RESEARCH ARTICLE

EASP WILEY

The effect of outcome severity on moral judgement and interpersonal goals of perpetrators, victims, and bystanders

Lisa Katharina Frisch[1] | Markus Kneer[1] | Joachim Israel Krueger[2] | Johannes Ullrich[1]

[1] Department of Psychology, University of Zurich, Zurich, Switzerland

[2] Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, Rhode Island, USA

Correspondence
Lisa Katharina Frisch, Department of Philosophy, University of Zurich, Binzmuehlestr. 14/15, 8050 Zurich, Switzerland.
Email: lisakatharina.frisch@uzh.ch

Abstract

When two actors have the same mental state but one happens to harm another person (unlucky actor) and the other one does not (lucky actor), the latter elicits a milder moral judgement. To understand how this outcome effect would affect post-harm interactions between victims and perpetrators, we examined how the social role from which transgressions are perceived moderates the outcome effect, and how outcome effects on moral judgements transfer to agentic and communal interpersonal goals. Three vignette experiments ($N = 950$) revealed similar outcome effects on moral judgement across social roles. In contrast, outcome effects on agentic and communal goals varied by social role: victims exhibited the strongest outcome effects and perpetrators the weakest, with bystanders falling in between. Moral judgement did not mediate the effects of outcome severity on interpersonal goals. We discuss the possibility that outcome severity raises normative expectations regarding post-harm interactions that are unrelated to moral considerations.

KEYWORDS
interpersonal goals, moral judgement, moral luck, needs-based model, outcome effect, social role

1 | INTRODUCTION

You are late for a job interview and decide to take the subway even though you do not have a face mask.

Unbeknownst to you, the father of your friend is on that train and inhales your aerosols. The next day you discover that you have Covid-19 and infected your friend's father, who is a patient at risk.

How much blame do you deserve?

What would your friend think?

Would your moral judgement predict how you guys want to interact with each other?

Would your answers to the above questions change if you had obtained a negative Covid-19 test result and thus not infected your friend's father?

Moral evaluations of such scenarios, known as moral luck, have been shown to be sensitive to outcome severity: holding mental states constant, people judge actions and agents more harshly when the consequences are more severe, thereby exhibiting outcome effects (Cushman, 2008; Cushman, Dreber, Wang, & Costa, 2009; Gino, Moore, & Bazerman, 2009; Gino, Shu, & Bazerman, 2010; Kneer & Machery, 2019; Lench, Domsky, Smallman, & Darbor, 2015; Young, Nichols, & Saxe, 2010; for a review, see Martin & Cushman, 2015). Previous research on outcome effects has mostly been limited to scenarios in which respondents take the perspective of an uninvolved bystander (but see Feltz, Harris, & Perez, 2012, for the effect of outcome on intention; Kneer, 2018, for outcome effects on epistemic states; and Gold, Pulford, & Colman, 2015; Schaich Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; Tassy, Oullier, Mancini, & Wicker, 2013,

for research on actor vs. observer differences in the trolley dilemma). However, it is unclear how people actually involved in the transgression (i.e., perpetrators and victims) respond to scenarios of moral luck.

Considering that moral judgement seems likely to influence how parties will interact with each other in the transgression's aftermath, it would make for a smoother resolution of conflicts if victims and perpetrators agreed on the amount of blame an action deserves, and if bystanders were to further validate their judgements. After all, consensus about the inequity the transgression caused is an important step towards reconciling the parties involved: from a social exchange perspective, reconciliation is achieved when victims and perpetrators correct the imbalance of material or symbolic resources that resulted from a transgression (e.g., Adams, 2016). A better understanding of how role-specific moral judgement might affect interpersonal interactions will inform restorative justice and reconciliation procedures as well as disputes and conflicts people face in their everyday lives.

In the present research, we use the outcome effect as a vehicle to understand post-transgression social interaction. Specifically, we (a) examine whether perpetrators, victims, and bystanders agree in their moral judgements following neutral and bad outcomes, and (b) explore how these outcome effects translate into the assessor's interpersonal goals of seeking control, dominance, and power (agentic goals) as well as caring for, cooperating with, and being connected with others (communal goals). We present the findings of three experiments (one main study reported in detail and two additional studies available online) testing the proposition that victims are the most sensitive to outcome information and perpetrators the least.

