

# Causation, Foreseeability, and Norms

Levin Güver (levin.guever@gmail.com)

University College London, Faculty of Law, 4–8 Endsleigh Gardens  
WC1H 0EG London, UK

Markus Kneer (markus.kneer@gmail.com)

University of Zurich, Department of Philosophy, Zollikerstrasse 117  
8008 Zurich, CH

## Abstract

A growing body of literature has revealed ordinary causal judgement to be sensitive to normative factors, such that a norm-violating agent is regarded more causal than their non-norm-violating counterpart. In this paper, we explore two competing explanations for this phenomenon: the Responsibility View and the Bias View. The Bias View, but not the Responsibility View, predicts features *peripheral* to the agent’s responsibility to impact causal attributions. In a series of three preregistered experiments (N = 1162), we present new evidence that the Norm Effect arises from such peripheral features, namely from nonpertinent or entirely silly norm violations. Furthermore, we show that this effect cannot be explained by recourse to the agent’s foreknowledge or desire of the outcome, nor by its foreseeability: the Norm Effect arises even when participants judge the norm-violating agent’s doing as equally foreseeable. This, we argue, provides evidence in favour of the Bias View.

**Keywords:** causation; norms; bias; blame; responsibility; foreseeability; negligence; experimental philosophy

## The Norm Effect

A growing body of psychological 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; Hitchcock & Knobe, 2009; Icard, Kominsky, & Knobe, 2017; Knobe & Fraser, 2008; Kominsky et al., 2015; Samland & Waldmann, 2016; Olier & Kneer, 2022) – a phenomenon that has since been termed the Norm Effect.<sup>1</sup> Imagine the following scenario (*Rollerblading*): Mark is rollerblading on a path while Lauren is walking ahead of him. Suddenly, a cat jumps in front of Lauren, 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?

Participants overwhelmingly point to the cat (Güver & Kneer, 2023). Now imagine a slight change to the vignette – the addition of a norm prohibiting Mark from rollerblading on the path – and ask again: who caused the accident? This slight change 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). In the following, we will focus on but two such explanations: the *Responsibility View* and the *Bias View*.

## The Responsibility View and the Bias View

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., 2022; Sytsma, Livengood, & Rose, 2012), where “responsible” is understood as a “normative evaluation” (Sytsma, 2019b). Thus, when people use the expression “cause”, they are referring 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. 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; 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

<sup>1</sup> The Norm Effect has been shown to arise not only in the context of prescriptive (or injunctive) norms, which make a claim as to 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). Unless stated otherwise, following mentions of norms refer to prescriptive norms.

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.

### Teasing the two views apart

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?

The answer lies in the precise demarcation of the term “responsible. As Sytsma clarifies, responsibility judgements are “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, holding instead 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 responsibility influence causal judgement or not.

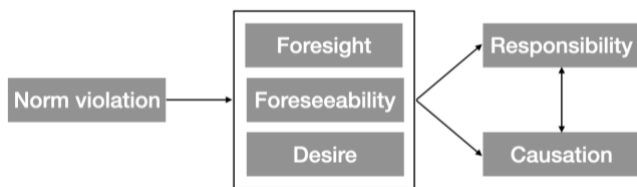


Figure 3: A more complex pathway model of the Responsibility View.

An early example of this approach can be found in Alicke (1992), whose findings seemed to suggest that persons with bad general character – a feature irrelevant to responsibility – were indeed deemed more causal. However, as Sytsma (2019b) has shown, it seems likely that participants in Alicke’s original study drew implicit 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 agents 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, is turn, relevant to the assessment of agential responsibility.

A more sophisticated version of the Responsibility View, such as the one proposed by Sytsma (2019b), accounts for the potential mediating role of several potentially inferred factors (Figure 3). Returning to our opening example, *Rollerblading*, participants may infer that Mark (i) wanted to crash into Lauren (*desire*), (ii) knew he’d crash into Lauren (*foresight*), or (iii) should have foreseen a crash (*foreseeability*).<sup>2</sup>

### The present studies

In the following, we will explore whether the Norm Effect arises from the violation of nonpertinent or silly norm violations (see Güver & Kneer, 2023). These are norms that are either irrelevant to the outcome at hand, or patently silly. As Sytsma (2019a) explicitly acknowledges, it “is imperative for [the Responsibility View] that the norm-violating action is connected to the outcome”. We will show that nonpertinent and silly norm violations *do* provoke the Norm Effect and that this effect cannot be exhaustively explained by recourse to potential intermediary factors such as desire, foresight, or foreseeability. All preregistrations, data, vignettes, and additional analyses are available under <https://osf.io/sxtf7/>.

