

## **UC Merced**

# **Proceedings of the Annual Meeting of the Cognitive Science Society**

### **Title**

The Impact of Ignorance Beyond Causation: An Experimental Meta-Analysis

### **Permalink**

<https://escholarship.org/uc/item/27v6t99v>

### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

### **ISSN**

1069-7977

### **Authors**

Kirfel, Lara  
Phillips, Jonathan

### **Publication Date**

2021

Peer reviewed

# The Impact of Ignorance Beyond Causation: An Experimental Meta-Analysis

Lara Kirfel ([ucjulki@ucl.ac.uk](mailto:ucjulki@ucl.ac.uk))

Department of Experimental Psychology, University College London, 26 Bedford Way  
London, WC1H 0AP England

Jonathan Phillips ([Jonathan.S.Phillips@dartmouth.edu](mailto:Jonathan.S.Phillips@dartmouth.edu))

Program in Cognitive Science, Dartmouth College, 23 N Main Street  
Hanover, NH 03755 USA

## Abstract

Norm violations have been demonstrated to impact a wide-range of seemingly non-normative judgments. Among other things, when agents' actions violate prescriptive norms they tend to be seen as having done those actions more freely, as having acted more intentionally, as being more of a cause of subsequent outcomes, and even as being less happy. The explanation of this effect continue to be debated, with some researchers appealing to features of actions that violate norms, and other researcher emphasizing the importance of agents' mental states when acting. Here, we report the results of a large-scale experiment that replicates and extends twelve of the studies that originally demonstrated the pervasive impact of norm violations. In each case, we build on the pre-existing experimental paradigms to additionally manipulate whether the agents knew that they were violating a norm while holding fixed the action done. We find evidence for a pervasive impact of ignorance: the impact of norm violations on non-normative judgments depends largely on the agent knowing that they were violating a norm when acting.

**Keywords:** normality; knobe effect; ignorance; knowledge; norms

## The Puzzling Impact of Normality

A large and growing body of research has documented that norm violations influence a wide range of intuitive judgments, including judgments of *intentional action* (Knobe, 2003), *causation* (Kominsky & Phillips, 2019), *freedom* (Young & Phillips, 2011), *happiness* (Phillips, De Freitas, Mott, Gruber, & Knobe, 2017), *doing vs. allowing* (Cushman, Knobe, & Sinnott-Armstrong, 2008), *pro-/con-attitude* ascriptions (Pettit & Knobe, 2009), and *modal* judgments (Knobe & Szabó, 2013). Such normality effects are not hard to demonstrate. Consider the following situation:

While sailing on the sea, a large storm came upon a captain and his ship. As the waves began to grow larger, the captain realized that his small vessel was too heavy and the ship would flood if he didn't make it lighter. The only way that the captain could keep the ship from capsizing was to cut the cargo loose which is weighing the ship down. The captain knows that the cargo contains his wife's expensive art collection because that is what he packed into the cargo.

Fully realizing the cargo contains his wife's expensive art collection, the captain cut the cargo loose and it fell into the sea. While the cargo containing his wife's expensive art collection sank to the bottom of the sea, the

captain was able to survive the storm and returned home safely.

Was the captain forced to throw his wife's cargo overboard? Intuitively, 'Yes.' (Phillips & Knobe, 2009). Now consider a variant in which the captains' actions violate a moral norm (changes in italics):

While sailing on the sea, a large storm came upon a captain and his ship. As the waves began to grow larger, the captain realized that his small vessel was too heavy and the ship would flood if he didn't make it lighter. The only way that the captain could keep the ship from capsizing was to cut the cargo loose which is weighing the ship down. The captain knows that the cargo contains his wife's expensive art collection because that is what he packed into the cargo. *However, he also learned that a number of illegal passengers have hidden in the cargo boxes before the ship left the harbor.*

Fully realizing the cargo contains *passengers*, the captain cut the cargo loose and it fell into the sea. While the cargo containing the *illegal passengers* sank to the bottom of the sea, the captain was able to survive the storm and returned home safely.

In cases like this, people judge the captain to have been much less forced (Phillips & Knobe, 2009). This impact of the normative status of an agent's action extends to the variety of judgment types mentioned before. A norm-violating agent is judged as acting more intentionally (Knobe, 2003), having more pro-attitudes (Pettit & Knobe, 2009), being more causal (Icard, Kominsky, & Knobe, 2017), being *less* happy (Phillips et al., 2017) and as making (vs. allowing) an outcome to occur, compared to an agent performing the same action but abiding to the norm (Phillips, Luguri, & Knobe, 2015; Pettit & Knobe, 2009). Hence, at this point there is little debate over *whether* norms influence a wide-range of judgments. Instead, discussion largely centers on why such an effect occurs.