Our data clearly show, quite contrary to predictions, that social interaction is uncoupled from moral judgement. The negligent perpetrators agree with victims in their moral judgement, accepting more blame to the extent that the outcome is more severe. However, perpetrators' communal behaviour reflects their overall negligence rather than the differential severity of the outcome. In the following, we first review the philosophical considerations that the literature offers to understand outcome effects among neutral bystanders. After reviewing the empirical research on moral luck, we derive predictions for the moderating effect of social role on moral judgement and goals for agentic and communal interactions. In the discussion we develop the idea that increased communal goals of perpetrators may be due to normative expectations associated with agent regret.

1.1 | Should outcome severity affect moral judgement?

How *should* outcome severity affect moral judgement? Two positions mark the endpoints of a spectrum in normative ethics: the first position is *Kantianism*, according to which "nothing in the world [...] could be called good without qualification except a good will" (Kant, 1785, p. 1). According to this view, an action with a beneficial outcome can be bad if undertaken for bad reasons. The second position is *Consequentialism*, which regards an action as good to the extent that its consequences

increase well-being. Strict consequentialists judge an action as morally better inasmuch as it yields a better outcome without regard to the intention. Whereas for Kantians an action's moral quality depends exclusively on the agent's intentions (or, broadly conceived, their mental states), consequentialists look primarily to actual or expected outcomes.

Outcome effects are thus consistent with consequentialist ethics, but they violate a fundamental maxim of Kantian ethics, the Control Principle. This principle states that agents are responsible only for those features of an action they can control: taking the subway without a face mask deserves a fixed amount of blame, whether the agent is unlucky (infecting others) or lucky (not infecting others). Philosophers assume widespread commitment to the Control Principle. And yet, they hypothesise, we are prone to judge unlucky agents more harshly than lucky agents, even when directly contrasting the two situations. If this were indeed the case, we would face a serious paradox, known as the *Puzzle of Moral Luck* (Nagel, 1979; Williams, 1981; for recent overviews, see Hartman, 2017; Nelkin, 2019). Empirical research has shown that in contrastive designs, where people are presented with both a lucky and an unlucky case side by side, they judge the actions and agents similarly (Kneer & Machery, 2019; Lench et al., 2015; Schwitzgebel & Cushman, 2012). The case for a deep puzzle of moral luck might thus have been overstated in the philosophical literature. However, in other designs the outcome effect has proven robust (Cushman, 2008; Gino et al., 2010; Kneer & Machery, 2019). This suggests that people, although they endorse the Control Principle in the abstract, are prone to violate it when presented with concrete cases where luck is at play.

1.2 | Outcome effects across social roles

Research on outcome effects has largely ignored potential differences across the parties involved in a transgression: perpetrators,[1] victims, and bystanders. How will social role impact the outcome effect? Schein and Gray (2018) assert that harm is a matter of perception, and perceived harm can differ from the vantage points of perpetrators, victims, and bystanders. From the actor perspective, people have direct access to their mental states and motives. From the observer perspective, by contrast, they must infer others' intentions, beliefs, and other mental states from behavioural cues. The epistemic authority of the first-person perspective can be deceptive as misperceptions of one's own intentions are common (cf. Epley & Dunning, 2000; regarding emotions, cf. Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; concerning prescient thoughts, cf. Pronin, Wegner, McCarthy, & Rodriguez, 2006). Importantly, however, the asymmetry in epistemic access creates an

---

[1] Although philosophers or legal experts are likely to disagree as to whether the term "perpetrator" (rather than the more neutral term "transgressor") is appropriate for an actor who is putting others at risk to be harmed, previous research on the social role of perpetrators may inform our understanding of the outcome effect. Laypeople judging such incidents may or may not make the same distinctions as philosophers or legal experts. Thus, we will refer to both lucky and unlucky actors as perpetrators in order to be able to merge the different literatures.

introspection illusion (Pronin, 2009).[2] This illusion has two sides: in self-assessment, people tend to overweight introspectively available information (i.e., mental state information and personal feelings); in other-assessment, they overweight extrospectively available evidence (i.e., behavioural and outcome information; see also Malle, 2005; Malle & Pearce, 2001). This perspective-driven weighting differential persists even when actors and observers have access to the same overall set of information (Pronin & Kugler, 2007).