### Experiment 1

The vignette, titled *Festival*, is based on a real criminal case (BBC News, 2015; Tapei Times, 2016).

### Participants

We recruited 305 participants on Amazon Mechanical Turk. Their IP address was restricted to the United States. As preregistered, 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).

### Methods and materials

In *Festival*, Mark attends a music festival where Lauren is responsible for the special stage effects. During the height of 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, ignited, 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 topsless.

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).

<sup>2</sup> This is consistent with the findings of Kirfel and Phillips (2021) that agents who unknowingly violate norms are not judged more causal than their norm-adhering counterparts: agents that

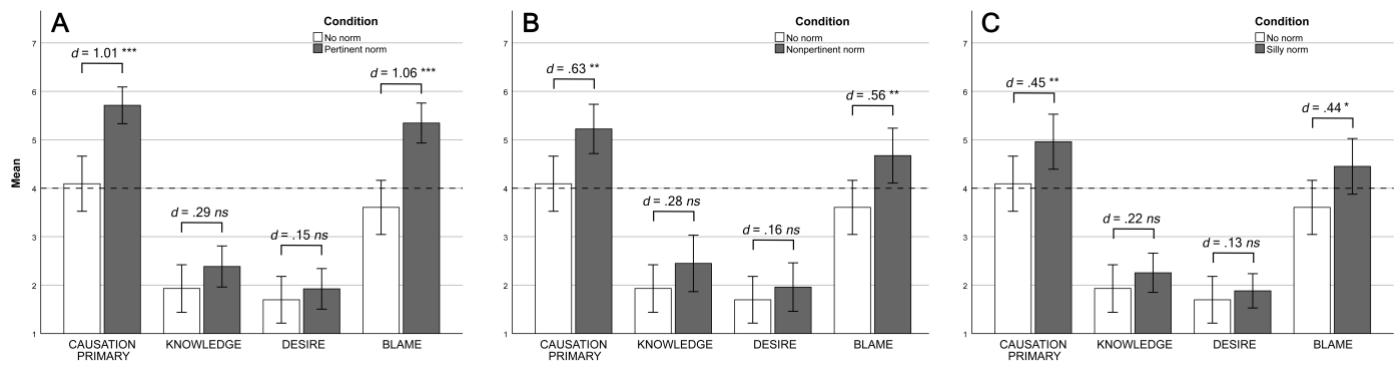


Figure 4: Comparison of means between the no norm and pertinent norm (A), nonpertinent norm (B), and silly norm (C) conditions. Effect sizes are given in terms of Cohen's *d*, \* indicates  $p < .05$ , \*\* indicates  $p < .01$ , \*\*\* indicates  $p < .001$ . Error bars denote 95% CI.

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, deserved punishment. All responses were recorded on 7-point Likert scales.

## Results

**ANOVAs** One-way ANOVAs revealed a significant influence of norm type on Mark's perceived causal contribution, blame, responsibility, and punishment (all  $ps < .011$ ), with large effect sizes for Mark's causal contribution and blameworthiness ( $\eta_p^2s > .156$ ) and moderate effect sizes for responsibility and punishment (responsibility:  $\eta_p^2 = .126$ ; punishment:  $\eta_p^2 = .115$ ). Both knowledge and desire, however, proved nonsignificant ( $ps > .234$ ).

We further ran independent samples *t*-tests for all dependent variables across the four norm type conditions.

**No norm v. pertinent norm** A comparison of the no norm and pertinent norm condition revealed that participants in the latter condition judged Mark significantly and pronouncedly more causal, blameworthy, responsible, and deserving of punishment ( $ps < .001$ ,  $ds > 1.00$ , large effects). There was no statistically significant difference across conditions for the knowledge and desire measures ( $ps > .159$ ; see Figure 4A).

