ELSEVIER

Original Articles

No luck for moral luck

Markus Kneer[a],[*],[1], Edouard Machery[b],[1]

[a] University of Zurich, Rämistrasse 66, 8001 Zurich, Switzerland
[b] University of Pittsburgh, 4200 Fifth Avenue, 15260 Pittsburgh, USA

ARTICLE INFO

ABSTRACT

Moral philosophers and psychologists often assume that people judge morally lucky and morally unlucky agents differently, an assumption that stands at the heart of the Puzzle of Moral Luck. We examine whether the asymmetry is found for reflective intuitions regarding wrongness, blame, permissibility, and punishment judgments, whether people's concrete, case-based judgments align with their explicit, abstract principles regarding moral luck, and what psychological mechanisms might drive the effect. Our experiments produce three findings: First, in within-subjects experiments favorable to reflective deliberation, the vast majority of people judge a lucky and an unlucky agent as equally blameworthy, and their actions as equally wrong and permissible. The philosophical Puzzle of Moral Luck, and the challenge to the very possibility of systematic ethics it is frequently taken to engender, thus simply do not arise. Second, punishment judgments are significantly more outcome-dependent than wrongness, blame, and permissibility judgments. While this constitutes evidence in favor of current Dual Process Theories of moral judgment, the latter need to be qualified: punishment and blame judgments do not seem to be driven by the same process, as is commonly argued in the literature. Third, in between-subjects experiments, outcome has an effect on all four types of moral judgments. This effect is mediated by negligence ascriptions and can ultimately be explained as due to differing probability ascriptions across cases.

1. Introduction

1.1. Moral luck

Sally had a few too many at the bar yet decides to drive home. On the way, a child stumbles in front of her car, is run over, and dies. John, having had just as much to drink as Sally, also drives home yet arrives at his house without incident. According to philosophers, people are inclined to judge Sally as more blameworthy than John, or her action as morally worse than John's. We will call the distinct assessment of the two actions or of the two agents the *Difference Intuition*. The Difference Intuition appears to be in conflict with the widely held *Control Principle*, according to which agents are only morally responsible for features of their actions that are under their control. The *Puzzle of Moral Luck* arises due to the fact that both the Difference Intuition and the Control Principle are plausible, yet appear in conflict with one another (Nagel, 1979; Williams, 1981; for a review, see Zimmerman, 2006; we limit ourselves to *resultant* moral luck, and do not address any of the other kinds discussed by Nagel, 1979).

What could explain the Difference Intuition? Much evidence

supports the hypothesis that outcome has an effect on moral judgment (Alicke & Davis, 1989; Alicke, Davis, & Pezzo, 1994; Baron & Hershey, 1988; Lowe & Medway, 1976; McKillip & Posavac, 1975; Walster, 1966; for a meta-analysis, see Robbennolt, 2000). According to the *Simple View* (not necessarily endorsed by the authors just listed), the Difference Intuition is just the product of an unmediated effect of outcome on moral judgment (Fig. 1).

More complex models propose that the effect of outcome on moral judgments is mediated by the attribution of certain mental states. According to the *Epistemic Single Pathway Model*, the difference in moral judgment results from a difference in epistemic states ascribed to the agent (Fig. 2; our typology extends the one proposed by Nichols, Timmons, & Lopez, 2014).

To take an example, Royzman and Kumar (2004) suggest that the Difference Intuition is due to an "epistemically corrupted evaluation" of the agent and her situation. Falling prey to the "I know, you know" bias (on which see Royzman, Cassidy, & Baron, 2003), people unconsciously and inappropriately project their superior knowledge about the outcome onto the agent: They are more inclined to ascribe knowledge about the harmful outcome h or awareness of a substantial probability

Fig. 1. The Simple View.

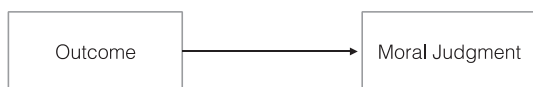Fig. 2. The Epistemic Single Pathway Model.

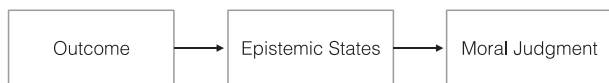
Fig. 4. Cushman's Dual Process Model of moral judgment.

Fig. 5. The Epistemic Dual Pathway Model.

of *h*'s occurrence to the agent in the unlucky condition than in the lucky condition. The more severe moral evaluation of the unlucky agent is thus not due to a direct impact of outcome on moral judgment, but instead due to a higher propensity to attribute the inculpating mental state of foresight (knowledge that *h* will occur) or recklessness (awareness of a substantial probability of *h*).

A third model, the *Probabilistic Single Pathway Model*, explains the difference in moral judgment across cases in terms of the hindsight bias (on the hindsight bias, see Fischhoff, 1975; for a review see Hawkins & Hastie, 1990; for a meta-analysis see Christensen-Szalanski & Willham, 1991). According to this third model, the impact of outcome on moral judgment is due not to an outcome-driven asymmetry in epistemic state ascriptions, but due to an outcome-driven asymmetry in the assessment of the likelihood of the harmful consequence. In other words, outcome knowledge triggers a bias not with respect to the assessment of the agent's mental states, but—in line with Fischhoff's original findings of "creeping determinism"—a bias with respect to the perceived probability of events. The effect of outcome on moral judgment would consequently not be mediated by foresight or recklessness attributions, but rather by negligence attributions, i.e. judgments about whether an agent *should* have been aware of a certain risk (a risk being the probability of a harmful event), see Fig. 3.

In retrospect, the probability of an accident is perceived as higher in the unlucky case, the unlucky agent is judged as more negligent than the lucky one, and the unlucky agent—who is perceived as more negligent—is judged more harshly. Given that it is commonplace to expound the Puzzle of Resultant Moral Luck with reference to acts of negligence, it is surprising that the Probabilistic Single Pathway Model has never been tested.

The analysis in terms of an epistemic or probabilistic *bias* is not the only way to account for the Difference Intuition. Instead, one might deem outcome a valid source of information for the ascription of epistemic states to the agent (for suggestions along these lines, see Heider, 1958; Richards, 1986; Thomson, 1993; Rosebury, 1995; for discussion see Nichols et al., 2014 and Young, Nichols, & Saxe, 2010). John's belief that no accident will occur, if well justified, can be taken to constitute knowledge. Sally's belief, by contrast, simply cannot amount to knowledge: Knowledge is a factive mental state and Sally's belief that no accident will occur turns out false.

The probabilistic model might also be given a rationalist spin: The fact that an accident did indeed occur in Sally's case might be taken as evidence that the probability of such an accident was high, whereas the fact that it did not occur might constitute good grounds to infer that the probability of an accident was low. Consequently, one might be disposed to hold that Sally should have been aware of a substantial risk of an accident (and hence acted negligently). It is, however, not the case that John should have been aware of a substantial risk of an accident, for the simple reason that the risk is not deemed substantial in the first place.

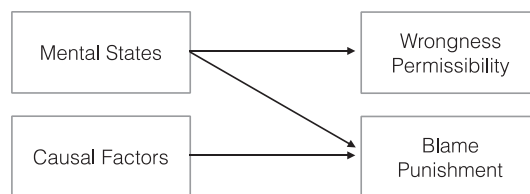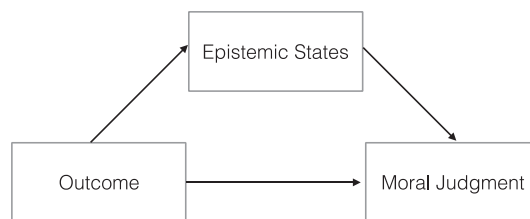Fig. 3. The Probabilistic Single Pathway Model.

Considerations of this sort engender a debate as to the appropriate type of probability for moral assessment: *ex ante* probability, according to which what is probable or not is assessed at the time of action, or *ex post* probability, where the probability of the event in question takes outcome information into consideration. Note that both in criminal law and ethics (except for hard-nosed consequentialist views) the definition of negligence draws on an *ex ante* concept of probability. Although it might thus be *in some sense* rational to ascribe different probabilities of an accident across the two cases, this does not yet make it rational to judge one agent more negligent than the other.

In contrast to the *Single Pathway* accounts just introduced, the *Dual Pathway Model* proposes that moral judgment is sensitive to both mental states (first pathway) and causal factors and outcomes (second pathway). Cushman (2008, 2013) proposes a view of this sort, though it is complicated by a further, orthogonal, distinction according to which blame and punishment judgments are sensitive to causal factors and mental states, whereas wrongness and responsibility judgments are sensitive mainly to mental states. Cushman calls his influential account the *Dual Process Model* of moral judgment (Fig. 4) since different types of judgment are processed with respect to different factors. It should not be confused with *Dual Pathway Models* of moral luck.

According to the *Epistemic Dual Pathway Model* (discussed, though not endorsed, by Nichols et al. (2014)), the impact of outcome is partially mediated by epistemic state ascriptions: Outcome directly influences moral judgment, but also affects the degree to which knowledge and belief are attributed, influencing moral judgment indirectly (Fig. 5). Just as for Single Pathway Models, an analogue drawing on probability ascriptions rather than epistemic state attributions can be envisioned (Fig. 6). Both types of Dual Pathway Models again allow for a rationalist or a bias-driven interpretation.[2]

----

[2] Yet other models exist. Young et al. (2010), for instance, propose an account according to which the asymmetry in blame ratings across the lucky and unlucky cases is driven by epistemic factors which (though dependent on truth-conferring features of outcome) are independent of the *severity* of outcome. The unlucky agent, they argue, frequently has a false outcome-related belief (e.g. that the bad consequence will not occur) whereas the lucky agent does not hold a false belief, and the more severe moral evaluation of unlucky agents is principally due to their false outcome-related beliefs. We will set this model aside for two reasons: First, we think that the model only applies to the subclass of moral luck cases in which a relevant outcome-related target belief *can* be clearly singled out. However, moral luck is believed to arise principally in situations of negligence, situations where the agent should have had, but *did not actually entertain* a belief about a substantial risk regarding a harmful outcome (and hence *a fortiori* did not have a true or false outcome-related belief). Second, in two experiments (Studies 3 and 4a) where we explicitly tested a wide
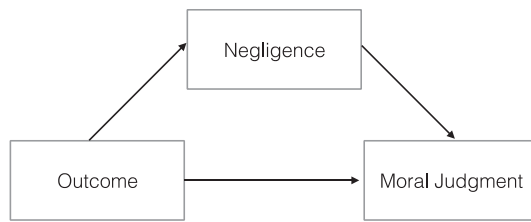
Fig. 6. The Probabilistic Dual Pathway Model.

1.2. Empirical data

Several empirical studies address moral luck directly (Cushman, Dreber, Wang, & Costa, 2009; Gino, Moore, & Bazerman, 2009; Lench, Domsky, Smallman, & Darbor, 2015; Young et al., 2010) or employ moral luck scenarios to explore related topics in moral psychology (Cushman, 2008; Gino, Shu, & Bazerman, 2010; Schwitzgebel & Cushman, 2012). Across a broad range of scenarios, the unlucky agent is robustly judged more harshly than the lucky agent (for a recent review of the empirical literature on moral luck, see Martin & Cushman, 2015). However, the literature suffers from a variety of shortcomings. First, most experiments use a between-subjects design, in which participants are assigned to either the good luck or the bad luck condition. While studies of this sort testify to an effect of outcome on moral judgment, they do not explore whether reflective judgments also manifest the Difference Intuition. Designs allowing participants to compare the two conditions are more likely to elicit such reflective judgments (e.g., Hsee, 1996; Baron, 2000; Baron & Ritov, 2004). In fact, in the few within-subjects studies (Spranca, Minsk, & Baron, 1991; Schwitzgebel & Cushman, 2012) about two thirds of the subjects judge the lucky and the unlucky agent identically in moral respects.

Second, most studies examine judgments about concrete moral luck situations, though less is known about people's abstract views about moral luck. Some evidence suggests that most people do not, in fact, endorse the Difference Intuition when explicitly asked about it (90% according to Nichols, 2009; 80% according to Schwitzgebel & Cushman, 2012; 85% and 63% in Studies 2 and 3 of Lench et al., 2015). The apparent conflict between explicitly endorsed abstract principles and judgments about concrete cases requires further exploration.

Third, there is a staggering variety in types of judgments that are invoked across the different experiments. Participants are asked to rate whether the agents' behavior is "wrong," "fair," "unethical," "blameworthy," "bad," "should be a crime," or "deserves punishment." This variety complicates the comparative assessment of results across studies, in particular since the different types of judgment are frequently conflated. Schwitzgebel and Cushman (2012), for instance, ask about blame in their study invoking concrete cases, yet explore punishment judgments when they ask participants about their explicitly endorsed abstract principles. Lench et al.'s (2015) questionnaire about participants' explicit commitment to the abstract principle of moral luck invokes at least four different types of moral judgment, which leaves it obscure what the dependent variable actually is. The Puzzle of Resultant Moral Luck and the Difference Intuition are often formulated in terms of blame and wrongness, and not just any blame-related judgment will do (for discussion, see Enoch & Marmor, 2007; Levy, 2005; but see, e.g., Kumar, 2018).

Fourth, the existing empirical literature is inconclusive as to which of the psychological models of moral luck, if any, is correct. Some studies suggest that outcome has a strong effect on the relevant judgments (Gino et al., 2009); others suggest it does not (Lench et al., 2015),

---

(footnote continued)

range of epistemic state ascriptions (knowledge, belief, and awareness of a substantial risk), we found no evidence for a significant difference across the lucky and unlucky cases.

although design and variables differed across experiments. In support of the Dual Pathway Model, Cushman (2008) reports that outcome strongly affects blame and punishment, while it affects wrongness and permissibility judgments only marginally. The findings of Gino et al. (2009), however, suggest that judgments of wrongness or "ethicality" are just as affected by outcome as blame and punishment. Young et al. (2010) report that moral evaluation is partially mediated by the perceived justification of the agent's belief, though outcome also seems to have a partial direct effect on blame. Entertaining counterfactuals—e.g. thinking about a negative outcome in a neutral outcome scenario—is sometimes found to affect moral judgment (Lench et al., 2015).