For many of these judgments types, specific proposals have been put forward aiming to explain why norm-violations influence this type of judgments in particular (Hindriks, 2014). For example, Díaz and Reuter (2020) argue that reduced attributions of happiness to norm-violating agents are the result of how "fitting" people perceive the concept of happi-

ness to be under such circumstances. Or, the increased attribution of intentional action to norm-violating agents has been suggested to reflect rational inferences about the agents' mental states or beliefs (Lombrozo & Uttich, 2010; Uttich & Lombrozo, 2010; Laurent, Reich, & Skorinko, 2019; Alfano, Beebe, & Robinson, 2012). And, more generally, the influence of norms on mental state ascriptions has been argued to be mediated by some form of moral or blame judgment (Hindriks, Douven, & Singmann, 2016; Nadelhoffer, 2004; Cova, Lantian, & Boudesseul, 2016).

Alternatively, however, a more unified explanation of normality's impact across judgments has been offered (Knobe, 2010; Phillips et al., 2015; Phillips & Knobe, 2018). Often conceived as alternative proposal to theories endorsing a mediating role of blame, this account suggests that the influence of normality on all these judgments—from intentionality to happiness—is driven by people's reasoning about possibilities (Phillips & Knobe, 2018; Knobe, in press).

The most thoroughly investigated case study of the impact of normality is focused on judgments of causation (Willemsen & Kirfel, 2019). Across a now large series of studies, the debate has centered on which of these two theoretical accounts can best account for the sensitivity of people's causal judgments to normality (Samland & Waldmann, 2016; Kominsky & Phillips, 2019; Sytsma, 2020; Alicke & Rose, 2012). We will briefly describe the accounts, reconstruct the debate, and finally, introduce a critical finding for testing the respective theories: the case of ignorant norm-violations.

### Causes, Counterfactuals and Blame

At the broadest level, the general finding is that when some event is one of multiple necessary conditions for a given outcome to occur, the more abnormal that event is, the more people judge it to be the cause of the outcome (Knobe & Fraser, 2008). If two cars crash in the middle of an intersection—one of which ran a red light and one of which did not—the *cause* of the accident is intuitively the driver who violated the traffic norm, not the one who didn't. But why?

One family of accounts emphasizes that people's causal judgments may simply be a form of moral responsibility judgment in disguise. If so, then it is of course not very surprising that they are influenced by prescriptive norm violations (Alicke, 2000; Samland & Waldmann, 2016; Livengood, Sytsma, & Rose, 2017). In other words, when people say that the driver who ran the red light was the cause of the accident, they simply mean that he should be blamed or held morally responsible.

In response to these accounts, researchers have pointed out that *descriptive* norm violations, e.g., events that occur despite being very unlikely, exhibit a remarkably similar pattern in intuitive causal judgments (Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015; Gerstenberg & Icard, 2020). If a forest fire starts in the presence of oxygen, dry leaves, and a lightning strike, people tend to judge that the least likely of these events—the lightning strike—was the cause.

Responsibility-based accounts are difficult to extend in a way that naturally covers the impact of descriptive norm violations, since these events often do not even involve intentional agents who can be held responsible or blamed.

An alternative family of approaches has been to argue that norms influence causal judgments because causal judgments rely on counterfactual possibilities (Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2020; Lewis, 1974; Pearl, 2009), and norms are well-known to influence counterfactual thought (Kahneman & Miller, 1986). On this account, people are inclined to judge that the driver who ran the red light was the cause of the accident because they are more inclined to think about what would have happened if he had not run the red light (and are correspondingly less inclined to think about what would have happened if the other driver had stopped at a green light) (Phillips et al., 2015). Moreover, this approach can be easily extended to descriptive norms: it is more relevant to consider possibilities in which lightning strikes do not occur than possibilities in which there are no dry leaves in the forest. And thus, it is not difficult to see why both kinds of norm violations will influence causal judgments in similar ways.