**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 towards the accident, as well as more deserving of punishment (all  $ps < .009$ , all  $ds > .55$ ). Crucially, the difference in knowledge and desire

ratings did not reach statistical significance ( $ps > .173$ ; 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 meriting of punishment (all  $ps < .039$ , all  $ds > 0.43$ , moderate effects). There was no statistically significant impact of norm type on knowledge and desire judgements ( $ps > .302$ ; see Figure 4C).

## 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. In order to distinguish between the Responsibility View and the Bias View, however, we must thus turn to the findings of the nonpertinent and silly norm conditions.

According to the Responsibility View, factors that are peripheral to the legitimate ascription of responsibility should not influence participants' causal judgements (Sytsma, 2019b). The nonpertinent and silly norm conditions were explicitly designed in such a way 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 has stressed, in the absence of a connection between action and outcome, an agent cannot be held responsible (Sytsma, 2019a). 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 difference in desire or knowledge ascriptions.

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

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.

Contrast	Variable	Between-subjects					Within-subjects				
		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	-1.18	0.855	0.02	[-.21;.18]

thus exaggerate his causal contribution, in an attempt to justify such blame.

If the Bias View were correct, as the results seem to suggest, reflection on behalf of the participants might be able to ameliorate the cognitive error that is taking place. Experiment 2 aims at eliciting such a reasoned judgement by employing a within-subjects design.

## Experiment 2

The second experiment employed a within-subjects design in which participants were presented with two conditions (no norm v. norm violation) of the *Festival* vignette side-by-side. The aim was to see whether this increases reflective judgement which could in turn reduce the Norm Effect by way of debiasing judgements (Baron, 2008; Hsee, 1996; Kneer & Machery, 2019; Kneer & Skoczen, 2022).

### Participants

358 participants were recruited online via Amazon Mechanical Turk. Their IP address was restricted to the United States. As preregistered, 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).

### 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.

## Results

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

We ran paired samples *t*-tests for a more detailed breakdown of the impact of norm type on the dependent variables. Table 1 contrasts a summary of the findings with the between-subjects findings from Experiment 1.

**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 ( $p < .001$ ,  $d = .70$ ). 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 ( $p = .009$ ).

**No norm v. nonpertinent norm** In comparing the no norm and nonpertinent norm conditions, we did not find any statistically significant differences in participants' assessments of the dependent variables (all  $ps > .173$ ), with very small effect sizes throughout (all  $ds < .15$ ).

**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  $ps > .123$ ), with tiny effect sizes throughout (all  $ds < .08$ , except for Lauren's causal contribution at  $d = .15$ ).

## Discussion

With respect to the nonpertinent and silly norm conditions, the results paint a clear picture: where in the between-subjects comparisons (Experiment 1) there were significant and medium-to-large effects for causation and the moral variables, the effect vanished in the within-subjects comparisons (Experiment 2), see Table 1. Upon reflection,

participants did not judge nonpertinent and silly norm-violating agents differently from no norm-violating ones, thus providing evidence in favor of the Bias View. These findings for causation track the results of between- and within-subjects contrasts for *mens rea* attributions reported by Kneer & Machery (2019) and Kneer & Skoczen (2023).

The situation is more complex in the case of pertinent norms. 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. Reflective judgement thus weakens the Norm Effect.

Nevertheless, a residual 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 ( $r = .82$ ), and the pertinent norm condition ( $r = .79$ ).

Proponents of the Responsibility View might further argue that we have not been testing the most appropriate mediators. Since our vignettes involve accidents, it comes as no surprise that participants do not ascribe knowledge or desire to the agent. What we should be testing instead is the carelessness or negligence of the agent, which is determined in relation to how reasonably foreseeable the accident was (Engelmann & Waldmann, 2021; Kirfel & Lagnado, 2021; Lagnado & Channon, 2008; Kneer & Machery, 2019; Kneer, 2022; Nobes & Martin, 2022; Sarin & Cushman, 2022). Participants might 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 its coming about (Dressler, 2015; Goldberg & Zipursky, 2010; Owen, 2009).

### Experiment 3

Experiment 3 set out to explore whether the findings of the previous experiments could be explained by recourse to the reasonable foreseeability of an accident, *i.e.* the agent's negligence. In order to account for the hindsight bias which frequently besets foreseeability judgements (Kamin & Rachlinski, 1995; Margoni & Surian, 2022; Kneer & Skoczen, 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 another scenario, the *Shooting range* vignette.