The introspection illusion suggests the following hypotheses: victims and bystanders, who evaluate the perpetrator from an outside perspective, will focus on the outcome, and they will neglect the perpetrator's mental states. For perpetrators, the distribution of relative weight will be reversed. Therefore, we predict a stronger outcome effect for victims and bystanders than for perpetrators. For victims, research has already shown that they tend to exaggerate the negative consequences of unfair treatment and view them as a persistent pattern of misbehaviour on the perpetrator's behalf (Baumeister, 1999; Baumeister, Stillwell, & Wotman, 1990). Yet, we do not expect the bystanders' judgements to be identical to the victims' judgements. Since victims are involved in the focal action (Adams & Inesi, 2016; Jones & Nisbett, 1972) and directly affected by the outcome (McCullough, Fincham, & Tsang, 2003), their tendency to focus on the severity of outcome should be stronger than that of bystanders. In short, we predict the rank order of outcome effects to be this: victims > bystanders > perpetrators.

1.3 | Agentic and communal interpersonal goals

Responses to moral transgressions rarely end with judgements of blameworthiness. The parties involved often continue to interact with each other. Human interaction can be parsimoniously described in terms of the broad dimensions of agency and communion (Abele & Wojciszke, 2007; Bakan, 1966; Fiske, Cuddy, Glick, & Xu, 2002; Locke, 2015; Reeder & Brewer, 1979; Wiggins, 1979, 1991). Agentic goals include seeking control, dominance, and power, whereas communal goals include caring for, cooperating with, and being connected with others.

The literature on outcome effects does not offer much guidance on the goals that people in different social roles are likely to pursue. We draw on the needs-based model (Shnabel & Nadler, 2008, 2015) to make predictions as to how interpersonal goals emerge after moral transgressions. The model assumes that transgressions threaten victims' and perpetrators' identities in different ways and activate different needs. Victims lack the power to influence their outcomes, whereas perpetrators lack moral integrity. In the context of reconciliation, victims are, therefore, primarily concerned with restoring their identity as powerful actors, whereas perpetrators are primarily concerned with restoring their identity as moral actors. It follows that victims and per-

petrators employ different strategies to restore their threatened identities: victims pursue agentic goals. Perpetrators, by contrast, pursue communal goals.

An important moderator of the effects of social role on agentic and communal goals is the extent to which the conflict parties view the action as immoral (Siem, Von Oettingen, Mummendey, & Nadler, 2013). For example, if perpetrators reject responsibility, the model does not predict the pursuit of communal goals. In fact, in many protracted conflicts there is competition for the victim role (e.g., Noor, Shnabel, Halabi, & Nadler, 2012) and where there is ambiguity with regard to the roles of victim and perpetrator, the agentic goals associated with the victim role are more pronounced (SimanTov-Nachlieli, Shnabel, Aydin, & Ullrich, 2018). Thus, the assumption bridging the needs-based model with the moral luck literature is that the harshness of moral judgement, capturing the perpetrator's moral debt to the victim, should determine whether and to what extent the perpetrator has communal goals vis-à-vis the victim and the victim has agentic goals vis-à-vis the perpetrator. If so, outcome severity, which we assume to affect moral judgement, should also affect the interpersonal goals of perpetrators and victims. Victims' interpersonal goals should also vary more strongly between situations with bad and neutral outcomes than perpetrators' interpersonal goals.

What kind of goals might bystanders have when interacting with victims and perpetrators? We propose that bystanders approach *perpetrators* with agency (signalling their behaviour as deviant) and *victims* with communion (manifesting solidarity). Thus, bystanders will presumably approach victims with perpetrator-type interpersonal goals, whereas they would approach perpetrators with victim-type interpersonal goals.

Although not specified in the needs-based model, there is evidence that an increase in agency is accompanied by a decrease in communion among victims and vice versa among perpetrators. For example, when respondents imagined interactions with members of an unjustifiably higher social class as opposed to interactions within their own class, they wanted to act more agentically *and* less communally. Conversely, when they imagined interactions with members of an unjustifiably lower social class, they wanted to act more communally *and* less agentically (Aydin, Ullrich, Siem, Locke, & Shnabel, 2019). We might expect such hydraulic patterns among perpetrators and victims in the context of transgression. Compared with bystanders, victims might simultaneously pursue more agency (i.e., their role-specific goal) and less communion, whereas perpetrators might simultaneously pursue more communion (i.e., their role-specific goal) and less agency.