However, a recent criticism of these counterfactual accounts has focused on cases in which an agent *unknowingly* violates a prescriptive norm (Samland & Waldmann, 2016). Samland and Waldmann (2016) argued that in such cases, because a norm is violated, counterfactual accounts should predict that whether or not the agent knows they are violating a norm, will not change normality's influence on causal judgments. They went on to show that in fact, participants' causal judgments were extremely sensitive to changes in the agents' knowledge: An agent who unknowingly violated a norm was *not* judged as more causal than a norm-abiding agent. While this work clearly demonstrated an important relationship between what agents know and the effect of normality on causal judgments, it did not provide decisive evidence against counterfactual explanations of these effects. As Kominsky and Phillips (2019) went on to show, participants' judgments of counterfactual relevance were also highly sensitive to the agents' mental states, and these counterfactual relevance judgments precisely predicted the differences in causal judgments. Importantly, recent work by Kirfel and Lagnado (2021) demonstrated that a surprisingly similar effect occurs in the case of descriptive norm violations. Specifically, agents are judged to be more causal of an eventual outcome when they did a statistically unlikely action that was necessary for that outcome, but this effect only occurs when the agent knows that their action leads to the outcome.

### The Impact of Ignorance Beyond Causation

Considered collectively, the upshot of this growing body of research is that the effect of norms on judgments of causation are broadly sensitive to agents' epistemic states. More specifically, ignorance cancels the influence of abnormality on causal judgements: abnormal but ignorant agents are not

perceived as more causal than normal agents. Given that causal judgments are just one type of judgments that are sensitive to normality, the question arises whether agents' epistemic states similarly affect the impact of norms across the wide range of different judgments. More precisely, it remains an important but unanswered question whether agent ignorance reduces the impact of normality on people's attributions of intentional action, freedom, happiness, and so on. This question is of theoretical interest because a systematic moderating effect of epistemic states on the impact of norms in a variety of judgment domains suggests that this impact of norms is itself driven by a common mechanism. The present study aims to provide a first answer to this question.

## Hypothesis

The experiment we report makes a novel contribution by considering the pervasive impact of normality as a whole (using a meta-analytic approach) and asking whether agents' epistemic states moderate the impact of normality for each of the different judgments that have previously been shown to be impacted by normality. More specifically, we both (1) attempt to replicate the effect of normality that has been previously demonstrated, and then (2) modify the original materials to allow us to ask whether the effect of normality is sensitive to changes in the agents' epistemic states. Drawing on recent work on the impact of ignorance on normality effects in causal judgments (Samland & Waldmann, 2016; Kirfel & Lagnado, 2021), we predict a similar effect of ignorant norm-violation on judgments of intentional action, freedom, happiness, etc. That is, we predict that the difference in judgments about a normal agent and an unknowingly norm-violating agent (here analysed as effect sizes) will be smaller than the difference between a normal and a knowingly norm-violating agent. In the current study, we will solely focus on the impact of *prescriptive* normality.

The results of this large-scale experiment will not only be informative for debates that have sought to explain the impact of normality in each separate case, but also will inform the broader question of whether researchers should seek a unified explanation of the pervasive impact of normality. In the General Discussion, we will return to the question of whether and how the two theoretical accounts outlined above—responsibility vs. counterfactuals—can provide an explanation of our findings.

## Methods

In an ongoing replication study and experimental meta-analysis, we selected 12 studies published between 2003 and 2019, containing 6 different paradigms with in total 29 statistical effects taken to indicate the influence of norm violations on different types of judgments.

### Selected Studies

To be included in this meta-analysis, studies needed to have investigated the impact of prescriptive norm violations, broadly construed (violating a conventional norm, causing

harm, etc.), on seemingly non-normative judgments. We identified 6 different judgment domains for which this was the case: *causation* (Kominsky & Phillips, 2019), *doing vs. allowing* (Cushman et al., 2008), *freedom* (Phillips & Knobe, 2009), *happiness* (Phillips et al., 2017), *mental state ascriptions* (Pettit & Knobe, 2009) and *modal judgments* (Knobe & Szabó, 2013).

**Causation** A series of studies finds that if the actions of two agents are necessary for an outcome to occur (a “conjunctive causal structure”), people judge the agent who violated a norm to be more of a cause of the outcome than the agent who acted according to the norm. In contrast, if both agents' actions are independently sufficient to bring about the outcome (a “disjunctive causal structure”), the agent who acted immorally is judged to be less of a cause of the outcome. We selected four scenarios (“battery”, “bridge”, “motion detector”, “computer”), which had both a conjunctive and disjunctive version from Kominsky and Phillips (2019) and the “pen” scenario from Knobe and Fraser (2008), which only had a conjunctive version.