#### Participants

1034 participants were recruited online on Amazon Mechanical Turk. Their IP address was restricted to the

United States. As preregistered, 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 = 20–100 years).

#### Methods and materials

The study took a 4 (norm type: no norm v. pertinent norm v. nonpertinent norm v. silly norm)  $\times$  2 (presentation of foreseeability question: *ex ante* v. *ex post*) between-subjects design. Participants were randomly sorted into one of 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 was at the shooting range despite it being closed. In the nonpertinent norm condition, it was prohibited to use the shooting range unless one wore protective gear such as gloves and glasses, and Mark did not wear any. In the silly norm condition, it was forbidden to bring any type of food or drinks to the shooting range, and Mark snuck 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.

#### Results

**ANOVA** A 4 (norm type)  $\times$  2 (presentation order) between-subjects ANOVA revealed a significant main effect of both order and norm type on foreseeability (both  $ps < .001$ ), with a small-to-moderate effect size for order ( $\eta_p^2 = .057$ ) and a moderate effect size for norm type ( $\eta_p^2 = .084$ ). 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 main effect of presentation order ( $p < .001$ ,  $\eta_p^2 = .024$ , a small effect). We further found significant and large main effects of norm type on all moral variables (all  $ps < .001$ , all  $\eta_p^2s > .256$ ), and small main effects for presentation order (all  $ps < .007$ , all  $\eta_p^2s < .029$ ).

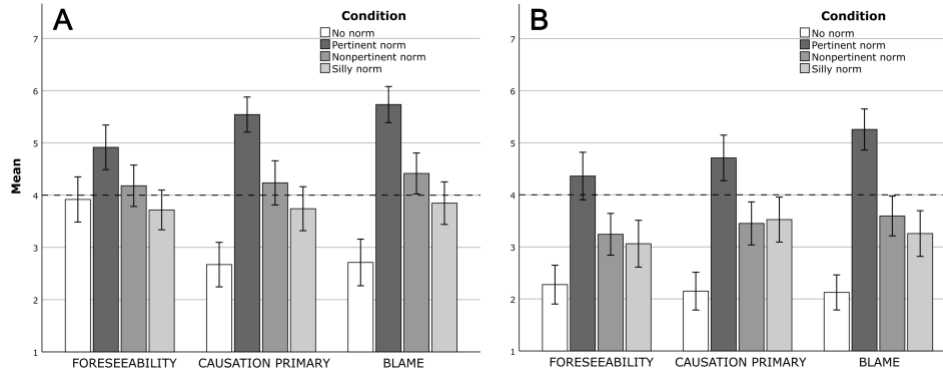


Figure 5: Comparison of means across the four *ex ante* (A) and *ex post* (B) conditions. Error bars denote 95% CI.

We explore the results of independent samples *t*-tests below (see Figure 5).

**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  $ps < .003$ ). Whereas the effect sizes for most variables were relatively stable across the *ex ante* and *ex post* presentation styles (effect sizes not differing in terms of more than .30), there was a great discrepancy in foreseeability judgements, where a moderate effect *ex ante* ( $d = 0.50$ ) doubled in size for the *ex post* ( $d = 1.08$ , a large effect).

**No norm v. nonpertinent norm** Participants in the *ex ante* presentation order assessed Mark's causal contribution, his blameworthiness, responsibility, and deserved punishment differently between the no norm and nonpertinent norm conditions (all  $ps < .001$ , all  $ds > .80$ , large effects). There was, however, no difference in their assessment as to how foreseeable the accident was ( $p = .376$ ). This changed in the *ex post* presentation order: participants now judged all dependent variables – including foreseeability – differently between the no norm and nonpertinent norm conditions (all  $ps < .002$ ), with moderate to large effect sizes throughout (all  $ds > .52$ ).

**No norm v. silly norm** As in the comparison above, when viewing the *ex ante* conditions, participants' judgements different significantly when it came to Mark's causal contribution, his blameworthiness, responsibility, and deserved punishment (all  $ps < .011$ , all  $ds < .70$ , small to moderate effects). Nevertheless, foreseeable remained nonsignificant ( $p = .488$ ). This is to be contrasted with the *ex post* conditions, where we found significant differences in all dependent variables – including, again, foreseeability – between the no norm and silly norm conditions (all  $ps < .009$ , all  $ds < .77$ , small to moderate effects).