2 | THE PRESENT RESEARCH

We conducted three vignette experiments online to examine how social role moderates the outcome effect on moral judgement and interpersonal goals. Here we focus on the main experiment with the highest statistical power because the other two studies were highly similar. Details on the additional studies can be found in an online Appendix, which is available together with data, analyses,

---

[2] For simplicity of exposition, we keep the term introspection *illusion* from the literature. We want, however, to remain non-committal concerning the question whether the underlying phenomena must be considered as biases or not.

preregistration (including hypotheses, study design, sample size, exclusion criteria, and planned analyses), and materials (including all measures, manipulations and exclusions) at https://osf.io/t7e2k/. We report the preregistered hypotheses and analyses in the main text (hypotheses concerning the interaction effect of outcome and social role) or in the Appendix (hypotheses concerning the main effect of social role).

Drawing on the introspection illusion (Pronin, 2009), we hypothesised that victims exhibit stronger outcome effects than bystanders, who in turn exhibit stronger outcome effects than perpetrators. Specifically, after bad rather than neutral outcomes, victims would indicate harsher moral judgements, stronger agentic goals, and weaker communal goals than bystanders interacting with the perpetrator. Compared with perpetrators, bystanders interacting with the victim would indicate harsher moral judgements, weaker agentic goals, and stronger communal goals when the outcome is bad rather than neutral. Note that we preregistered hypotheses only for the comparisons of outcome effects between victims and bystander as well as perpetrators and bystanders. The comparison of outcome effects between victims and perpetrators seemed reasonable post hoc and should be considered as exploratory.

2.1 | Method

In a 4 × 2 design, we varied social role (perpetrator vs. bystander interacting with victim vs. bystander interacting with perpetrator vs. victim) between participants and outcome (neutral vs. bad) within participants.

2.1.1 | Participants

As determined a priori in our preregistration, we recruited 800 participants on Amazon Mechanical Turk, with IP address locations being restricted to the USA. We excluded all individuals who failed at least one of the preregistered criteria (i.e., attention checks, comprehension checks, manipulation checks, native language, time to complete the survey, completion of the survey; for details see Appendix), which amounted to $n = 52$ in the perpetrator condition, $n = 50$ in the bystander interacting with victim condition, $n = 55$ in the bystander interacting with perpetrator condition, and $n = 67$ in the victim condition. Final data of 576 participants (265 women and 311 men, Mdn$age = 35$, range 20–82) remained for analysis. A sensitivity analysis conducted with G*Power (Erdfelder, Faul, & Buchner, 1996) showed that this sample size provides 80% power to detect an interaction effect size of $\eta_p^2 = 0.02$ using an alpha of .05.

2.1.2 | Procedure

The experiment adhered to ethical guidlines specified in the APA Code of Conduct as well as national ethics guidlines. After providing informed consent, participants were randomly assigned to the

between-subjects conditions of the factorial design. Participants then read two versions of a vignette (see below), starting with the one leading to a neutral outcome. To be able to compare our results with previous research in which social role was not manipulated, we used a (slightly adapted) scenario from Kneer and Machery (2019). Order effects were likely, as negative information influences subsequent evaluations more heavily than neutral information (negative information bias; Ito, Larsen, Smith, & Cacioppo, 1998; see also Young & Tsoi, 2013). When participants consider a bad outcome first, they may have difficulty ignoring information about the bad outcome when subsequently judging an action that ends in a neutral outcome (see also Study A.2 in the Appendix). We therefore always presented the neutral outcome vignette first. Finally, we held the mental state of the actor fixed across conditions—they were always described as acting without an intention to cause harm. Prior research has shown that people infer intentions from outcomes (e.g., Fincham, 1982; Knobe, 2003; for a discussion see Malle et al., 2014). Therefore, we sought to rule out the possibility that participants use outcome severity as a proxy for intentions by explicitly stating that the actor did not intend to harm the victim. In a prior study (see Study A.1 in the Appendix) we manipulated mental states (i.e., intended vs. unintended) and found that this factor did not interact with social role or outcome severity.