Finally, there has been little work exploring whether intuitions regarding moral luck correlate with psychological, ideological and demographic factors. An exception is Gino et al. (2009), who report that participants primed to engage in slow, analytic cognition manifest the Difference Intuition to a lesser degree than those primed to engage in intuitive cognition.

In the following, we will report a systematic series of studies that explore moral luck intuitions for two distinct scenarios and address these shortcomings. For each scenario, we ran a between-subjects experiment, a within-subjects experiment, and a contrastive design experiment (i.e., an experiment in which participants are presented with a single question comparing the morally lucky and unlucky agent instead of two questions across and within participants; see Section 3). We also collected data on people's abstract views about moral luck in order to compare them with their concrete, case-based judgments. Following Cushman (2008, 2013), each experiment focused on four distinct dependent variables: wrongness, blame, permissibility, and punishment. Furthermore, several studies explicitly targeted the question which, if any, of the psychological models surveyed above best explains the effect of outcome on moral judgment. We also employed a variety of questionnaires to investigate whether moral luck intuitions correlate with particular psychological, ideological, or demographic characteristics, namely the Rational-Experiential Inventory (Epstein, Pacini, Denes-Raj, & Heier, 1996; Pacini & Epstein, 1999), the Belief in a Just World Scale (Rubin & Peplau, 1975), the 12-item Social and Economic Conservatism Scale (Everett, 2013), and the 20-item Moral Foundations Questionnaire (Graham, Haidt, & Nosek, 2009; Graham et al., 2011). The results are reported in Sections 3.2.2 and 7.3 of the Appendix.

The article will proceeds as follows: Studies 1a, 1b, and 2 explore the robustness of the effect of outcome on moral judgment by employing a between-subjects design, a within-subjects design, and a contrastive design respectively. Study 3 investigates whether the effect of outcome on moral judgment is mediated by epistemic states and probability ascriptions. Studies 4a, 4b, 4c, and 4d, report replications of the findings. Study 5 focuses on participants' abstract views about moral luck. The general discussion examines the significance of these findings for moral psychology and moral philosophy and points out future avenues of research on the topic.

2. Study 1a (between-subjects design) and Study 1b (within-subjects design)

2.1. Methods and materials

For Study 1a, we used a mixed-factorial design (between-subjects factor: Moral Luck—lucky vs. unlucky; within-subjects factor: Judgment Type—wrongness vs. permissibility vs. blame vs. punishment). Having passed an attention check, participants were randomly presented with one of the following two vignettes:

Bad Luck
Anna is at home, giving her 2-year-old son a bath. She fills the bath, while her son stands near the tub. The phone rings in the next room. Anna tells her son to stand near the tub while she answers the phone. Anna believes her son will stand near the tub for a few

minutes and wait for her to return. Anna leaves the room for 5 min. When Anna returns, her son is in the tub, dead, face down in the water.

Good Luck

Beth is at home, giving her 2-year-old son a bath. She fills the bath, while her son stands near the tub. The phone rings in the next room. Beth tells her son to stand near the tub while she answers the phone. Beth believes her son will stand near the tub for a few minutes and wait for her to return. Beth leaves the room for 5 min. When Beth returns, her son is still standing near the tub, where she left him. The boy then enjoys his bath.

Participants were asked to answer four moral judgment questions. The wrongness question read: "How wrong was Anna [Beth] to leave her son alone in the above scenario?". Responses were recorded on a 7-point Likert scale anchored at 1 with "not wrong at all" and at 7 with "extremely wrong." The three other questions (with adapted labels of the scale) read:

(1) To what extent was Anna [Beth] blameworthy for leaving her son alone in the above scenario? (not at all blameworthy/extremely blameworthy)
(2) To what extent was it permissible for Anna [Beth] to leave her son alone in the above scenario? (entirely permissible/entirely impermissible)
(3) How much punishment does Anna [Beth] deserve for leaving her son alone in the above scenario? (no punishment at all/very severe punishment)

These four questions were presented in randomized order on separate screens. They were followed by a comprehension check and a demographic questionnaire.

Study 1b was identical in all respects, except that participants were presented with *both* vignettes (bad luck always first), and then had to judge *both* Anna and Beth's actions in terms of all four measures (wrongness, blame, permissibility, and punishment), presented in randomized order. The wrongness question, for instance, read "How wrong were Anna and Beth to leave their sons alone in each of the above scenarios?". Participants had to rate Anna's action, and thereafter Beth's action, on separate Likert scales ranging from 1 ("not wrong at all") to 7 ("extremely wrong").

2.2. Participants

For Study 1a, 241 participants were recruited online via Amazon Mechanical Turk. The IP address location was restricted to the USA. Participants who failed the attention check or the comprehension question were excluded, leaving a sample of 196 participants (male: 40.8%; mean age: 30.1, range: 18–73).

For Study 1b, 120 participants were recruited online via Amazon Mechanical Turk. The IP address location was again restricted to the USA. Participants who failed the attention check or the comprehension question were excluded, leaving a sample of 95 participants (male: 48.4%; mean age: 37.2, range: 20–69). Analyses with the complete data sets are reported in Sections 1 and 2 of the Appendix (excluding inattentive subjects did not make an important difference).

2.3. Results for Study 1a

A mixed-design ANOVA determined that, aggregating across the four dependent variables, participants judged the action of the morally unlucky agent to be worse than the action of the morally lucky agent ($F(1,194) = 2831.62$, $p < .001$, $\eta^2 = .94$, Fig. 7). Furthermore, it revealed that, aggregating across the two moral luck conditions, the difference in judgment type was significant ($F(3,582) = 162.77$,

$p < .001$, $\eta^2 = .46$). Bonferroni-corrected post hoc tests showed that the only significant differences were (1) between answers to the punishment question and answers to the three other questions and (2) between answers to the permissibility question and the wrongness question (uncorrected $ps < .001$). We did not find any evidence that participants responded differently to the other pairs of questions (blame and wrongness: uncorrected $p = .50$; wrongness and permissibility: uncorrected $p = .01$; blame and permissibility: uncorrected $p = .008$). Importantly, the two main effects were qualified by an interaction ($F(3,582) = 9.89$, $p < .001$, $\eta^2 = .04$). To analyze this interaction, we compared the difference between the morally lucky and the morally unlucky conditions for each of the four questions. We then computed the effect size (Cohen's d) for each effect (Morris and DeShon, 2002) as well as the confidence intervals for each of these four effect sizes (Wuensch, 2012). The results are summarized in Table 1.

2.4. Results for Study 1b

To examine the influence of moral luck on moral judgments, we analyzed participants' answers by means of a two-way (Moral Luck: lucky vs. unlucky; Judgment Type: wrongness vs. permissibility vs. blame vs. punishment) repeated-measures ANOVA. We found that aggregating across the four judgment types, participants' mean responses for the lucky condition differed significantly from the unlucky condition ($F(1,94) = 25.00$, $p < .001$, $\eta^2 = .21$; Fig. 7). We also found that, aggregating across the two moral luck conditions, participants' mean answers to the wrongness, permissibility, blame, and punishment questions differed significantly ($F(3,282) = 105.55$, $p < .001$, $\eta^2 = .34$). Bonferroni-corrected post hoc tests revealed that the only significant differences were between answers to the punishment question and answers to the three other questions (uncorrected $ps < .001$; all the other $ps > .5$). That is, participants answered the punishment question differently from the wrongness, permissibility, and blame questions. We did not find any evidence indicating that participants responded differently to the wrongness, permissibility, and blame questions. Importantly, the two main effects we observed were qualified by a two-way interaction ($F(3,282) = 16.18$, $p < .001$, $\eta^2 = .15$). To analyze this interaction, we compared the difference between the morally lucky and the morally unlucky conditions for each of the four questions. We then computed the effect size for each effect as well as the confidence intervals for each of these four effect sizes, using the 2016 ESCI software. The results are summarized in Table 1.

To investigate whether the observed aggregate influence of moral luck on wrongness, blame, and punishment judgments is widespread among participants or is rather due to a small minority, we compared the proportion of participants manifesting and failing to manifest the Difference Intuition. Participants who judged the lucky and the unlucky agents identically were counted as manifesting "no Difference Intuition" (*No DI* for short in Fig. 8). The vast majority of participants did not share the Difference Intuition for wrongness (89%), blame (87%), and permissibility (96%), though the proportion was substantially smaller for punishment (64%). All proportions differed significantly from chance (binomial test $p < .001$, two-tailed). Punishment judgments differed significantly from wrongness, blame, and permissibility judgments (binomial tests, test proportion = .64, $p < .001$). Permissibility judgments differed significantly not only from punishment judgments, but also from wrongness judgments (binomial test, test proportion = .96, $p = .010$) and blame judgments (binomial test, test proportion = .96, $p < .001$). Wrongness judgments did not differ significantly from blame judgments (binomial test, test proportion = .89, $p = .702$, all two-tailed).

2.5. Discussion

Taken together, Studies 1a and 1b revealed four main findings. First, at the aggregate level an effect of outcome on moral judgment was
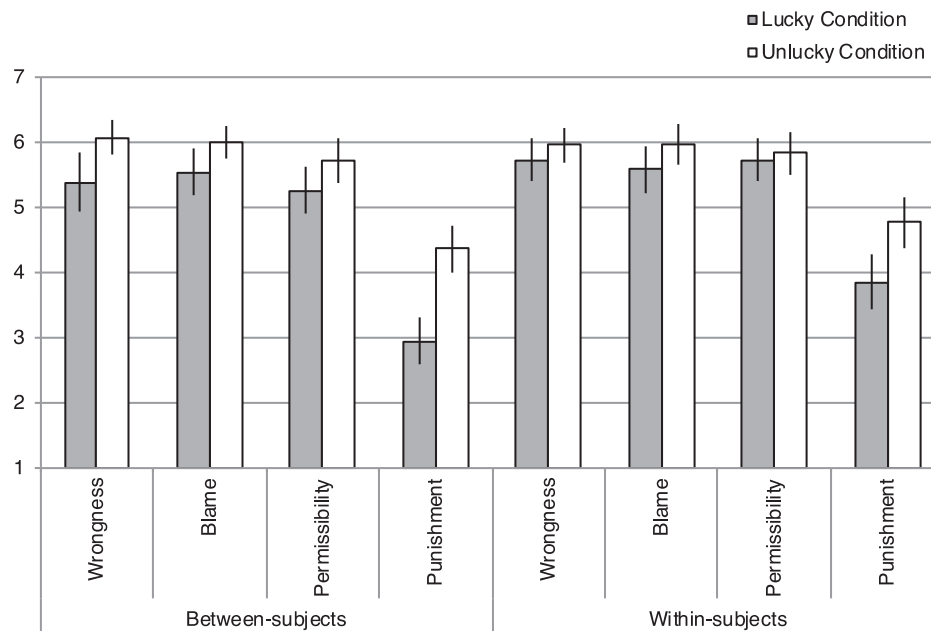
Fig. 7. Mean wrongness, blame, permissibility, and punishment judgments for the between-subjects and within-subjects designs; error bars denote 95% confidence intervals.

Table 1

Effect of outcome on wrongness, blame, permissibility, and punishment judgments in a between-subjects and a within-subjects design; 95% confidence intervals are given for the effect sizes.

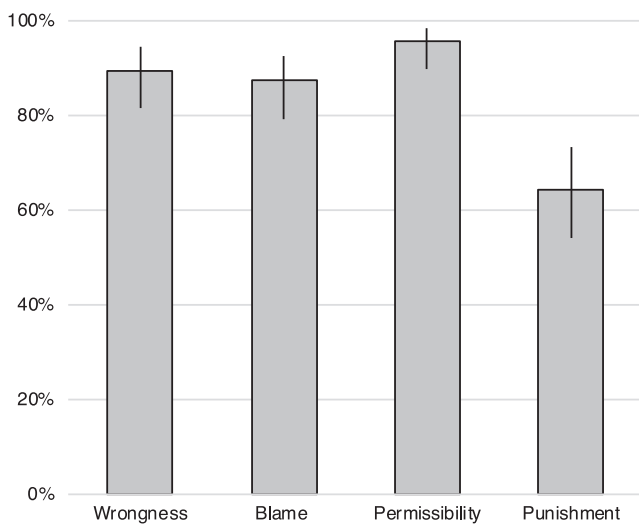|	Between-subjects Design				Within-subjects Design			
Variable	t(194)	*p*	Cohen's *d*	95% CI	t(94)	*p*	Cohen's *d*	95% CI
Wrongness	3.08	.002	.44	[.16;.72]	2.80	.006	.16	[.004;.27]
Blame	2.09	.04	.39	[.17;.58]	3.28	.001	.24	[.09;.38]
Permissibility	1.85	.07	.26	[−.02;.55]	1.55	.12	.06	[−.02;.14]
Punishment	5.52	< .001	.79	[.50;1.08]	5.84	< .001	.47	[.30;.64]

Fig. 8. Proportions of participants who judged the lucky and unlucky agents identically (*No DI*) with respect to wrongness, blame, permissibility, and punishment; errors bars denote 95% confidence intervals; Wilson method, see Brown, Cai, & DasGupta, 2001.

found in both between- and within-subjects designs, although the effect sizes in the within-subjects design were smaller than in the between-subjects design. Second, for a very large majority of people the outcome of the agent's actions did not influence permissibility, blame, and

wrongness judgments when they read about both a morally lucky and unlucky situation. This result is in line with previous findings by Spranca, Minsk, and Barron (1991) and Schwitzgebel and Cushman (2012) reported in the introduction. It suggests that outcome does not influence most people's reflective (i.e. within-subjects design) judgments about concrete cases of moral luck. Third, the influence of moral luck varied across types of moral judgment: We found a medium to large effect of outcome on punishment judgments, but only a small to medium effect of outcome on wrongness and blame judgments. Admittedly, the confidence intervals for the effect sizes overlap in Study 1a, which prevents us from concluding that the effect size of punishment is significantly larger than the effect sizes for wrongness and blame. But the lower bound of the confidence intervals for the effect size of punishment is larger than the higher bound of the effect size of wrongness in Study 1b. We will come back to this question in Study 4. We failed to find any significant effect of outcome on permissibility judgments. Fourth, in light of the existing literature, the way the four different types of moral judgment related to one another is unexpected. Cushman (2008, 2013) reports that punishment patterns with blame, but we observed that blame patterns with wrongness and differs from punishment.