**Doing vs. allowing** Work on the ‘doing/allowing’ distinction shows that morally bad behavior is more likely to be construed as actively ‘doing’ than as passively ‘allowing’. We selected the “Dr. Bennet” scenario from Cushman et al. (2008) for our study, in which a doctor removes a homeless man from life support.

**Freedom** As discussed in the introduction, studies on people's judgements about the freedom to act show that agents who acted immorally (vs. neutrally) are more thought to have acted freely (i.e., were not forced to do that action). Young and Phillips (2011) found that this effect is also affected by the moral focus of the force judgment: People agree more with the active form of the sentence “X forced Y to act” than the passive form “Y was forced by X to act”. For our study, we used the original “ship” vignette from Phillips and Knobe (2009) as described above, as well as the *active* and *passive* version of the “ship” and “doctor” scenario from Young and Phillips (2011).

**Happiness** Previously, research found that even when an agent is described as satisfying all of the psychological criteria for happiness (high positive affect, low negative affect, high life satisfaction), participants are disinclined to rate the agents as being “happy” when they believe the agent to be living an immoral lives (though not when living morally good or neutral life). We selected the “nurse” scenario from Phillips et al. (2017) as paradigmatic test case of this effect for our study.

**Mental State Ascriptions** This line of research, also known as the “side-effect effect”, shows that an agent who brings about a side effect is judged as having intended this side effect to a greater extent when this effect is bad vs. good. Subsequent studies have shown that this pattern occurs for the attribution of other mental states (e.g. desire) as well

(Pettit & Knobe, 2009), and find an inverse effect for attribution of opposition: People judge the agent to have opposed the effect *less* when the side effect is morally bad vs. good. We selected the original “chairman” scenario for testing “intentionality” from Knobe (2003) and “decision/desire” from Pettit and Knobe (2009), as well as the “manager” scenario (“advocate / in favour of”), “CEO” scenario (“opposed to”) and “bomb” scenario (“intended to”) from Pettit and Knobe (2009). In addition, we selected the “gizmo”, “scrubs” and “truck trailers” scenarios from Uttich and Lombrozo (2010) who tested the attributions of intentionality to agents violation conventional norms. We subsume the various effects in this area under the term “mental states ascriptions”.

**Modal Proxies** Knobe and Szabó (2013) demonstrated that the effect of norm violations found in previous research on force, intention, causation extended to ‘modal proxies’ of these judgments. For example, just as people would say an agent was more forced to do a morally neutral action than an immoral action, they more agreed with the sentence “Given the circumstances, the agent had to do that action” when the action was morally neutral than when the action was immoral. We selected the “captain”, “pen” and “bulls-eye” vignette for our study.

### Pre-replication procedure

Each of the 29 selected scenarios included two experimental conditions (see the “ship” scenario from Phillips and Knobe (2009) from the introduction): One “Normal” condition in which the agent acts morally good or neutral, and one “Norm Violating” condition in which the agent acts morally bad. We created a third experimental condition for each of the 29 scenarios that matched the “Norm Violation” condition in all aspects, except for the agent’s epistemic state about the the normality of their action. In the “Ignorant Norm-Violation” condition, the agent’s action violates a norm (causes harm, etc.) but the agent is unaware that their action violates a norm. To illustrate, here is the “Ignorant Norm Violation” condition of Phillips and Knobe (2009)’s ship scenario (differences again indicated by italics):

While sailing on the sea, a large storm came upon a captain and his ship. As the waves began to grow larger, the captain realized that his small vessel was too heavy and the ship would flood if he didn’t make it lighter. The only way that the captain could keep the ship from capsizing was to cut the cargo loose which is weighing the ship down. The captain *thinks* that the cargo contains his wife’s expensive art collection because that is what he packed into the cargo. However, *completely unbeknownst to the captain*, a number of illegal passengers have hidden in the cargo boxes before the ship left the harbor.

*Without* realizing the cargo contains passengers, the captain cut the cargo loose and it fell into the sea. While the cargo containing the illegal passengers sank to the bot-

tom of the sea, the captain was able to survive the storm and returned home safely.

In order to match all three conditions and to allow for a consistent manipulation of the agent’s knowledge of the normality of their action, we modified aspects of some the original vignettes. We also adapted the precise phrasing of the Dependent Variable Questions of some studies and standardised the rating scales (see below). Accordingly, our replication approach varied from a relatively direct replication in some cases, to something closer to that of a conceptual replication in other cases (Hendrick, 1990; Lynch Jr, Bradlow, Huber, & Lehmann, 2015). Importantly, the focus of this study centers on investigating the moderating role of epistemic states on the effects of morality, rather than the strict replicability of the original effects.