## Discussion

The aim of Experiment 3 was to investigate the role of foreseeability in the context of norm violations. *Ex ante*

contrasts of the no norm condition with the nonpertinent and silly norm conditions reveals participants to have judged the accident as equally foreseeable ( $ps > .375$ ). Nevertheless, the Norm Effect persisted, as participants continued to note stark differences with respect to the agent's causal contribution ( $ds < .82$ ) and moral variables ( $ds < 1.00$ ). The findings raise trouble for the Responsibility View, as they show that the Norm Effect continues to seep in, even when foreseeability is held fixed across conditions. Furthermore, although participants did judge the accident in the pertinent norm condition as more foreseeable, the effect size is comparatively small ( $d = .50$ ) in relation to the huge effect on perceived causation ( $d = 1.67$ ) and the moral variables ( $ds > 1.83$ ). In short, the Norm Effect on causation in the nonpertinent and silly norm conditions cannot be explained by aid of foreseeability judgments, and foreseeability can at best account partially for the effect in the pertinent norm condition.

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.

## Conclusion

The Responsibility View and the Bias View come apart in their treatment of factors peripheral to moral responsibility: the former, unlike the latter, holds that such factors will not influence causality judgments. In three experiments, we have shown that they do, and that this influence cannot be explained by recourse to legitimate responsibility-enhancing factors such as desire, foreknowledge, or foreseeability.

While our results provide considerable evidence in favour of the Bias View, they are not entirely conclusive just yet, as the interpretation of the pertinent norm v. no norm contrasts is debatable. Adherents of the Responsibility View could further argue that yet other mediators should be explored, though the burden of proof, at this point, lies with them.

## Acknowledgments

We would like to thank the four anonymous reviewers for their helpful comments. Work on this project was supported by a Swiss National Science Foundation Ambizione grant (PZ00P1\_179912, PI: Kneer).

## 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., & 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.
- BBC News. (2015, June 28). Taiwan Formosa Water Park explosion injures hundreds. *BBC News*. Retrieved from <https://www.bbc.com/news/world-asia-33300970>.
- Dressler, J. (2015). *Understanding Criminal Law* (7th ed.). LexisNexis.
- 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>.
- Gerstenberg, T., & Icard, T. (2020). Expectations affect physical causation judgments. *Journal of Experimental Psychology: General*, 149(3), 599–607.
- 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.
- 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.
- Henne, P., O’Neill, K., Bello, P., Khemlani, S., & De Brigard, F. (2021). Norms Affect Prospective Causal Judgments. *Cognitive Science*, 45(1), e12931.
- Henne, P., O’Neill, K., Bello, P., Khemlani, S., & De Brigard, F. (2021). Norms Affect Prospective Causal Judgments. *Cognitive Science*, 45(1), e12931.
- 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. (2021). Causal judgments about atypical actions are influenced by agents’ epistemic states. *Cognition*, 212, 104721.
- 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.
- 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., & 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., & Fraser, B. (2008). Causal Judgment and Moral Judgment: Two Experiments. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 2* (pp. 441–447). MIT Press.
- Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, 137, 196–209.
- Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, 108(3), 754–770.
- 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.
- 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.
- 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.
- 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.
- Roose, 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.
- Samland, J., & Waldmann, M. R. (2016). How Prescriptive Norms Influence Causal Inferences. *Cognition*, *156*, 164–176.
- Samland, J., & Waldmann, M. R. (2016). How Prescriptive Norms Influence Causal Inferences. *Cognition*, *156*, 164–176. <https://doi.org/10.1016/j.cognition.2016.07.007>
- 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>.
- 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. (2022). *Mutual Entailment between Causation and Responsibility* [Preprint]. Retrieved from <http://philsci-archive.pitt.edu/20497/>.
- Tapei Times. (2016, April 26). ‘Color Play Asia’ Organiser Found Guilty. *Taipei Times*. Retrieved from <https://www.taipetimes.com/News/front/archives/2016/04/27/2003644910>.
- Willemsen, P., & Kirfel, L. (2019). Recent empirical work on the relationship between causal judgements and norms. *Philosophy Compass*, *14*(1), e12562.