Across the first two studies, we have varied the factor of explicit comparability. Study 1a used a between-subjects design (low comparability), Study 1b a within-subjects design (medium comparability). In Study 2 we explore participants' responses to a question that explicitly contrasts the two agents (high comparability), i.e. a question that asks whether Anna and Beth should be judged identically. We will refer to this type of design as a "contrastive design."

3. Study 2: Contrastive design

Study 2 employed a contrastive design to further increase participants' reflective comparison of the two types of cases.

3.1. Methods and materials

Following an attention check, participants were presented with both vignettes of Study 1b (the order was fixed: bad luck first). The formulation of the wrongness, blame, permissibility, and punishment questions differed from Study 1b. In Study 2, participants were presented with statements explicitly comparing Anna and Beth with respect to each of the four moral measures:

(1) It was just as wrong for Anna to leave her son as it was for Beth.
(2) Anna is just as blameworthy as Beth for leaving her son.
(3) It was just as impermissible for Anna to leave her son as it was for Beth.
(4) Anna deserves just as much punishment as Beth for leaving her son.

Each statement (presented in randomized order) was followed by a 7-point Likert scale anchored at 1 with "completely agree" and at 7 with "completely disagree."

3.2. Participants

120 participants were recruited online via Amazon Mechanical Turk. We restricted the IP address location to the USA. Participants who failed the attention check or the comprehension question were removed, leaving a sample of 89 participants (male: 48.3%; mean age: 36.9, SD = 12.1; range: 18–65). Analyses with the complete data sets are reported in the Appendix.

3.3. Results

A one-way repeated-measures ANOVA determined that participants' mean answers to the wrongness, blame, permissibility, and punishment questions differed significantly ($F(3,264) = 39.51$, $p < .001$, $\eta^2 = .24$; Fig. 9). Post hoc tests using the Bonferroni correction revealed that the only significant comparisons were those between answers to the punishment question and answers to the three other questions (uncorrected $ps < .001$; wrongness and permissibility, uncorrected $p = .38$; wrongness and blame, uncorrected $p = .05$; blame and permissibility,
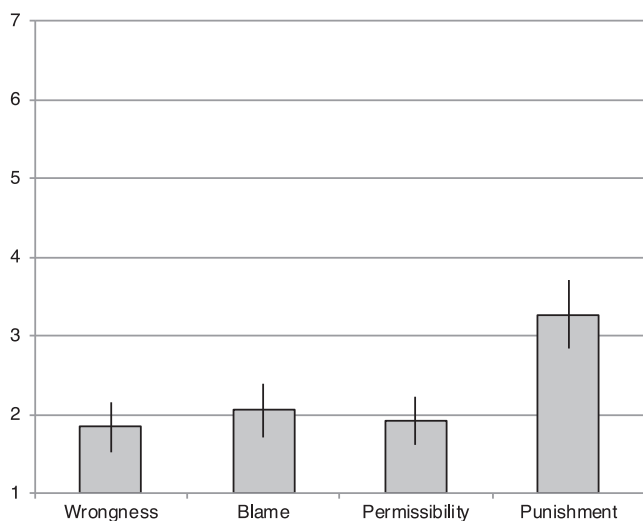

Fig. 9. Mean disagreement with the claim that the two agents should be judged identically with respect to wrongness, blame, permissibility, and punishment; error bars denote 95% confidence intervals.

Table 2
Proportions of participants who agreed (Likert scale < 4) or completely agreed (endpoint 1) that the two agents should be judged identically with respect to wrongness, blame, permissibility, and punishment.

Measure	Wrongness	Blame	Permissibility	Punishment
Likert scale < 4	87%	87%	82%	52%
Endpoint 1	62%	60%	57%	32%

uncorrected $p = .34$). Participants disagreed more with the claim that the morally lucky and morally unlucky agents deserved the same punishment than they disagreed with the claims that their actions were equally wrong, equally impermissible, and equally blameworthy, and we did not find any evidence that they responded differently to these three latter claims. The means for all four measures were significantly below the neutral midpoint 4 and significantly above the endpoint 1 (complete agreement), see Section 3.1 of the Appendix.

Following Lench et al. (2015), we calculated the percentage of participants who agreed with the claim that the two agents should be judged identically. We also aggregated the number of participants who responded with "completely agree" (Likert scale endpoint 1). The results, presented in Table 2, were consistent with the findings from Study 1b: For wrongness, blame, and permissibility, the large majority of participants (over 80%) agreed that the two agents should be judged identically (Likert scale < 4). Over half of the participants chose the endpoint of the Likert scale. As regards punishment, by contrast, only about half of the participants agreed that the two agents deserve the same punishment, and less than a third completely agreed with an assessment of this sort.

3.4. Discussion

Study 2 replicates and extends the findings of Study 1b. People rather uniformly hold that the two agents should be judged the same in terms of wrongness, blame, and permissibility. Mean agreement levels differ significantly from the scale midpoint; over 80% of the responses are on the "agree" spectrum of the Likert scale and the majority "completely agrees" with an identical assessment in terms of wrongness (62%), blame (60%), and permissibility (57%). As in Studies 1a and 1b, outcome had a much larger impact on punishment judgments than on wrongness, blame, and permissibility judgments, and only a minority of participants (31%) completely agreed that the actions of the morally lucky and unlucky agents deserve equal punishment. Consistent with Studies 1a and 1b, blame judgments patterned with wrongness judgments (and in this experiment with permissibility, too), but differed from punishment judgments.

According to Studies 1b and 2, which employed a within-subjects and contrastive design respectively, most people do not share the Difference Intuition for wrongness, blame, and permissibility. This raises the question of whether there is a Puzzle of Resultant Moral Luck in the first place, a question we address in the general discussion. While the majority of previous moral luck experiments have consistently found that outcome influences moral judgment, this is presumably due to the fact that they employ a between-subjects design. Our between-subjects results from Study 1a replicate these between-subjects findings, which constitute an important phenomenon in its own right: In ordinary life, moral and legal judgment resembles the between-subjects design, as we rarely compare actual situations to counterfactual ones with alternative outcomes. In Study 3 we turn to the question of what explains the effect of outcome on moral judgment in non-comparative contexts.

4. Study 3: Mediators of the impact of outcome on moral judgment

What explains the different moral assessment of the lucky and

unlucky agents in non-comparative contexts? In the introduction we have surveyed several accounts. According to the Simple View, the difference in moral judgment arises due to a brute effect of outcome on moral judgment. Alternatively, the effect of outcome on moral judgment might be mediated either by ascriptions of epistemic states or else perceptions of the probability of a harmful outcome across cases. The mediation could be complete (Single Pathway Model) or partial (Dual Pathway Model), in which case parts of the outcome effect on judgment would be direct, and parts of it mediated by epistemic state ascription or probability assessment.

4.1. Methods and materials

In a 2 (Moral Luck—lucky vs. unlucky) × 2 (Judgment Type—wrongness vs. punishment) between-subjects design study employing the two vignettes from Study 1a, participants were randomly assigned to one of the four conditions. They responded either to the wrongness question or the punishment question on a 7-point Likert scale. Thereafter, participants were presented with six questions regarding the agent's epistemic states (belief and knowledge), negligence, and the probability of an accident (the order was randomized). The questions (except for the last one) asked to what extent participants agreed or disagreed with the following statements on a 7-point Likert scale anchored at 1 with "completely disagree" and at 7 with "completely agree":

(1) Anna knew an accident would happen.
(2) Anna knew there was a high probability of an accident.
(3) Anna believed an accident would happen.
(4) Anna believed that there was a high probability of an accident.
(5) Anna should have believed that there was a high probability of an accident.
(6) On a scale of 0–100%, how high do you think was the probability of an accident? N.B. 0% means the accident was completely improbable, 100% means it was certain to occur.

Before the main task, all participants completed an attention check. Following the main task, they were presented with a comprehension check.

4.2. Participants

650 participants were recruited online via Amazon Mechanical Turk. We restricted their IP address locations to the USA. Participants failing the attention or comprehension checks were excluded, leaving a sample of 566 participants (male: 51%, mean age: 37.3, SD = 12.4, range: 18–84).

4.3. Results

A 2 (Outcome: lucky vs. unlucky) × 2 (Judgment Type: wrongness vs. punishment) ANOVA revealed a significant main effect for judgment type ($F(1,562) = 134.23$, $p < .001$, $\eta^2 = .193$), a significant main effect for outcome ($F(1,562) = 59.20$, $p < .001$, $\eta^2 = .095$), and a significant interaction ($F(1,562) = 4.50$, $p = .034$, $\eta^2 = .008$). Wrongness judgments in the unlucky condition ($M = 6.00, SD = 1.52$) significantly exceeded those in the lucky condition ($M = 5.17$, $SD = 1.66$), $t(280) = 4.341$, $p < .001$, $d = .52$. Punishment judgments in the unlucky condition ($M = 4.60$, $SD = 1.97$) also significantly exceeded those in the lucky condition ($M = 3.15$, $SD = 1.83$), $t(282) = 6.407$, $p < .001$, $d = .76$ (Fig. 10).

Turning to the ascription of epistemic states and probability judgments (Figs. 10 and 11), *knew p* ascriptions did not differ significantly across the unlucky ($M = 2.22$, $SD = 1.65$) and lucky conditions ($M = 2.25$, $SD = 1.68$), $t(564) = -.216$, $p = .83$, $d = -.03$; *knew probably p* ascriptions did not differ significantly across the unlucky

($M = 3.38$, $SD = 2.00$) and lucky conditions ($M = 3.14$, $SD = 2.05$), $t(564) = 1.423$, $p = .16$, $d = .09$; *believed p* ascriptions did not differ significantly across the unlucky ($M = 2.06$, $SD = 1.50$) and lucky conditions ($M = 2.00$, $SD = 1.47$), $t(564) = .482$, $p = .63$, $d = .04$; *believed probably p* ascriptions did not differ significantly across the unlucky ($M = 2.61$, $SD = 1.80$) and lucky conditions ($M = 2.50$; $SD = 1.81$), $t(564) = .701$, $p = .48$, $d = .06$. By contrast, *should have believed that probably p* ascriptions (i.e., judgments of negligence) differed significantly across the unlucky ($M = 5.72$, $SD = 1.70$) and lucky conditions ($M = 5.18$, $SD = 1.99$), $t(564) = 3.506$, $p < .001$, $d = .30$, and *probability* judgments differed significantly across the unlucky ($M = 58.84$, $SD = 28.41$) and lucky conditions ($M = 47.40$, $SD = 27.74$), $t(564) = 4.839$, $p < .001$, $d = .41$.

Several multiple mediation analyses were conducted using Preacher and Hayes's (2008) macro for SPSS with 20,000 bootstrap samples, testing for all potential mediators simultaneously. Negligence attribution proved a significant mediator between outcome and wrongness ascription. None of the other proposed mediators were significant, although the *a path* of probability reached significance (Fig. 12). A follow-up analysis revealed probability to be a strong mediator between outcome and negligence (Fig. 13).

For punishment, another mediation analysis with 20,000 bootstrap samples was conducted. Both negligence attribution and probability judgments proved significant mediators between outcome and punishment (Fig. 14). Probability again proved a strong (close-to-complete) mediator between outcome and negligence (Fig. 15).

4.4. Discussion

Section 1 introduced several competing explanations of the influence of outcome on moral judgments. One type of model assumes a single pathway of the impact of outcome on moral judgment, which can either be direct (the Simple View), mediated by epistemic state ascriptions (the Epistemic Single Pathway Model) or by probability assessments (the Probabilistic Single Pathway Model). According to the Dual Pathway Model, *some* of the impact of outcome on judgment is direct, and *some* of it is mediated (either by epistemic state ascriptions or probability assessments). Study 3 shows that the impact of outcome on moral judgment is not epistemically mediated, which undermines the corresponding Single and Dual Pathway Models. Participants ascribed the same degree of outcome knowledge and risk awareness to the unlucky and lucky agent. Hence, the effect of outcome on moral judgment does not result from different attributions of the *mens rea* of foreknowledge (knowledge that a harmful outcome will occur) or recklessness (awareness of the risk of a harmful outcome) across conditions. (For the definitions of the *mens reas* of foreknowledge and recklessness, see the US Model Penal Code (1985), 2.02 sections b and c.)

The data makes a strong case in favor of the probabilistic models. On average, when people are presented with a single situation (either the lucky or the unlucky situation), they agree more with the statement that unlucky Anna should have been aware of a high probability of an accident than with the statement that lucky Beth should have been aware of such a risk. That is, people judge the unlucky agent as more negligent than the lucky agent (see the US Model Penal Code, 2.02 (d) for the standard definition of negligence). Negligence attributions proved a partial mediator of the impact of outcome on wrongness and punishment judgments. The impact of outcome on negligence, in turn, is fully mediated by the diverging assessments of the probability of an accident. Overall, the findings are consistent with a Probabilistic Dual Pathway Model, where some of the effect of outcome on wrongness and punishment is mediated by negligence, and some might be direct ("might" because we cannot rule out that mediators distinct from the ones here tested were at play).