### Participants, Design and Procedure

We recruited 1554 participants via Amazon Mechanical Turk ( $M_{age} = 40.70$ ,  $SD_{age} = 12.77$ ,  $N_{female} = 736$ ). Our study employed a 3 *normality* (normal vs. norm violation vs. ignorant norm violation)  $\times$  3 (*scenario*) design. Both norm and scenario were manipulated within participants. That is, each participant saw one example of each of the normality conditions in randomised order, and for each normality condition, the scenario participants saw was randomly sampled from the 29 different scenarios included in the study. After reading each scenario, participants responded to three different questions:

**Dependent Variable** A rating of the dependent variable from the original study, which was sometimes adapted to the normality condition and sometimes not. For the ship scenario, for example, the dependent variable was an agreement rating with the statement: “The ship captain was forced to cut the cargo loose and let it fall into the sea.” on a 7-point Likert scale (1-‘strongly disagree’, 7-‘strongly agree’).

**Knowledge Check** A knowledge check question, asking about the central agent’s knowledge of the abnormality of their action (e.g. Please rate how much you agree or disagree with the following statement: “The captain knew that a number of illegal passengers were hiding in the cargo boxes.”) on a 7-point Likert scale (1-‘strongly disagree’, 7-‘strongly agree’).

**Should know** A question about what the agent should have known with regards to the abnormality of their action (e.g. Please rate how much you agree or disagree with the following statement: “The captain *should* have known that a number of illegal passengers were hiding in the cargo boxes.”) on a 7-point Likert scale (1-‘strongly disagree’, 7-‘strongly agree’). This question served to examine whether the manipulation of epistemic states in our experiments not only influenced people’s beliefs about what the agents actually knew, but also their normative beliefs about what the agents should have known.

