, D. (2018). Statistical norm effects in causal cognition. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, *40*, 617–622.

Kirfel, L., & Lagnado, D. (2021a). Causal judgments about atypical actions are influenced by agents' epistemic states. *Cognition*, *212*, 104721.

Kirfel, L., & Lagnado, D. (2021b). Causation by Ignorance. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*, 966–972.

Kirfel, L., & Lagnado, D. (2021c). *Changing Minds — Epistemic Interventions in Causal Reasoning*. PsyArXiv. Retrieved from https://doi.org/10.31234/osf.io/db6ms

Kirfel, L., & Phillips, J. (2021). The Impact of Ignorance Beyond Causation: An Experimental Meta-Analysis. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*, *43*, 1595–1601.

Kirfel, L., & Phillips, J. (2023). The pervasive impact of ignorance. *Cognition*, *231*, 105316.

Kirfel, L., Bunk, X., Ro'i, Z., & Gerstenberg, T. (2023). Father, don't forgive them, for they could have known what they're doing. *Proceedings of the 45th Annual Conference of the Cognitive Science Society*, *45*, 980–987.

Kneer, M. (2022). Reasonableness on the Clapham Omnibus: Exploring the outcome-sensitive folk concept of *reasonable*. In *Judicial Decision-Making: Integrating Empirical and Theoretical Perspectives* (pp. 25–48). Cham: Springer International Publishing.

Kneer, M., & Bourgeois-Gironde, S. (2017). Mens Rea Ascription, Expertise and Outcome Effects: Professional judges Surveyed. *Cognition*, *169*, 139–146.

Kneer, M., & Machery, E. (2019). No luck for moral luck. *Cognition*, *182*, 331–348.

Kneer, M., & Skoczeń, I. (2023). Outcome effects, moral luck and the hindsight bias. *Cognition*, *232*, 105258.

Knobe, J. (2023). Difference and robustness in the patterns of philosophical intuition across demographic groups. *Review of Philosophy and Psychology*, 1-21.

Knobe, J., & Fraser, B. (2008). Causal Judgment and Moral Judgment: Two Experiments. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 2* (pp. 441–447). MIT Press.

Knobe, J., & Shapiro, S. (2021). Proximate cause explained. *The University of Chicago Law Review*, *88*(1), 165-236.

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(1), 1–26.

Lagnado, D. A. (2021). *Explaining the Evidence: How the Mind Investigates the World*. Cambridge University Press.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754–770.

Lagnado, D. A., & Gerstenberg, T. (2017). Causation in Legal and Moral Reasoning. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (pp. 565–601). Oxford University Press.

Lidén, M., Gräns, M., & Juslin, P. (2019). 'Guilty, no doubt': detention provoking confirmation bias in judges' guilt assessments and debiasing techniques. *Psychology, Crime & Law*, *25*(3), 219-247.

Liu, J. Z., & Li, X. (2019). Legal techniques for rationalizing biased judicial decisions: Evidence from experiments with real judges. *Journal of Empirical Legal Studies*, *16*(3), 630-670.

Livengood, J., & Rose, D. (2016). Experimental Philosophy and Causal Attribution. In J. Sytsma & W. Buckwalter (Eds.), *A Companion to Experimental Philosophy* (pp. 434–449). Blackwell.

Livengood, J., Sytsma, J., & Rose, D. (2017). Following the FAD: Folk Attributions and Theories of Actual Causation. *Review of Philosophy and Psychology*, *8*(2), 273–294.

Macleod, J. (2019). Ordinary Causation: A Study in Experimental Statutory Interpretation. *Indiana Law Journal*, *93*(3), 957–1030.

Margoni, F., & Brown, T. R. (2023). Jurors use mental state information to assess breach in negligence cases. *Cognition*, *236*, 105442.

Margoni, F., & Surian, L. (2022). Judging accidental harm: Due care and foreseeability of side effects. *Current Psychology*, *41*(12), 8774–8783.

Morris, A., Phillips, J., Gerstenberg, T., & Cushman, F. (2019). Quantitative causal selection patterns in token causation. *PLOS ONE*, *14*(8), e0219704.

Murray, S., Krasich, K., Irving, Z., Nadelhoffer, T., & De Brigard, F. (2023). Mental control and attributions of blame for negligent wrongdoing. *Journal of Experimental Psychology: General*, *152*(1), 120.

Nobes, G., & Martin, J. W. (2022). They should have known better: The roles of negligence and outcome in moral judgements of accidental actions. *British Journal of Psychology*, *113*(2), 370–395.