A limitation of the studies reported so far is that we used a single vignette to study moral luck, and one may worry that our results
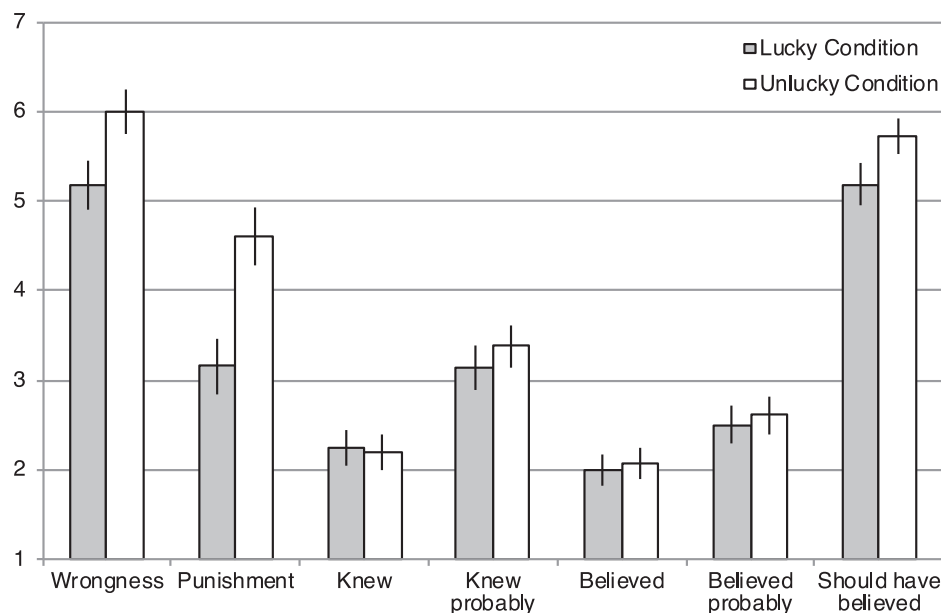
Fig. 10. Mean ascription of wrongness, punishment, epistemic states, and negligence; error bars denote 95% confidence intervals.
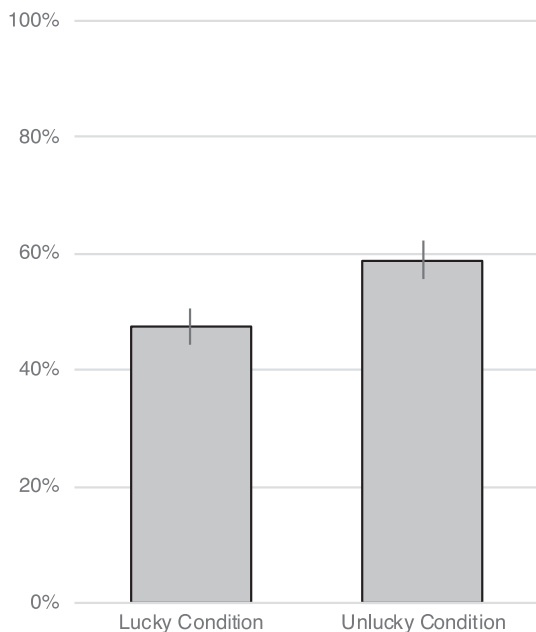

Fig. 11. Mean ascription of probability of the occurrence of a harmful outcome; error bars denote 95% confidence intervals.

depend on the specific formulations of this vignette. Study 4 addresses this question and replicates our findings.

5. Replications

The goal of the following series of studies was to replicate the previous results with a different scenario and an alternative within-subjects design inspired by Experiment 1 of Spranca et al. (1991).

5.1. Study 4a: Between-Subjects replication

5.1.1. Methods and materials

In a mixed-factorial design study (between-subjects factor: Moral Luck—lucky vs. unlucky; within-subjects factor: Judgment Type—wrongness vs. permissibility vs. blameworthiness vs.

punishment), participants were randomly presented with one of the following two vignettes (variations in square brackets):

Bad [Good] Luck

John [Frank] is an artisan toymaker. For his new puppet collection, he changes the paint supplier. He picks a company from abroad that has previously had problems with lead in their paint. Lead is poisonous. John [Frank] believes that the paint no longer has the lead problem, and does not have it tested. Unfortunately, the paint does contain lead. John's puppets sell out and three children who play with them die from lead poisoning. [Fortunately, the paint does not contain any lead. Frank's puppets sell out and none of the children playing with them die from lead poisoning.]

Participants were asked to answer four moral questions about the agents (John or Frank) and their four moral judgment questions. The wrongness question read: "How wrong was John [Frank] in using the new paint without testing it for lead?" Responses were recorded on a 7-point Likert scale anchored at 1 with "not wrong at all" and at 7 with "extremely wrong." The three other questions, whose scale labels were adapted, read:

(1) To what extent is John [Frank] blameworthy for using the new paint without testing it for lead? (not at all blameworthy/extremely blameworthy)
(2) To what extent was it permissible for John [Frank] to use the new paint without testing it for lead? (entirely permissible/entirely impermissible)
(3) How much punishment does John [Frank] deserve for using the new paint without testing it for lead? (no punishment at all/very severe punishment)

The order of the four questions was randomized. The main task was preceded by an attention check and followed by a comprehension question. Having completed the task, participants were presented with questions about the agent's epistemic states and the likelihood of a bad outcome. On a 7-point Likert scale anchored at 1 with "completely disagree" and at 7 with "completely agree" participants had to report their agreement with statements 1–5, and to report their answer to question 6 on a scale of 0% ("no probability at all") to 100% ("certain"). The items were presented in randomized order:
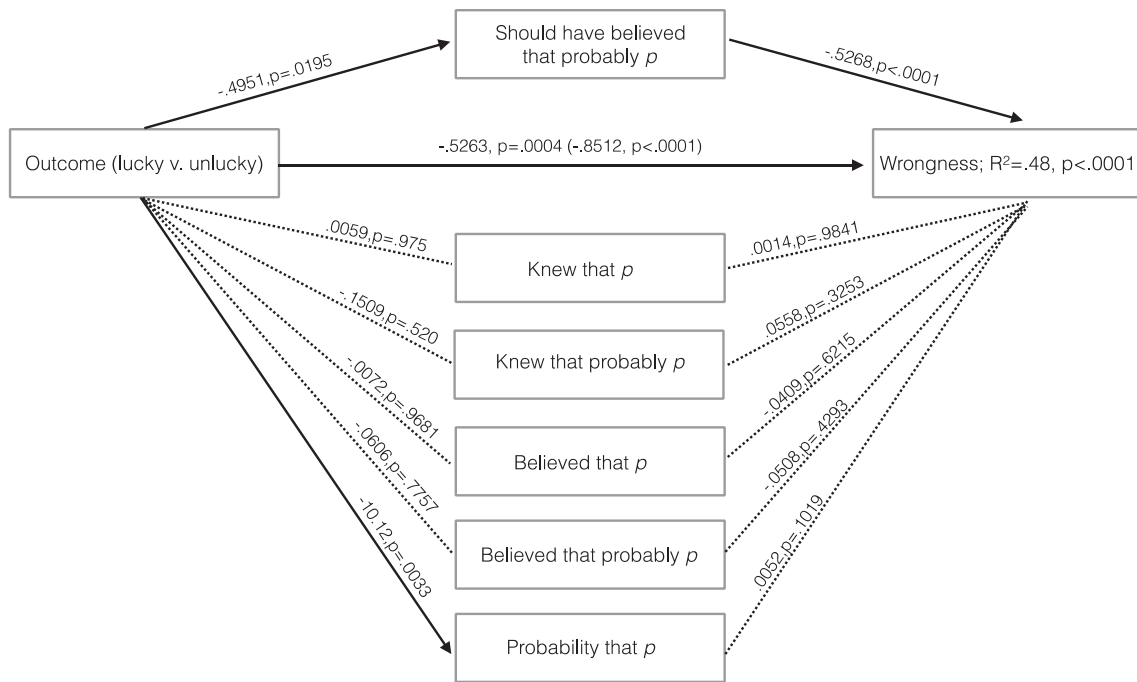
Fig. 12. Mediation of the relationship between outcome (lucky vs. unlucky) and wrongness judgments by probability, negligence, knowledge, and belief ascriptions.
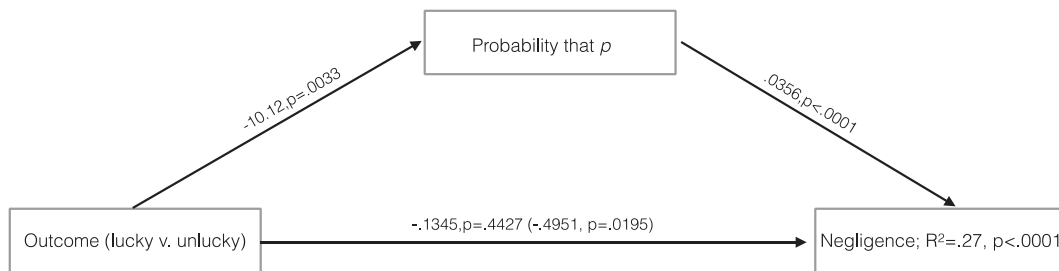

Fig. 13. Mediation of the relationship between outcome (lucky vs. unlucky) and negligence ascriptions ("should have believed there was a high probability of an accident") by probability judgments.

(1) John [Frank] knew the paint still contained lead.
(2) John [Frank] knew there was a high probability that the paint would still contain lead.
(3) John [Frank] believed the paint still contained lead.
(4) John [Frank] believed that there was a high probability that the paint would still contain lead.
(5) John [Frank] should have believed that there was a high probability that the paint would still contain lead.
(6) On a scale of 0–100%, how high do you think was the probability of the paint still containing lead? N.B. 0% means no probability at all, 100% means it was certain that the paint would contain lead.

At the end of the study, participants were also asked to complete a biographical questionnaire.

5.1.2. Participants

332 participants were recruited online via Amazon Mechanical Turk. We restricted their IP address location to the USA. Participants who failed the attention check or the comprehension check were removed, leaving a sample of 254 participants (male: 46.9%; mean age: 37.2, range: 19–81).

5.1.3. Results

A mixed design ANOVA determined that, aggregating across the four judgment types, participants considered the action of the unlucky agent as worse than the action of the lucky agent ($F(1,252) = 24.46$, $p < .001$, $\eta^2 = .09$), and that, aggregating across the two moral luck conditions, they answered the four moral questions differently ($F(3,756) = 83.11$, $p < .001$, $\eta^2 = .25$; Fig. 16). Post hoc tests using the Bonferroni correction revealed that the significant differences were (1) between answers to the punishment question and answers to the three other questions, and (2) between answers to the permissibility question and answers to the three other questions. We did not find any evidence that the participants responded differently to the other pairs of questions.

Importantly, the two main effects were qualified by an interaction ($F(3,756) = 23.38$, $p < .001$, $\eta^2 = .09$). To analyze this interaction, we compared the difference between the lucky and the unlucky conditions for each of the four questions. We then computed the effect size (Cohen's d) for each effect as well as the confidence intervals for each of these four effect sizes (Wuensch 2012). The results are summarized in Table 3.

Figs. 16 and 17 present the results for epistemic state ascriptions and probability assessments; Table 4 reports effect sizes and confidence intervals for each of the six measures.

A series of multiple mediation analyses was conducted to investigate whether the relation between outcome (lucky vs. unlucky) and moral judgment (wrongness, blame, and punishment) was mediated by epistemic state ascriptions or assessments of probabilities. Since for permissibility there was no significant outcome effect, we did not run a
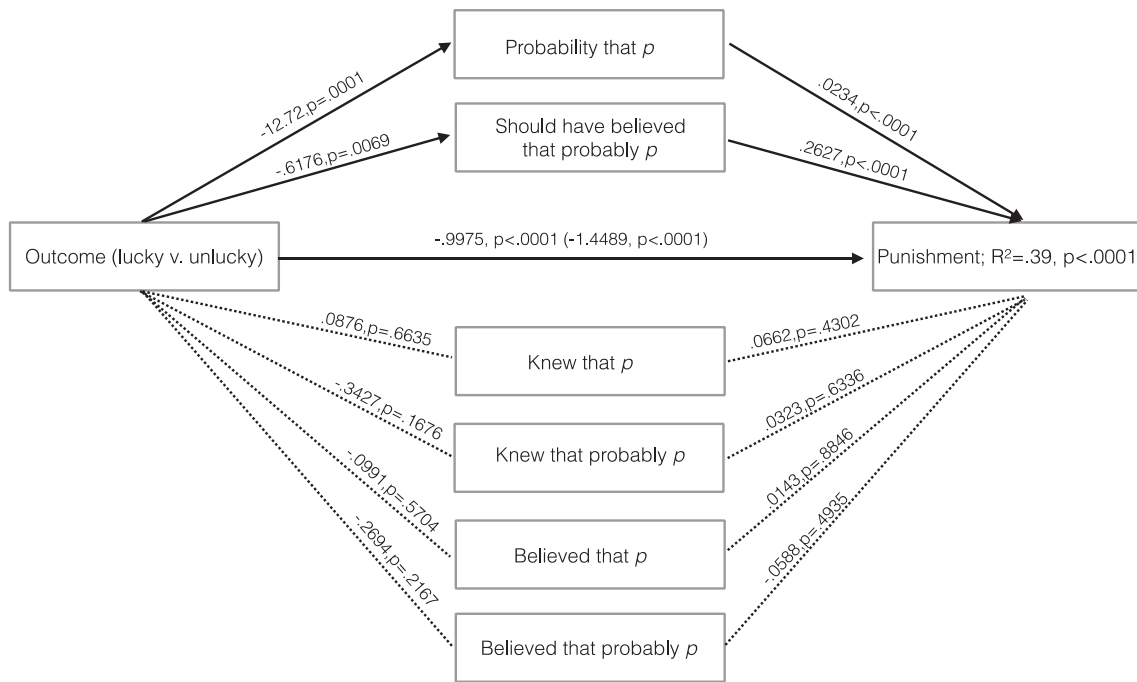
Fig. 14. Mediation of the relationship between outcome (lucky vs. unlucky) and punishment judgments by probability, negligence, knowledge, and belief ascriptions.

mediation analysis. The simultaneously tested potential mediators were that Frank/John (i) *knew that p*, (ii) *knew that probably p*, (iii) *believed that p*, (iv) *believed that probably p*, (v) *should have believed that probably p*, and (vi) the perceived *probability of p*, where *p* stands for "the paint contained lead." The mediation analyses were conducted using Preacher and Hayes's (2008) SPSS macro with 20,000 bootstrap samples.