## Replicated and New Effect Sizes by Study

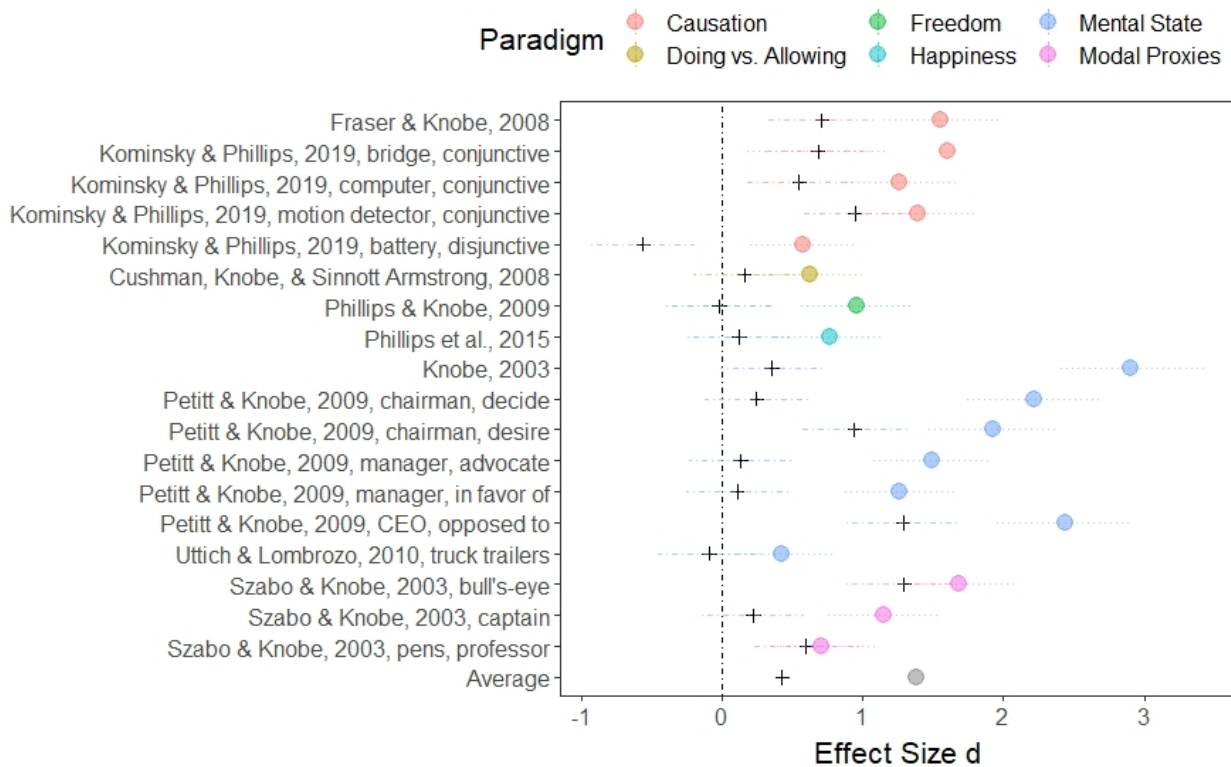


Figure 1: **Effect Sizes by Study**: Replication effect sizes of the original effects (normal vs. norm-violation) are marked by dots, effect sizes of the new effect (normal vs. ignorant norm-violation) are marked by crosses. Error bars depict 95% confidence intervals of effect sizes

### Analysis approach

Study materials and analyses were pre-registered at <https://osf.io/g52zs>. For each individual study, we first tested whether we replicated the originally observed effect, i.e., whether there was significant difference in dependent variable between the “Normal” and “Norm Violation” conditions. A replication was considered successful when  $p < .05$  and the effect was in the same direction as the original effect. We collected the effect sizes for those effects that were replicated (all converted to Cohen’s  $d$ ). To simplify our analysis, we reduced all normality effects to simple effects. That is, interaction effects such as the causal structure  $\times$  normality interaction effect observed in Kominsky and Phillips (2019) was decomposed into two separate simple effects. In all cases where we were able to replicate the original effect, we then performed the same statistical test, but replaced the “Norm Violation” condition with the newly created “Ignorant Norm Violation” condition and recorded the new effect size that measures the difference between these conditions. We adopted a meta-analytic approach towards our hypothesis that norm effects are influenced by agents’ epistemic states, predicting that effect sizes of the statistical tests for “Neutral vs. Ignorant Norm Violation” will be smaller than in the “Neutral vs. Norm Violation” tests. In order to evaluate this hypothesis, we first aligned all norm effects by reversing the direc-

tion of the effect size direction for those studies in which the norm manipulation has been shown to lead to a reduction in the DV rating (e.g. agent is judged as *less* forced in abnormal vs. neutral condition). We then built a linear mixed-effects null model including a random intercept for the study being replicated and extended ( $1 \mid \text{study}$ ) and a random intercept and slope for the impact of epistemic states across paradigms ( $\text{epistemicState} \mid \text{paradigm}$ ), and compared it to a model had the same random effects structure but included a fixed effect for “Epistemic State”. This factor coded for whether effect size was for a case in which the norm violation was known vs. unknown. The fixed effect was determined to be significant if the fit of the the model that included the fixed factor for epistemic state differed significantly from the model including only the random effects. The same procedure was also used for both kinds of knowledge ratings.

### Results

Overall, we successfully replicated 18 out of 29 effects of normality on non-moral judgments (62%): 5 effects on judgments of *causation* (out of 9), the effect on *doing vs allowing*, 1 effect on judgments about *freedom* (out of 5), the effect on judgments about *happiness*, 7 effects on *mental state ascriptions* (out of 10) and all three effects on *modal proxies*. While this replication rate is low, it is important to note that many

of our tests were not direct replications because they involved modifying the original materials to allow for a close match between the new conditions in which the agent was ignorant of the normative status of their action. The size of the original effects also varied greatly. Because we conducted this experiment as a single large-scale study which randomly assigned participants to conditions, we were under-powered to detect some of the smaller effects, and over-powered to detect larger effects.

**Knowledge Check** Our analysis revealed a significant effect for whether the norm violation occurred knowingly vs. unwittingly on effect sizes of knowledge ratings  $\chi^2(1) = 16.13$ ;  $p < .001$  ( $b = 2.73$ ,  $SE = .31$ ,  $t = 8.66$ ). The effect on ratings of the agent's knowledge about the abnormality of their behaviour was larger when the norm violation was intentional ( $M = 2.63$ ,  $SD = 1.35$ ) vs. ignorant ( $M = -0.01$ ,  $SD = 1.30$ ). While this is not surprising, it serves as an important manipulation check, demonstrating that we successfully manipulated participants' perceptions of the agents' knowledge.