Olier, J. G., & Kneer, M. (2023). Ordinary causal attributions, norms, and gradability [In preparation].

Owen, D. (2009). Figuring Foreseeability. *Wake Forest Law Review*, *44*(5), 1277–1308.

Prochownik, K. (2022). Causation in the law, and experimental philosophy. In P. Willemsen & A. Wiegmann (Eds.), *Advances in Experimental Philosophy of Causation* (pp. 165–188). Bloomsbury Publishing.

Rachlinski, J. J. (1998). A Positive Psychological Theory of Judging in Hindsight. *The University of Chicago Law Review*, *65*(2), 571–625.

Rachlinski, J. J. (2000). Heuristics and Biases in the Courts: Ignorance or Adaptation? *Oregon Law Review*, *79*(1), 61–102.

Roese, N. J., & Vohs, K. D. (2012). Hindsight Bias. *Perspectives on Psychological Science*, *7*(5), 411–426.

Rogers, R., Alicke, M. D., Taylor, S. G., Rose, D., Davis, T. L., & Bloom, D. (2019). Causal deviance and the ascription of intent and blame. *Philosophical Psychology*, *32*(3), 404–427.

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

Rose, D., & Danks, D. (2012). Causation: Empirical Trends and Future Directions. *Philosophy Compass*, *7*(9), 643–653.

Samland, J., & Waldmann, M. R. (2016). How Prescriptive Norms Influence Causal Inferences. *Cognition*, *156*, 164–176.

Samland, J., Josephs, M., Waldmann, M. R., & Rakoczy, H. (2016). The role of prescriptive norms and knowledge in children's and adults' causal selection. *Journal of Experimental Psychology: General*, *145*(2), 125–130.

Sarin, A., & Cushman, F. (2022). One thought too few: Why we punish negligence [Preprint]. PsyArXiv. Retrieved from https://doi.org/10.31234/osf.io/mj769.

Solan, L. M., & Darley, J. M. (2001). Causation, Contribution, and Legal Liability: An Empirical Study. *Law and Contemporary Problems*, *64*(4), 265–298.

Spamann, H., & Klöhn, L. (2016). Justice is less blind, and less legalistic, than we thought: Evidence from an experiment with real judges. *The Journal of Legal Studies*, *45*(2), 255-280.

Strohmaier, N., Pluut, H., Van den Bos, K., Adriaanse, J., & Vriesendorp, R. (2021). Hindsight bias and outcome bias in judging directors' liability and the role of free will beliefs. *Journal of Applied Social Psychology*, *51*(3), 141-158.

Summers, A. (2018). Common-Sense Causation in the Law. *Oxford Journal of Legal Studies*, *38*(4), 793–821.

Stich, S. P., & Machery, E. (2023). Demographic differences in philosophical intuition: A reply to Joshua Knobe. *Review of Philosophy and Psychology*, *14*(2), 401-434.

Sytsma, J. (2019a). Structure and norms: Investigating the pattern of effects for causal attributions [Preprint]. Retrieved from http://philsci-archive.pitt.edu/16626/.

Sytsma, J. (2019b). The Character of Causation: Investigating the Impact of Character, Knowledge, and Desire on Causal Attributions [Preprint]. Retrieved from http://philsci-archive.pitt.edu/16739/.

Sytsma, J. (2021). Causation, Responsibility, and Typicality. *Review of Philosophy and Psychology*, *12*(4), 699–719.

Sytsma, J. (2022). The Responsibility Account. In P. Willemsen & A. Wiegmann (Eds.), *Advances in Experimental Philosophy of Causation* (pp. 145–164). Bloomsbury Publishing.

Sytsma, J., Livengood, J., & Rose, D. (2012). Two Types of Typicality: Rethinking the Role of Statistical Typicality in Ordinary Causal Attributions. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(4), 814–820.

Sytsma, J., Willemsen, P., & Reuter, K. (2023). Mutual entailment between causation and responsibility. *Philosophical Studies*, *180*(12), 3593–3614.

Tobia, K. (2021). Law and the Cognitive Science of Ordinary Concepts. In B. Brozek, J. Hage, & N. Vincent, *Law and Mind: A Survey of Law and the Cognitive Sciences* (pp. 86–96). Cambridge University Press.

VerSteeg, R. (2011). Perspectives on Forseeability in the Law of Contracts and Torts: The Relationship between Intervening Causes and Impossibility. *Michigan State Law Review*, *2011*, 1497–1519.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Willemsen, P., & Kirfel, L. (2019). Recent empirical work on the relationship between causal judgements and norms. *Philosophy Compass*, *14*(1), e12562.