For all three DVs, negligence ("should have believed that probably *p*") and probability were significant mediators, whereas the other proposed mediators proved insignificant. For both wrongness and blame the unmediated, significant relation between outcome and moral judgment (the c path) was rendered insignificant once the mediators were taken into account (the c' path; Figs. 18 and 19).

Importantly, the relation between outcome and negligence was itself mediated by probability. The significant unmediated impact of outcome on negligence (the c path) was rendered insignificant once probability was taken into account as a mediator (the c' path; Fig. 20).

Punishment was once again an outlier. Whereas for wrongness and blame the c' path was insignificant, for punishment it remained significant even after the mediators had been taken into account (Fig. 21).

The mediation analyses replicate the findings from Study 3: Whereas we could find no evidence that the relation between outcome and moral judgment is mediated by epistemic state ascriptions, negligence and probability were found to be significant mediators. The effect of outcome on wrongness and blame judgments fit the Probabilistic

Single Pathway Model: Once negligence has been controlled for as a mediator, the impact of outcome on wrongness and blame is no longer significant. Punishment judgments, by contrast, fit the Probabilistic Dual Pathway Model: The effect of outcome on punishment is partly mediated by negligence and probability ascriptions, and partly direct (at least with respect to the mediators tested, though others might be possible).

5.2. Study 4b: Within-subjects replication

5.2.1. Methods and materials

To replicate Study 1b, the Toymaker scenario was used in a within-subjects design identical in all respects to the design of Study 1b. Participants were presented with both the Bad Luck and Good Luck vignettes (in that order) and had to assess the action of John and Frank with respect to all four moral judgment types (wrongness, blame, permissibility, and punishment). To take an example, participants were asked "How wrong were John and Frank in using the new paint without testing it for lead?" and had to answer separately for John and Frank on a 7-point scale anchored at 1 with "not wrong at all" and at 7 with "extremely wrong." The order of the four questions was again randomized. The main task was preceded by an attention check and followed by a comprehension question.

Having completed the main task, participants were asked explicitly about their abstract views regarding moral luck. They read the
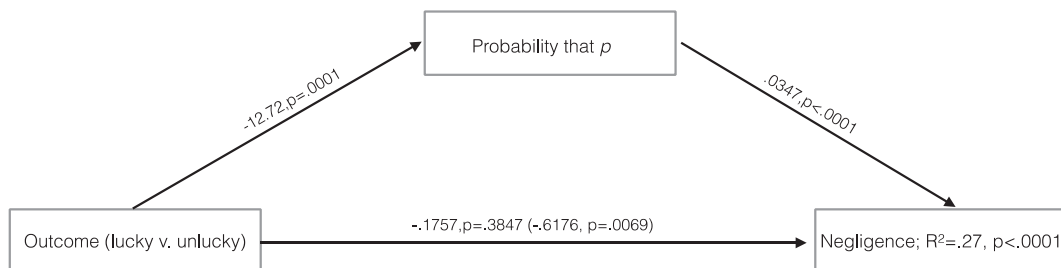

Fig. 15. Mediation of the relationship between outcome (lucky vs. unlucky) and negligence ascriptions ("should have believed there was a high probability of an accident") by probability judgments.
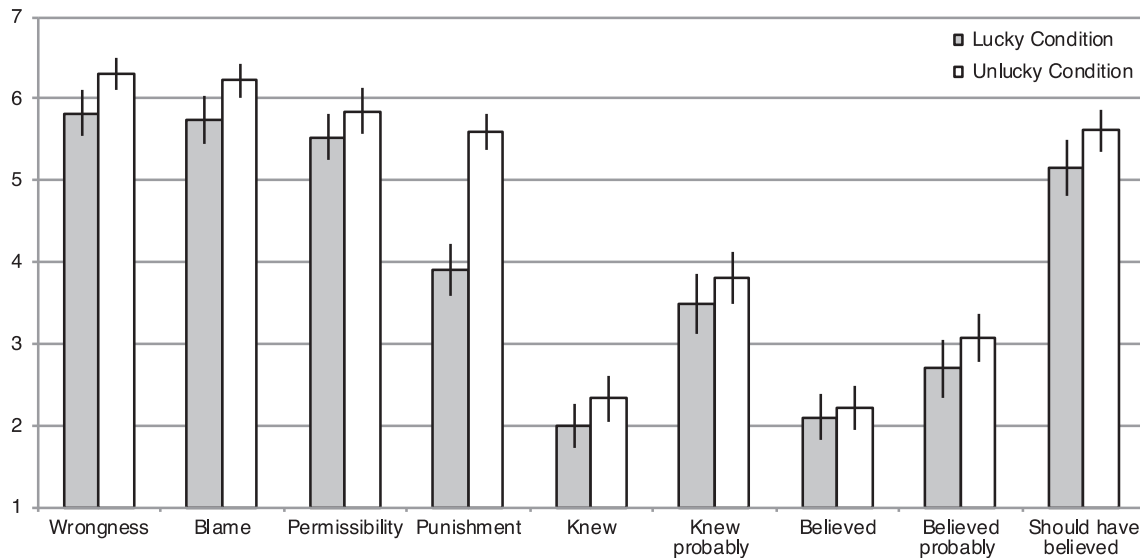
Fig. 16. Mean wrongness, blame, permissibility, and punishment judgments as well as mean epistemic state and negligence ascriptions; error bars denote 95% confidence intervals.

Table 3
Effect of outcome on wrongness, blame, permissibility, and punishment judgments.

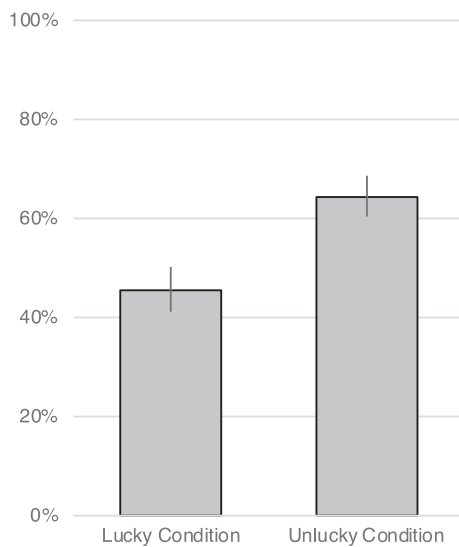| Variable	t(252)	p	Cohen's d	95% CI
Wrongness	2.80	.006	.36	[.16;.63]
Blame	2.67	.008	.33	[.13;.63]
Permissibility	1.53	.13	.06	[−.08;.48]
Punishment	8.50	< .001	1.07	[.84;1.39]

Fig. 17. Mean probability judgments; error bars denote 95% confidence intervals.

following prompt and had to respond to four questions concerning wrongness, blame, permissibility, and punishment on separate screens in randomized order:

> Suppose two people do the exact same thing, with the exact same frame of mind. Then, due entirely to matters of chance beyond their control, one of them produces a very bad outcome, but the other does not. To what extent do you agree or disagree with the following statements:

Table 4
Effect of outcome on ascriptions of epistemic states, negligence, and probability.

Variable	t(252)	p	Cohen's d	95% CI
Knew	1.67	.10	.21	[−.08;.48]
Knew probably	1.15	.25	.15	[−.17;.52]
Believed	.64	.53	.08	[−.20;.36]
Believed probably	1.70	.09	.05	[−.25;.39]
Should have believed	2.11	.04	.36	[.01;.60]
Probability	6.13	< .001	.77	[.51;1.02]

(1) The action of the one person is exactly as wrong as that of the other person.
(2) Both people are equally blameworthy.
(3) The action of the one person is as permissible or impermissible as that of the other person.
(4) Both people deserve an equal amount of punishment.

Responses were collected on a 7-point Likert scale anchored at 1 with "completely disagree" and at 7 with "completely agree". Following the main task, participants completed the 20-item Moral Foundations Questionnaire (Graham et al., 2009, 2011).

5.2.2. Participants

396 participants were recruited online via Amazon Mechanical Turk. We restricted their IP address location to the USA. Participants who failed an attention check or the comprehension question were removed, leaving a sample of 320 participants (male: 48.4%; mean age: 45.6, range: 20–75).

5.2.3. Results

To examine the influence of moral luck on moral judgments, we analyzed participants' answers by means of a two-way (Moral Luck: lucky vs. unlucky; Judgment Type: wrongness vs. permissibility vs. blameworthiness vs. punishment) repeated-measures ANOVA (Fig. 21). We found that aggregating across the four questions, participants' answers for the unlucky agent (John) differed significantly from those concerning the lucky agent (Frank), $F(1,319) = 78.0077$, $p < .001$, $\eta^2 = .20$. We also found that, aggregating across the two levels of Moral Luck, participants' mean answers to the wrongness, permissibility, blameworthiness, and punishment questions differed significantly
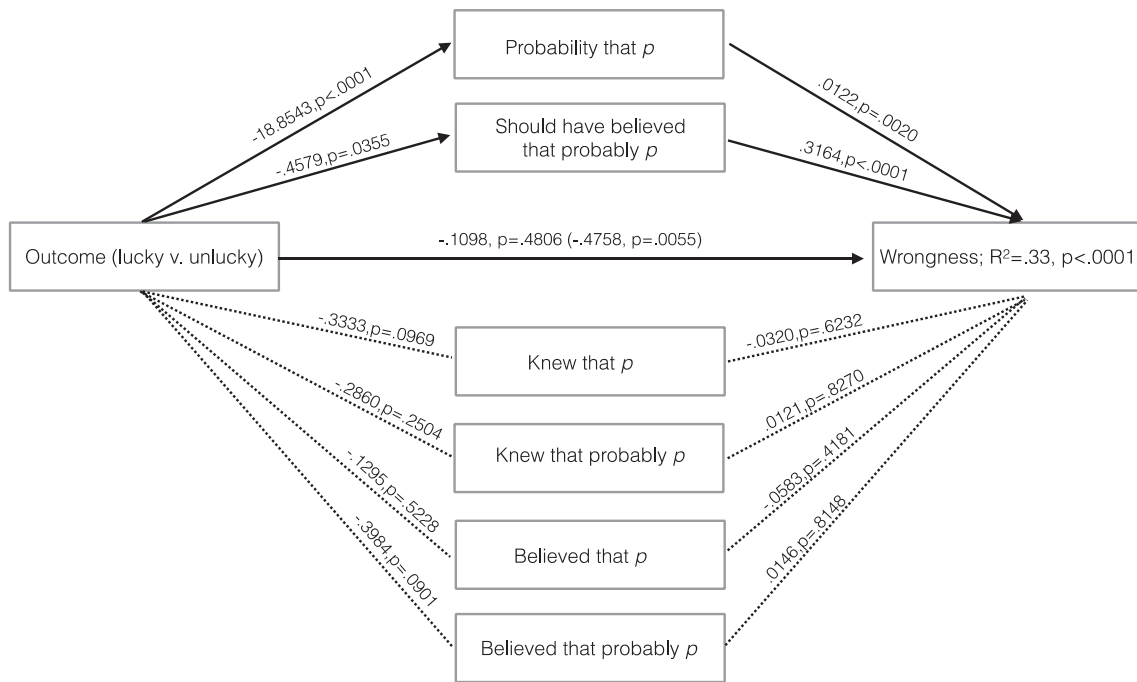
Fig. 18. Mediation of the relationship between outcome (lucky vs. unlucky) and wrongness judgments by probability, negligence, knowledge, and belief ascriptions.

($F(3,957) = 66.14$, $p < .001$, $\eta^2 = .17$). Post hoc tests using the Bonferroni correction revealed that the significant differences were (1) between answers to the punishment question and answers to the three other questions, and (2) between answers to the wrongness question and answers to the three other questions (Fig. 22). We did not find any evidence that participants responded differently to the last pair of questions.

Importantly, the two main effects we observed were qualified by a two-way interaction ($F(3,957) = 60.61$, $p < .001$, $\eta^2 = .16$). To analyze this interaction, we compared the difference between the morally lucky and the morally unlucky conditions for each of the four questions. Probably because of rounding issues, we were unable to compute 95%

confidence intervals for these effect sizes using 2016 ESCI. The results are summarized in Table 5.