**Should Know Check** Additionally, the extent to which people judged that the agent should have known that their behaviour was counter-normative was also predicted by our manipulations of the agents' knowledge,  $\chi^2(1) = 6.59$ ;  $p = .010$  ( $b = 1.44$ ,  $SE = .43$ ,  $t = 3.38$ ). Differences in whether the agent should have known were larger when the agent knowingly violated a norm ( $M = 1.63$ ,  $SD = 1.17$ ) than when they did so unknowingly ( $M = -0.03$ ,  $SD = 1.03$ ). This suggests that our manipulations not only successfully manipulated participants' perceptions of whether the agents did in fact know that their actions violated a norm, but also whether they *should* have known that.

**Dependent Variable** Given these results, we can now turn to the critical test of our hypothesis: whether our manipulations of the agents' epistemic states affected the pervasive impact of morality. We found that they did. Once again, the likelihood ratio test indicated that a model including a fixed effect for "Epistemic State" provided a better fit for effect sizes of the dependent variable than a model without it  $\chi^2(1) = 9.26$ ;  $p = .002$  ( $b = .85$ ,  $SE = .17$ ,  $t = 4.95$ ). The average replication effect size, i.e. the effect size for the original effect of norm violation was larger ( $M = 1.39$ ,  $SD = 0.66$ ) than the average new effect size, i.e. the effect of a ignorant norm violation ( $M = 0.43$ ,  $SD = 0.48$ ) (Figure 1).

## Discussion

Norm violations have been previously demonstrated to influence a wide range of intuitive judgments, including judgments of causation, freedom, happiness, doing vs. allowing, mental state ascriptions, and modal claims. A continuing debate centers on why normality has such a pervasive impact, and whether one should attempt to offer a unified explanation of these various effects (Hindriks, 2014).

At the broadest level, the current results demonstrate that the pervasive impact of normality likely warrants a unified

explanation at some level. Across a wide range of intuitive judgments and highly different manipulations of an agents' knowledge, we found that the impact of normality on non-normative judgments was diminished when the agent did not know that they were violating a norm. That is, we found evidence for a correspondingly pervasive impact of ignorance.

We see two broad approaches for accounting for this pervasive impact of ignorance. On the one hand, responsibility-based accounts may try and use this fact to their advantage by arguing that knowledge is directly relevant to moral responsibility, and thus that these effects merely reflect that participants responses truly are a form of disguised responsibility judgments (Alicke & Rose, 2012; Livengood et al., 2017; Samland & Waldmann, 2016). Such move would however require a plausible theory about why a variety of seemingly non-normative judgments such as happiness or modal claims are used by participants to express or indicate blame and responsibility (Samland & Waldmann, 2016). In addition, this account still faces the challenge of explaining the similar impact of descriptive rather than prescriptive norms (Gerstenberg & Icard, 2020; Kominsky et al., 2015) and moreover the specific similarity of the role of ignorance in those cases (Kirfel & Lagnado, 2021).

On the other hand, researchers may try to extend a unified counterfactual-based account to explaining the pervasive impact of normality (Phillips et al., 2015; Phillips & Knobe, 2018). However, these accounts face important challenges too. One is to show how and why these epistemic states play the correct role in shaping counterfactual reasoning. While some progress on this has been made in the case of causal judgments (Kominsky & Phillips, 2019), no general account has been offered. This remains an important area for future work. A second challenge is to explain the role of counterfactual or other forms of modal thought in each of the judgments where normality has been found to have an impact. While the relevant work has again been done in some of the cases, see e.g., the work by Knobe and Szabó (2013) and Phillips et al. (2015), there are other cases where the connection to counterfactuals is less clear, as in the case of assessments of happiness (Phillips, Misenheimer, & Knobe, 2011) or other mental state attributions (Pettit & Knobe, 2009). This too remains an important area for future work.

However, whichever of these theories turns out to be correct in the end, this work should inspire a new target on the impact of normality, since one needs not only to explain the pervasive impact of norms, but also the pervasive impact of ignorance.

## References

- Alfano, M., Beebe, J. R., & Robinson, B. (2012). The centrality of belief and reflection in knobe-effect cases: A unified account of the data. *The Monist*, 95(2), 264–289.
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological bulletin*, 126(4), 556.