To investigate whether the observed aggregate influence of moral luck on wrongness, blame, permissibility, and punishment judgments is widespread among participants or due to a small minority, we compared the proportion of participants who manifested a Difference Intuition with those who did not. Participants who judged the lucky and the unlucky agents identically were counted as *No DI* (Fig. 23). As in Study 1b, the vast majority of participants did not share the Difference Intuition for wrongness (89%), blame (83%), and permissibility (90%); all distributions differed significantly from chance (binomial test, $p < .001$, two-tailed). Consistent with the findings from Study 1b, the
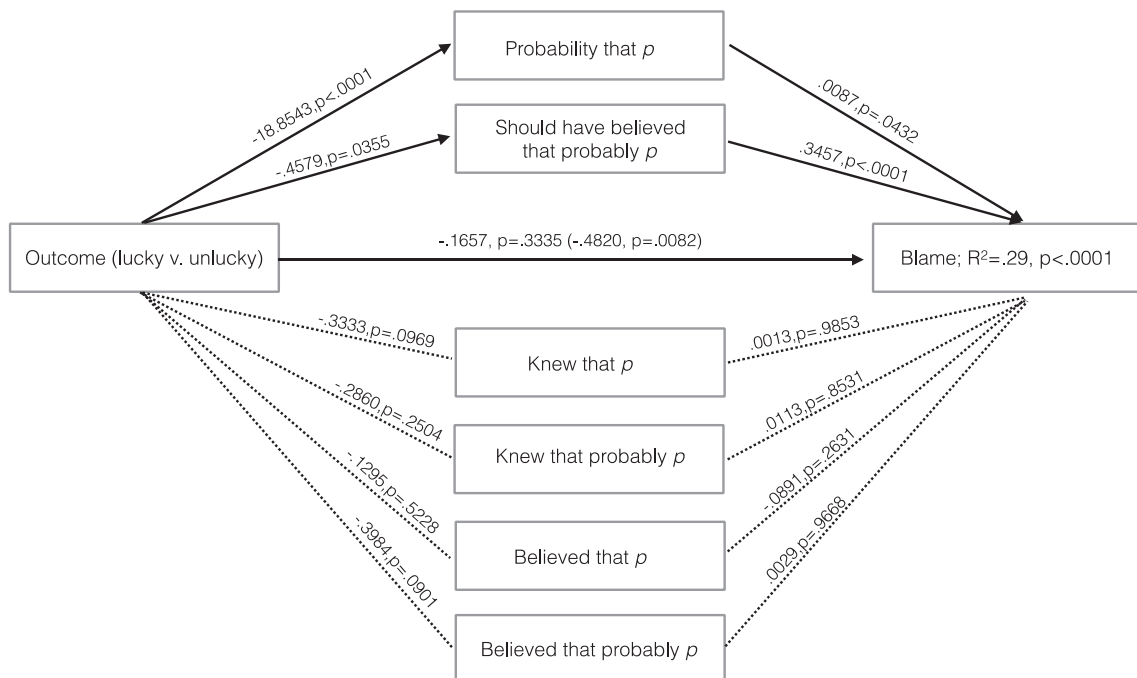

Fig. 19. Mediation of the relationship between outcome (lucky vs. unlucky) and blame judgments by probability, negligence, knowledge, and belief ascriptions.
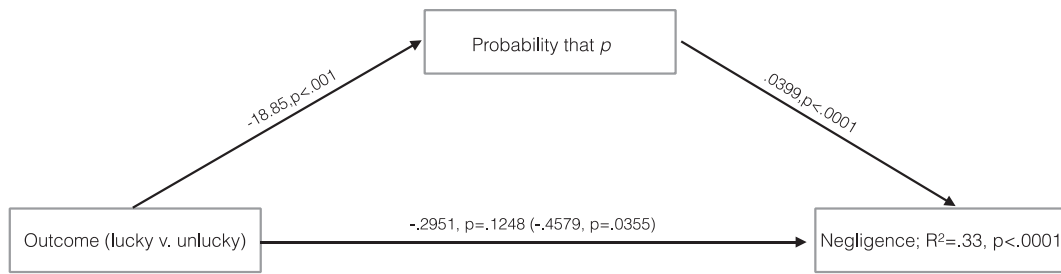
Fig. 20. Mediation of the relationship between outcome (lucky vs. unlucky) and negligence ascriptions ("should have believed there was a high probability of an accident") by probability judgments.

ratio was considerably smaller for punishment (55%), which did not differ significantly from chance (binomial test, $p = .103$, two-tailed), and which differed significantly from wrongness, blame, and permissibility (binomial tests, test proportion = .55, $p < .001$, two-tailed). Blame also differed significantly from wrongness (binomial test, test proportion = .83, $p = .004$, two-tailed) and permissibility (binomial test, test proportion = .83, $p < .001$, two-tailed).

Next, we turned to people's abstract views about moral luck. A repeated measures ANOVA with a Greenhouse-Geisser correction revealed a significant difference across types of moral judgments (F (2.543, 752.80) = 29.58, $p < .001$, $\eta^2 = .091$; Fig. 24). The order of display was not significant ($F(23, 296) = 1.26$, $p = .193$, $\eta^2 = .089$). In Bonferroni corrected comparisons, all pairs differed significantly from one another except wrongness and blame judgments. The means for all four measures differed significantly from the midpoint 4 of the Likert scale, as well as from the endpoint 7 (see Appendix section 5.1).

To investigate the fit of judgments about moral luck in the concrete and abstract conditions, we tested for correlations between, on the one hand, the difference in the judgments about the unlucky and unlucky agent and, on the other, responses in the abstract condition for all four measures. For all but permissibility, abstract rejection of moral luck correlates with a less pronounced difference in the assessment of the unlucky agent in contrast to the lucky one (for N = 320, wrongness: $r = -.38$, $p < .001$; blame: $r = -.28$, $p < .001$; permissibility: $r = -.10$, $p = .090$; punishment: $r = -.54$, $p < .001$, all two-tailed).
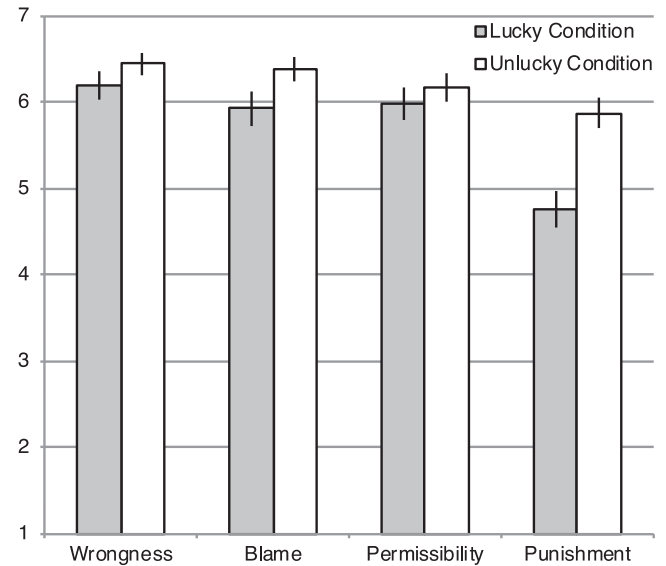

Fig. 22. Mean wrongness, blame, permissibility, and punishment judgments in the within-subjects design for the Toymaker scenario; error bars denote 95% confidence intervals.
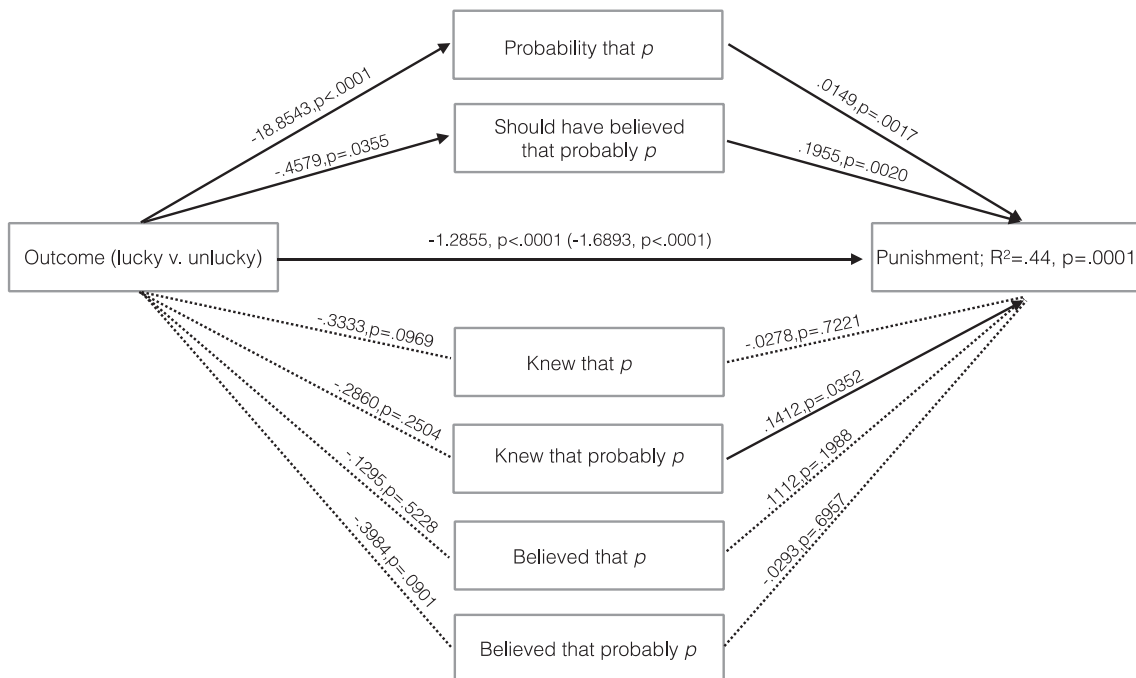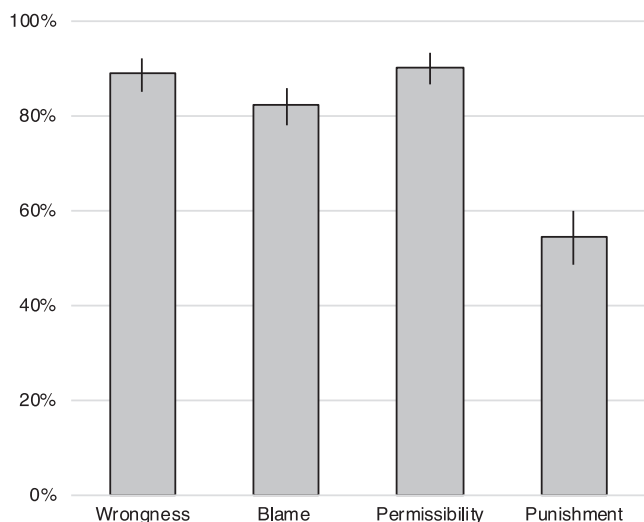

Fig. 21. Mediation of the relationship between outcome (lucky vs. unlucky) and punishment by probability, negligence, knowledge, and belief ascriptions.

Table 5
Effect of outcome on wrongness, blame, permissibility, and punishment judgments.

Variable	t(319)	p	Cohen's d
Wrongness	4.20	< .001	.19
Blame	3.99	< .001	.30
Permissibility	5.86	< .001	.12
Punishment	11.27	< .001	.64

Fig. 23. Proportions of participants who judged the two agents identically (*No DI*) for wrongness, blame, permissibility, and punishment in the Toymaker scenario; errors bars represent 95% confidence intervals; Wilson method, see Brown et al., 2001.
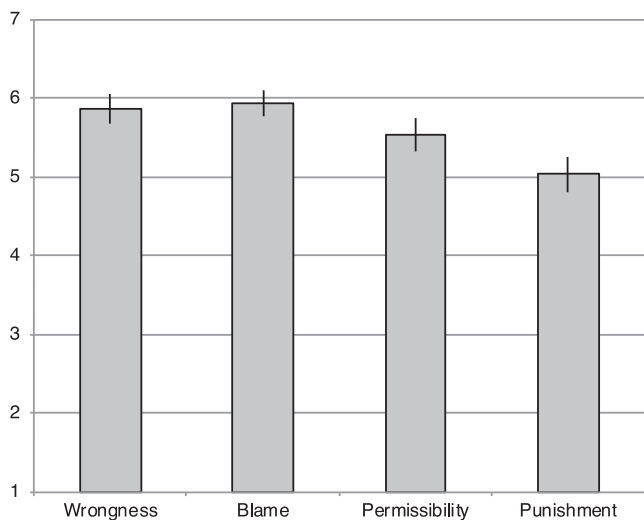
Fig. 24. Abstract denial of moral luck for wrongness, blame, permissibility, and punishment; error bars denote 95% confidence intervals.

Table 6
Proportions of participants who agreed (above midpoint 4) and completely agreed (endpoint 7) with the abstract denial of moral luck for wrongness, blame, permissibility, and punishment.

Measure	Wrongness	Blame	Permissibility	Punishment
Answers > 4	81%	83%	76%	64%
Endpoint 7	54%	51%	54%	36%

The results suggest that in concrete cases, people judge largely in harmony with their abstract views about moral luck.

Following Lench et al. (2015), we aggregated the number of participants who agreed that the two agents should be judged identically if their actions differed only with respect to outcome-related factors beyond their control (answers > 4) and those who completely agreed with this claim (endpoint 7). The results are summarized in Table 6.

To further explore the fit between concrete and abstract judgments about moral luck, we aggregated the number of participants who judged the two agents identically in the two concrete cases and who, when asked explicitly about their abstract principles, agreed that the two agents should be judged the same, thus uniformly disavowing the Difference Intuition. We also aggregated the number of participants who judged the two agents differently in the two concrete cases, and who, when asked explicitly about their abstract principles, disagreed with the claim that the two agents should be judged the same, thus uniformly endorsing the Difference Intuition. The sum of the two proportions constitutes the overall ratio of participants whose judgments about concrete cases fit their abstract principles, i.e. who judged *uniformly* (Fig. 25).

For the vast majority of participants judgments about wrongness, blame, and permissibility were in harmony with their abstract principles, whereas just about half of punishment judgments were uniform. Perhaps even more importantly, for wrongness, blame, and permissibility, over three quarters of the participants uniformly disavow the Difference Intuition.

5.3. Study 4c: Alternative within-subjects design

Given the potentially controversial nature of our findings—an absence of a Difference Intuition regarding wrongness, blame, and permissibility for the vast majority of people—we ran a second replication with a different design. Following Experiment 1 of Spranca, Minsk, and Barron (1991), 140 participants were given a single vignette (a variation of the Toymaker scenario) with the instruction to imagine two distinct endings: one in which some children die of lead poisoning, another in which no harm occurs.[3] Subsequently they were asked to judge wrongness, blame, responsibility, and punishment for the first and the second ending. This design has the advantage that a single agent is used, which makes it perfectly transparent to the reader that the general situation of the protagonist and their mental states are held fixed. As in Study 1b and 4b, participants were also asked about their abstract views regarding moral luck. Methods, materials, and detailed results are reported in the Appendix (Section 6).