- Alicke, M. D., & Rose, D. (2012). Culpable control and causal deviance. *Social and Personality Psychology Compass*, 6(10), 723–735.
- Cova, F., Lantian, A., & Boudesseul, J. (2016). Can the knobe effect be explained away? methodological controversies in the study of the relationship between intentionality and morality. *Personality and Social Psychology Bulletin*, 42(10), 1295–1308.
- Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition*, 108(1), 281–289.
- Díaz, R., & Reuter, K. (2020). Feeling the right way: Normative influences on people's use of emotion concepts. *Mind & Language*.
- Gerstenberg, T., Goodman, N., Lagnado, D., & Tenenbaum, J. (2020). A counterfactual simulation model of causal judgment.
- Gerstenberg, T., & Icard, T. (2020). Expectations affect physical causation judgments. *Journal of Experimental Psychology: General*, 149(3), 599.
- Hendrick, C. (1990). Replications, strict replications, and conceptual replications: are they important? *Journal of Social Behavior and Personality*, 5(4), 41.
- Hindriks, F. (2014). Normativity in action: how to explain the knobe effect and its relatives. *Mind & Language*, 29(1), 51–72.
- Hindriks, F., Douven, I., & Singmann, H. (2016). A new angle on the knobe effect: Intentionality correlates with blame, not with praise. *Mind & Language*, 31(2), 204–220.
- Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, 161, 80–93.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological review*, 93(2), 136.
- Kirfel, L., & Lagnado, D. (2021). Causal judgments about atypical actions are influenced by agents' epistemic states. *Cognition*, 212, 104721. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0010027721001402> doi: <https://doi.org/10.1016/j.cognition.2021.104721>
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(3), 190–194.
- Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33(4), 315–329. doi: 10.1017/S0140525X10000907
- Knobe, J. (in press). *Morality and possibility*. The Oxford Handbook of Moral Psychology. Oxford: Oxford University Press.
- Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. *Moral psychology*, 2, 441–8.
- Knobe, J., & Szabó, Z. G. (2013). Modals with a taste of the deontic. *Semantics and Pragmatics*, 6, 1–1.
- Kominsky, J. F., & Phillips, J. (2019). Immoral professors and malfunctioning tools: Counterfactual relevance accounts explain the effect of norm violations on causal selection. *Cognitive science*, 43(11), e12792.
- Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, 137, 196–209.
- Laurent, S. M., Reich, B. J., & Skorinko, J. L. (2019). Reconstructing the side-effect effect: A new way of understanding how moral considerations drive intentionality asymmetries. *Journal of Experimental Psychology: General*, 148(10), 1747.
- Lewis, D. (1974). Causation. *The journal of philosophy*, 70(17), 556–567.
- 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.
- Lombrozo, T., & Uttich, K. (2010). Putting normativity in its proper place. *Behavioral and Brain Sciences*, 33(4), 344.
- Lynch Jr, J. G., Bradlow, E. T., Huber, J. C., & Lehmann, D. R. (2015). Reflections on the replication corner: In praise of conceptual replications. *International Journal of Research in Marketing*, 32(4), 333–342.
- Nadelhoffer, T. (2004). Blame, badness, and intentional action: a reply to knobe and mendlow.
- Pearl, J. (2009). *Causality*. Cambridge university press.
- Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. *Mind & language*, 24(5), 586–604.
- Phillips, J., De Freitas, J., Mott, C., Gruber, J., & Knobe, J. (2017). True happiness: The role of morality in the folk concept of happiness. *Journal of Experimental Psychology: General*, 146(2), 165.
- Phillips, J., & Knobe, J. (2009). Moral judgments and intuitions about freedom. *Psychological Inquiry*, 20(1), 30–36.
- Phillips, J., & Knobe, J. (2018). The psychological representation of modality. *Mind & Language*, 65–94. Retrieved from <http://dx.doi.org/10.1111/mila.12165> doi: 10.1111/mila.12165
- Phillips, J., Luguri, J. B., & Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, 145, 30–42.
- Phillips, J., Misenheimer, L., & Knobe, J. (2011). The ordinary concept of happiness (and others like it). *Emotion Review*, 3(3), 320–322.
- Samland, J., & Waldmann, M. R. (2016). How prescriptive norms influence causal inferences. *Cognition*, 156, 164–176.
- Sytsma, J. (2020). Causation, responsibility, and typicality. *Review of Philosophy and Psychology*, 1–21.
- Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, 116(1), 87–100.
- Willemsen, P., & Kirfel, L. (2019). Recent empirical work on the relationship between causal judgements and norms. *Philosophy Compass*, 14(1), e12562.
- Young, L., & Phillips, J. (2011). The paradox of moral focus. *Cognition*, 119(2), 166–178.