The findings were consistent with the results of Studies 1b and 4b. First, at the aggregate level, an effect of outcome on moral judgment was found for all four measures. Second, for the vast majority of people, the outcome of the agent's actions did not influence permissibility, blame, and wrongness judgments when they read about both a morally lucky and unlucky situation or when they report their abstract views on the matter. Third, the influence of moral luck varies across types of moral judgment: We found a large effect of outcome on punishment judgments, but only medium effects of outcome on wrongness, blame, and permissibility judgments. Fourth, wrongness, blame, and permissibility form one cluster of moral judgments, while punishment alone forms a distinct cluster. Finally, abstract views were by and large correlated with judgments about concrete cases.

5.4. Study 4d: Contrastive design

Study 4d replicated Study 2. 386 participants recruited on Amazon Mechanical Turk were presented with both the Good Luck and Bad Luck versions of the Toymaker vignette. Thereafter they had to state to what

---

[3] We are grateful to Jonathan Baron for suggesting this experimental design.
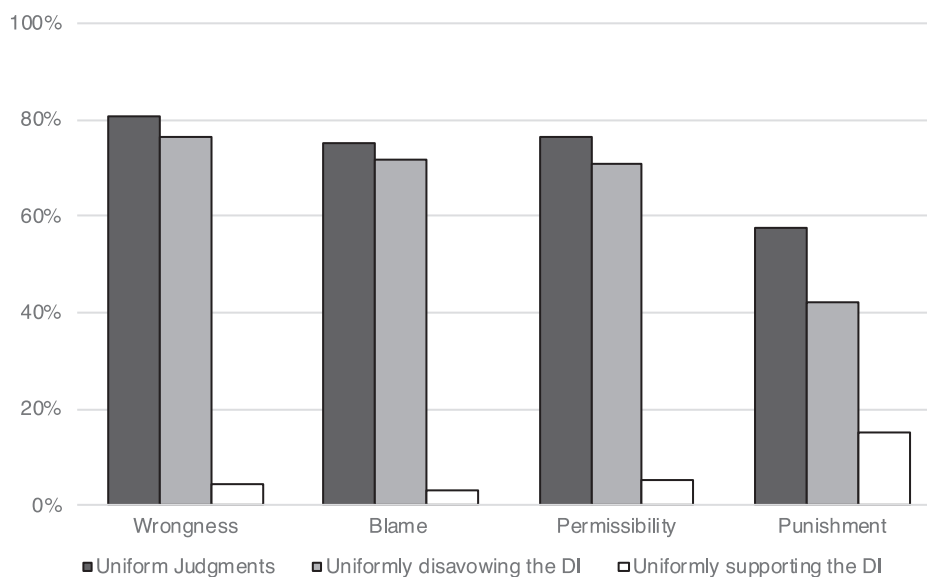
Fig. 25. Proportions of uniform judgments across outcomes for wrongness, blame, permissibility, and punishment.

extent they agreed with comparative claims about Frank and John's behavior, focusing on wrongness, blame, permissibility, and punishment. The statements took the exact same form as in Study 2. For wrongness, for instance, participants had to assess the claim "It was just as wrong for John to use the new paint without testing it for lead as it was for Frank" on a 7-point Likert scale anchored at 1 with "completely agree" and at 7 with "completely disagree." As in Study 2, the four statements were presented in randomized order. The main task was preceded by an attention check and followed by a comprehension question. Methods, materials, and detailed results can be found in the Appendix (Section 7).

All results from Study 2 were replicated: First, for wrongness, blame, and permissibility, over 80% of the participants agreed (Likert-scale < 4) and over 60% completely agreed (endpoint 1) that the two agents should be judged in the same fashion. The ratios for punishment were significantly lower: 62% agreed and 41% completely agreed with an identical assessment of the two agents. Second, for wrongness, blame, and permissibility mean disagreement with moral luck was pronounced both in the case-based task and the abstract question; for punishment mean disagreement was present though again significantly lower. Third, for all four measures, abstract endorsement of moral luck correlates with a more pronounced difference in the assessment of the two agents in the case-based task.

5.5. Discussion

5.5.1. Evidence for the Difference Intuition across experimental designs

The results for the between-subjects Study 4a replicated those of Study 1a perfectly. The effect size for the influence of outcome on punishment judgment was again large ($d = 1.07$), whereas it was about a third as pronounced yet also significant for wrongness and blame judgments ($d = .36$ and $d = .33$). Furthermore, the lower bound of the 95% confidence interval for the effect size for punishment judgments was larger than the higher bounds of the 95% confidence intervals for the effect sizes for the three other judgments. As in Study 1a, outcome had no statistically significant effect on permissibility judgments.

The basic pattern of the within-subjects Study 1b was replicated by Study 4b: The influence of outcome was again large for punishment judgment ($d = .61$) and moderate for wrongness ($d = .30$) and blame ($d = .35$) judgments. In contrast to the other studies, however, we found a small effect for permissibility judgments ($d = .15$). Since the confidence intervals for the effect of outcome on permissibility judgments in our previous studies barely included 0, we suspect that there is

a real, although small effect of outcome on permissibility judgments. Study 4c, which used an alternative design invoking two distinct endings rather than two different vignettes, was also consistent with the findings of Study 1b and 4b.

The results of the contrastive design Study 4d were consistent with those of Study 2. There was solid agreement with the claims that the two agents were equally blameworthy, their actions equally wrong, and equally impermissible, although for all three measures the average differed significantly from the endpoint 1. Most participants "completely agreed" with the claim that the actions of the morally lucky and unlucky agents are equally permissible, wrong, and blameworthy. Punishment judgments differed from the three other types of judgment. Consistent with the findings of Study 2, there was only moderate agreement with the statement that the two agents deserve the same degree of punishment. Only a minority of participants "completely agreed" with the claim that the morally luck and unlucky agents deserve equal punishment.

Taken together, Studies 4b, 4c, and 4d confirm that for the vast majority of people outcome has no influence on permissibility, wrongness, and blame judgment when they are presented with both the Good Luck and Bad Luck scenarios. By contrast, even in such comparative circumstances, outcome *still* influences judgments about punishment. Study 4a confirms that in a non-comparative task, where people judge only one of the two types of agents, outcome influences wrongness, blame, and punishment judgments (the latter more than the former two).

5.5.2. Mediation analyses

The second part of Study 4a, which focused on potential mediators, replicated the findings of Study 3. For all four types of moral judgment, the epistemic mediators proved once more insignificant. As in Study 3, we found strong support in favor of the *Probabilistic Model*: For wrongness, blame, and punishment, probability and negligence ascriptions mediated the impact of outcome on moral judgment, and negligence was near-completely mediated by probability. Importantly, the impact of outcome on wrongness and blame were reduced to insignificance once the mediators had been taken into account, which suggests that for wrongness and blame a Single Probabilistic Pathway Model might be most appropriate. Punishment, by contrast, again best fits a Dual Probabilistic Pathway Model. While the effect of outcome on moral judgment was significantly mediated by probability, a strong direct effect of outcome on moral judgment remained (note again that it might disappear if other potential mediators were tested for).

6. Study 5: Abstract views on moral luck

In Studies 4b, 4c and 4d, we examined not only people's judgments regarding moral luck in concrete cases, but also their abstract views on the subject. Since in all three studies, the questions about participants abstract views about moral luck were preceded by the concrete, case-based judgment tasks, responses to the latter might have influenced responses to the former. We ran another study with 121 participants recruited on Amazon Mechanical Turk, who were presented with the abstract questions (Section 5.2.1), this time not preceded by judgments tasks invoking concrete cases. The results are by and large the same as in Studies 4b, 4c and 4d (see Section 8 of the Appendix for further detail): About three quarters of participants agree that the two agents should be judged identically in terms of wrongness, blame, and permissibility, whereas for punishment it is just over half of the participants.

Together with Studies 4b, 4c and 4d, Study 5 suggests that there is no abstract/concrete difference for moral luck as there is in other areas of moral psychology (Nichols & Knobe, 2008). On average, people agree with the abstract claim that the morally lucky and unlucky individuals should be treated similarly. Punishment differs from the three other moral judgment types: wrongness, blame, and permissibility.

7. General discussion

7.1. Moral luck and moral competence

Our studies suggest that when confronted with both a Good Luck and a Bad Luck scenario, a very large majority of people judge the two agents and their actions identically in terms of wrongness, blame, and permissibility. That is, for most people outcome does not influence these moral judgments in such circumstances. This is true both when people are asked to make a moral judgment about a concrete situation and when they are asked to consider abstract views about moral luck.

Our mediation analyses cast light on the mechanisms behind the influence of outcome on moral judgment when people consider only one situation. According to one view, people ascribe different epistemic states to the two agents: The difference in the moral assessment of the morally lucky and unlucky agents would thus be due to different ascriptions of foresight (knowledge that the bad outcome will occur) or recklessness (awareness of a substantial probability of the bad outcome occurring). Our results suggest that such models are incorrect since people ascribe foresight and recklessness to similar degrees across the lucky and unlucky situations. Rather, our mediation analyses suggest that, in non-comparative situations, outcome affects blame and wrongness judgments because people view the unlucky agent as more negligent than the lucky agent, and they view the former as more negligent because they consider the probability of the bad outcome's occurrence in the unlucky situation as higher than in the lucky situation. Once these mediators are taken into consideration, the effect of outcome on blame and wrongness judgments turns insignificant.

The mediation analysis results of Study 4a suggest a Probabilistic Single Pathway Model for blame and wrongness (though for wrongness, the results of Study 3 are more in tune with a Probabilistic Dual Pathway Model). Both Studies 3 and 4a suggest a Probabilistic Dual Pathway Model for punishment. These findings can be interpreted in two distinct ways: On a *rationalist approach*, outcome information can be taken to constitute evidence regarding an event's likelihood: It is only reasonable, one might argue, for a judging subject to assess the probability of a contextually salient event higher in a situation in which it has just occurred than in a situation in which it just failed to occur. This justified difference in perceived probability gives rise to an equally innocuous perceived difference in negligence ascriptions across lucky and unlucky agents. But if the unlucky agent was indeed more negligent, it is appropriate to consider her more blameworthy than the lucky, and less negligent, agent.

On a *bias approach*, by contrast, the differing perceptions of accident probability might be seen as a performance error. A performance error is a judgment that is not reflective of people's domain-specific competence, but rather results from the characteristics of the processes involved in making the judgment. For instance, the judgment that a center-embedded sentence like "The rat the cat the dog bit chased escaped" is not acceptable is a performance error: it does not reflect people's grammatical competence, but rather results from the processing limitations of working memory. Other types of sentences such as garden-path sentences elicit performance errors, too. A garden-path sentence is grammatical but sounds unacceptable because its beginning is similar to a misleading salient syntactic construction and thus elicits an interpretation that ends up being incorrect. "The horse raced past the barn fell" and "The old man the boat" are classic examples of garden-path sentences. On reflection, however, competent speakers can see that such sentences are acceptable because they overcome the similarity between their beginning and the misleading syntactic constructions. Certain types of moral judgments in non-comparative contexts, we propose, give rise to a similar phenomenon. On reflection, we hypothesize, the folk conceptions of probability and negligence in moral contexts pattern with their legal analogues (the basic legal framework regarding *mens rea* and culpability presumably derives from folk morality and psychology, see Moore (2011)). According to the folk (and legal) view, what matters for moral evaluation is the *ex ante* probability of an accident (the probability at the context of action) and a concept of negligence defined in terms of the latter, not *ex post* probability (the probability at the context of assessment). When attempting to assess the *ex ante* probability of an accident in non comparative situations, however, people fall prey to the hindsight bias (a performance error), as their judgments are inappropriately influenced by outcome information. They exaggerate the accident probability in the unlucky case, deem the unlucky agent as more negligent than the lucky one, and consequently judge her more harshly. In short, the influence of outcome on blame and wrongness judgments results from a domain-general, irrational bias regarding probability assessments, and is not expressive of people's reflective moral views. The distinct assessment of the morally lucky and unlucky agents in non-comparative contexts, this is to say, is not a manifestation of moral competence; it is a performance error.

Even if the effect of outcome on blame and wrongness judgments in non-comparative contexts is a performance error, it is a phenomenon of great significance for everyday life. Moral and legal judgments typically take place in such non-comparative contexts, and one would thus expect that outcome frequently influences judgment in everyday circumstances (cf. also Kneer & Bourgeois-Gironde, 2017 for the effect of outcome on *mens rea* attribution amongst professional judges). The phenomenon should thus be of concern to legal professionals, but also to all of us when we engage in moral judgment.

7.2. Punishment vs. blame and wrongness

Punishment judgments systematically diverged from wrongness and blame judgments in all our studies: (i) Outcome has a systematically larger effect on aggregate punishment judgments than on aggregate wrongness and blame judgments in both the between-subjects and within-subjects designs. (ii) In contrast to wrongness and blame judgments, a sizeable proportion of people judge that the morally lucky and unlucky agents should be punished differently in comparative contexts and when reporting their abstract principles. (iii) The mechanisms behind punishment judgments differ from the mechanisms behind wrongness and blame judgments, as revealed by the mediation analyses. The influence of outcome on wrongness and blame judgments is principally due to the different ascriptions of negligence, which in turn are driven by distinct perceptions of outcome probability across conditions (Study 4a, but see Study 3). By contrast, the influence of outcome on punishment judgments is only partly due to the varying ascriptions of negligence. With the mediators tested, outcome continues

to have a direct effect on punishment judgments even after negligence and probability ascriptions have been taken into account.

The systematic divergence of punishment judgments from wrongness and blame judgments suggests a Dual Process Model of moral judgment. Blame and wrongness judgments result from a first process, while punishment judgment results from a second process. The key distinction between these two processes regards a partial direct effect of outcome on moral judgment. Whether an outcome is bad or good was shown to have a considerable direct effect on punishment judgments (Studies 3 and 4a). In Study 4a, however, no direct effect of outcome on wrongness and blame judgments could be detected; the entire effect of outcome on wrongness and blame was mediated by negligence (though there was some direct effect on wrongness in Study 3).

Our findings are partly consistent with current influential Dual Process Models of moral judgment (Cushman, 2008; see Fig. 4 above), according to which wrongness judgments are primarily influenced by mental state ascriptions (belief and desire), while punishment judgments are influenced by mental state ascriptions *and* the agent's causal contribution to the outcome. The respective effect sizes of outcome for punishment and wrongness judgments are consistent with this model: Since the mental states of the agents were held fixed across the lucky and unlucky conditions and perceived to be the same for both agents (Studies 3 and 4a), their influence on moral judgment in the tested cases must be limited. Outcome, by contrast, varies across conditions. Consistent with Cushman's theory, outcome had a more pronounced effect on punishment judgments (deemed sensitive to both mental *and* causal factors) than on wrongness (deemed sensitive principally to mental states). In addition, while in Study 4a the impact of outcome on wrongness was mediated in its entirety by probability and negligence ascriptions (in conformity with the Probabilistic Single Pathway Model), we detected a considerable direct effect of outcome on punishment judgments (which thus fits the Probabilistic Dual Pathway Model).

In one respect, however, our findings are inconsistent with current Dual Process Models of moral judgments. These models predict that blame judgments pattern with punishment judgments: Blame judgments are supposed to be influenced by outcome the way punishment judgments are. The findings reported in this article as well as further findings on the action/omission effect (Kneer et al., in preparation a) suggest that this need not be the case; rather, blame often patterns with wrongness judgments. Cushman (p.c.) draws our attention to the fact that in his experiments, the questions were stated using the expression "blame" whereas our questions were phrased in terms of "blameworthy." As of yet unpublished findings of different researchers seem to converge on the fact that judgments invoking "blame" are more outcome-sensitive than judgments cast in terms of "blameworthy."

7.3. Philosophical implications

The Control Principle and the Difference Intuition are inconsistent. A puzzle arises if, as philosophers have commonly assumed, both the Control Principle and the Difference Intuition are widely held. The philosophical literature on the topic is vast since the Puzzle of Moral Luck is deemed one of the most foundational challenges to systematic ethics. As Bernard Williams observed, widespread support for the Difference Intuition "cannot leave the concept of morality where it was" (1993, 54).

The Difference Intuition, as it features in the philosophical literature, requires the comparative assessment of two agents, who bring about distinct outcomes in otherwise identical situations. The between-subjects experiments frequently found in the empirical literature on the topic are thus not, strictly speaking, pertinent to the debate regarding the Puzzle of Resultant Moral Luck. We have used a variety of designs (within-subjects, contrastive, and Baron's design with alternative endings in Study 4c) that are actually suited to test the Difference Intuition. In contrast to what philosophers standardly assume, very few people share the Difference Intuition. In the within-subjects studies for the Bathroom and the Toymaker scenarios, the vast majority of participants

judged the two agents identically for wrongness (89% in the Bathroom scenario/89% in the Toymaker scenario), blame (87%/83%), and permissibility (96%/90%). The same pattern of results was found in the contrastive design experiments and when participants were explicitly consulted on their abstract views about moral luck for all three measures. It is unclear whether the few participants apparently in agreement with the Difference Intuition genuinely endorse it or instead constitute statistical noise. In the latter case, there simply is no Puzzle of Resultant Moral Luck; in the former, its significance would be drastically reduced. Moral luck would no longer constitute a fundamental challenge to the very possibility of systematic ethics, but merely an oddity, or perhaps a bias, of the moral judgment of a small minority.

Whereas no Difference Intuition could be detected for wrongness, blame, and permissibility judgments in comparative contexts, a substantive minority of people hold that the unlucky agent deserves a harsher punishment. One might thus conclude that the Puzzle of Moral Luck exists after all, though its scope is restricted to the punishment judgments of a substantive minority. However, the problem of moral luck is often cast in terms of blame, that is, a property the ascription of which should exclusively depend on moral responsibility, rather than 'blame-related' notions such as deserved punishment, which may also depend on considerations regarding, for instance, deterrence, prevention, or what is socially desirable (Enoch & Marmor, 2007; but see Kumar, 2018 for a punishment-centered approach to the Puzzle of Moral Luck). Thus, the fact that unlucky reckless drivers are punished more harshly than lucky ones in most Western legal systems, and that a substantive minority deems this appropriate, need not raise a puzzle.

7.4. Limitations

Our studies suffer from two main limitations. First, the nature of permissibility judgments is not entirely clear. The effect of outcome on permissibility judgments was only significant in some studies although the confidence intervals of the effect sizes barely included zero when this effect failed to reach significance. We lean towards the view that outcome does influence permissibility judgments in non-comparative contexts, although its effect size is smaller than the effect size for blame and wrongness judgments. If this were correct, a question arises as to whether permissibility judgments result from the same process as wrongness and blame judgments or from a distinct process. Future research should investigate this matter.

Second, the studies reported in this article as well as past studies have principally investigated the influence of moral luck on the moral judgments of Westerners (an important exception is Barrett et al., 2016, who explored the impact of intent and accidental factors on moral judgment in small-scale societies). In light of the variation of psychological findings across cultures in general (Fessler & Machery, 2012; Henrich, Heine, & Norenzayan, 2010) and moral psychology in particular (Abarbanell & Hauser, 2010; Hannikainen et al., in preparation), the universality of our findings should be investigated cross-culturally and cross-linguistically (Kneer et al., in preparation b).

8. Conclusion

Philosophers have argued that the Puzzle of Resultant Moral Luck challenges the very possibility of systematic ethics. Our findings suggest there is no such challenge since most people simply do not share the Difference Intuition. A very large majority of people do not make different wrongness, blame, and permissibility judgments when they are reflectively considering the Good Luck and the Bad Luck scenarios side by side, and this pattern of concrete judgments is consistent with their abstract principles. It is only when people consider a single situation in isolation that their judgment may be influenced by factors that the agent does not control, such as the action's outcome. While such circumstances are common in everyday life, our mediation analyses suggest that the resulting judgments are performance errors: They result from the hindsight bias,

that is, from a retroactive overestimation of the probability that an outcome would occur if it does indeed occur, which leads people to view the morally unlucky agent as more negligent than the lucky one.

Punishment judgments differ from the other types of moral judgments including blame, an unexpected finding in light of standard dual-process models, according to which punishment judgments pattern with blame judgments. We found that both in comparative contexts and when consulted about their abstract principles, people consider an unlucky agent more deserving of punishment than a lucky agent. While this finding might call for a refinement of current Dual Process Models of moral judgment in moral psychology, it is not philosophically puzzling since on many normative views, deserved punishment does depend on the consequences of actions.

Acknowledgments

Appendix A. Supplementary materials

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cognition.2018.09.003.

References

Abarbanell, L., & Hauser, M. D. (2010). Mayan morality: An exploration of permissible harms. *Cognition, 115*(2), 207–224.

Alicke, M., & Davis, T. (1989). The role of a posteriori victim information in judgments of blame and sanction. *Journal of Experimental Social Psychology, 25*(4), 362–377.

Alicke, M., Davis, T. L., & Pezzo, M. V. (1994). A posteriori adjustment of a priori decision criteria. *Social Cognition, 12*(4), 281–308.

American Law Institute (1985). *Model penal code: Official draft and explanatory notes: Complete text of model penal code as adopted at the 1962 Annual Meeting of The American Law Institute at Washington, D.C., May 24, 1962.* Philadelphia: American Law Institute.

Baron, J. (2000). *Thinking and deciding.* Cambridge University Press.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*(4), 569.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*(2), 74–85.

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., & Scelza, B. A. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences, 113*(17), 4688–4693.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 101–117.

Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes, 48*(1), 147–168.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*(2), 353–380.

Cushman, F. (2013). Action, outcome, and value a dual-system framework for morality. *Personality and Social Psychology Review, 17*(3), 273–292.

Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *PloS one, 4*(8), e6699.

Enoch, D., & Marmor, A. (2007). The case against moral luck. *Law and Philosophy, 26*(4), 405–436.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of personality and social psychology, 71*(2), 390.

ESCI Exploratory Software for Confidence Intervals (2016), accompanies. Cumming, G. & Calin-Jagerman, R. (2017), Introduction to the New Statistics: Routledge.

Everett, J. A. (2013). The 12 item social and economic conservatism scale (SECS). *PloS one, 8*(12), e82131.

Fessler, D. M., & Machery, E. (2012). Culture and cognition. In E. Margolis, R. Samuels, & S. P. Stich (Eds.). *The Oxford handbook of philosophy of cognitive science.* Oxford University Press.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 288.

Gino, F., Moore, D. A., & Bazerman, M. H. (2009). No harm, no foul: The outcome bias in ethical judgments. *Harvard Business School NOM Working Paper,* 08–080.

Gino, F., Shu, L. L., & Bazerman, M. H. (2010). Nameless + harmless = blameless: When seemingly irrelevant factors influence judgment of (un)ethical behavior. *Organizational Behavior and Human Decision Processes, 111*(2), 93–101.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology, 96*(5), 1029.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366.

Hannikainen, I., Machery, E., Stich, S. P., Rose, D., Alai, M., Angelucci, A., … Zhu, J. (In preparation). Sourcehood, alternate possibilities, and the moderating role of reflection.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*(3), 311.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

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

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.

Kneer, M., & Bourgeois-Gironde, S. (2017). Mens rea ascription, expertise and outcome effects: Professional judges surveyed. *Cognition, 169,* 139–146.

Kneer, M., Olivola, C. Y., Kim, H., Machery, E., Stich, S. P., Rose, D. … Zhu, J. (In preparation a). The action/omission distinction across cultures.

Kneer, M., Machery, E., Stich, S. P., Rose, D., Alai, M., Angelucci, A. … Zhu, J. (In preparation b). Moral luck across cultures.

Kumar, V. (2018). Empirical vindication of moral luck. *Nous*.

Lench, H. C., Domsky, D., Smallman, R., & Darbor, K. E. (2015). Beliefs in moral luck: When and why blame hinges on luck. *British Journal of Psychology, 106*(2), 272–287.

Levy, N. (2005). The good, the bad, and the blameworthy. *Journal of Ethics and Social Philosophy, 1*(2), 1–16.

Lowe, C. A., & Medway, F. J. (1976). Effects of valence, severity, arid relevance on responsibility and dispositional attribution. *Journal of Personality, 44*(3), 518–538.

Martin, J. W., & Cushman, F. (2015). The adaptive logic of moral luck. In J. Sytsma, & W. Buckwalter (Eds.). *The Blackwell companion to experimental philosophy.* Wiley-Blackwell.

McKillip, J., & Posavac, E. J. (1975). Judgments of responsibility for an accident. *Journal of Personality, 43*(2), 248–265.

Moore, M. S. (2011). Intention as a marker of moral culpability and legal punishability. In A. Duff, & S. P. Green (Eds.). *Philosophical foundations of criminal law.* Oxford: Oxford University Press.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105.

Nagel, T. (1979). *Mortal Questions (Vol. 89, Vol. 3).* Cambridge University Press.

Nichols, S., & Knobe, J. (2008). Moral responsibility and determinism. In J. Knobe and S. Nichols (Eds.), Experimental Philosophy (pp. 105-128).

Nichols, S., Timmons, M., & Lopez, T. (2014). Using experiments in ethics–ethical conservatism and the psychology of moral luck. In M. Christen, C. vanSchaik, J. Fischer, M. Huppenbauer, & C. Tanner (Eds.). *Empirically informed ethics: Morality between facts and norms* (pp. 159–176). Springer International Publishing.

Nichols, S. (2009). Ethics and the psychology of moral luck. In Presented at the Pacific American Psychological Association, Vancouver, BC.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*(6), 972.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891.

Richards, N. (1986). Luck and desert. *Mind, 65,* 198–209.

Robbennolt, J. K. (2000). Outcome severity and judgments of "Responsibility": A meta-analytic review 1. *Journal of Applied Social Psychology, 30*(12), 2575–2609.

Rosebury, B. (1995). Moral responsibility and moral luck. *Philosophical Review, 104,* 499–524.

Royzman, E., Cassidy, K. W., & Baron, J. (2003). "I know, you know": Epistemic egocentrism in children and adults. *Review of General Psychology, 7*(1), 38.

Royzman, E., & Kumar, R. (2004). Is consequential luck morally inconsequential? empirical psychology and the reassessment of moral luck. *Ratio, 17*(3), 329–344.

Rubin, Z., & Peplau, L. A. (1975). Who believes in a just world? *Journal of Social Issues, 31*(3), 65–89.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language, 27*(2), 135–153.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27,* 76–105.

Walster, E. (1966). Assignment of responsibility for an accident. *Journal of Personality and Social Psychology, 3*(1), 73.

Thomson, J. J. (1993). Morality and bad luck. In D. Statman (Ed.). *Moral luck* (pp. 195–215). SUNY Press.

Williams, B. (1981). *Moral luck: Philosophical papers 1973–1980.* Cambridge: Cambridge University Press.

Williams, B. (1993). Moral luck. In D. Statman (Ed.). *Moral Luck.* SUNY Press.

Wuensch, K. (2012). Using SPSS to obtain a confidence interval for Cohen's d. http://core.ecu.edu/psyc/wuenschk/SPSS/CI-d-SPSS.pdf.

Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Review of Philosophy and Psychology, 1*(3), 333–349.

Zimmerman, M. J. (2006). Moral Luck: A Partial Map. *Canadian Journal of Philosophy, 36*(4), 585–608.