Vignettes differed solely in their crucial words to manipulate treatment conditions. The following vignette was presented to participants in both bystander conditions:

Beth takes care of Mary's 2-year-old son. She fills the bath, while Mary's son stands near the tub. The phone rings in the next room. Beth tells Mary's son to stand near the tub while she answers the phone. Beth believes Mary's son will stand near the tub for a few minutes and wait for her to return. Beth leaves the room for 5 minutes. When Beth returns, Mary's son is still standing near the tub. The boy then enjoys his bath.

The vignettes in the bad outcome condition concluded with the sentence: *when Beth returns, Mary's son is in the tub, dead, face down in the water.* In the victim condition, participants read the vignettes in the parent's role. In the perpetrator condition, they read the vignettes in the caretaker's role. Otherwise, the vignettes were identical across conditions. Both bystander conditions differed only in the assessment of interpersonal goals: bystanders interacting with victims were asked to indicate their interpersonal goals towards the victim (the parent), whereas bystanders interacting with the perpetrator were asked to imagine an interaction with the perpetrator (the caretaker). Neither the vignette, nor the assessment of moral judgements, differed across these two conditions.

2.1.3 | Measures

Moral Judgement
After each experimental vignette, participants rated the permissibility and the wrongness of the action, and how much blame and punishment the actor deserved (i.e., "To what extent was it permissible for Beth to leave Mary's son alone in the above scenario?," "How wrong was Beth to leave Mary's son alone in the above scenario?," "To what extent is
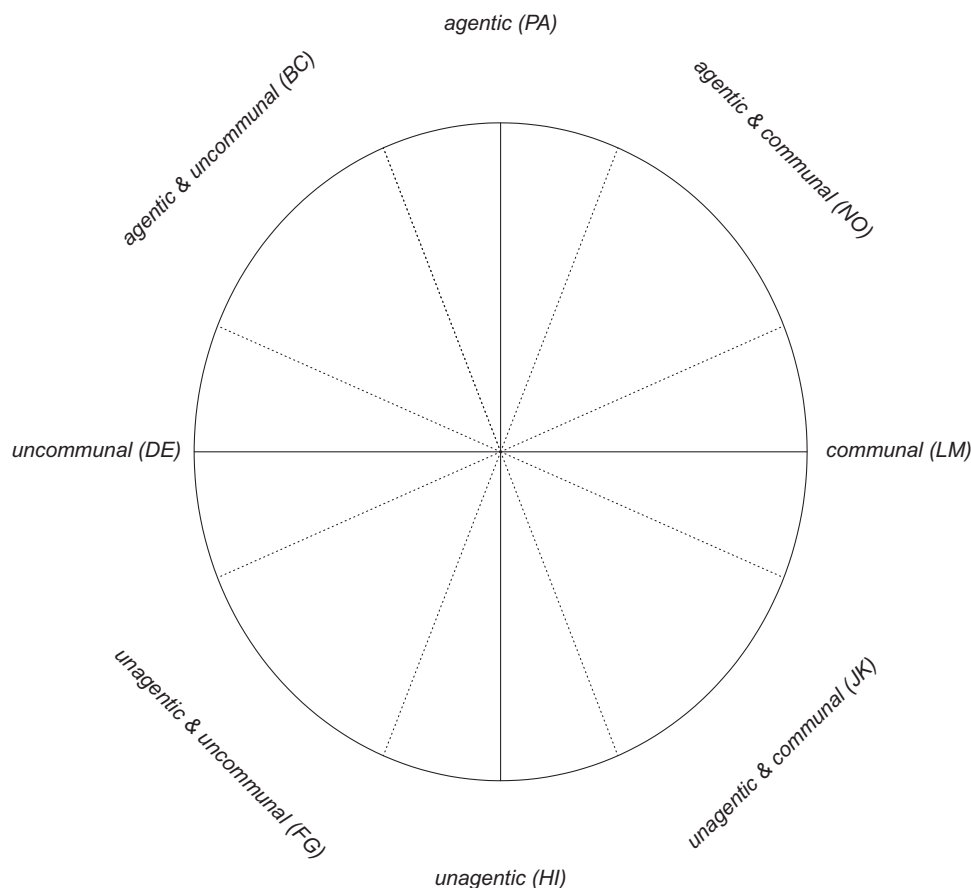
FIGURE 1 Circumplex of interpersonal goals (octants depicted with dotted lines)

Beth blameworthy for leaving Mary's son alone in the above scenario?," "How much punishment does Beth deserve for leaving Mary's son alone in the above scenarios?"). Scales ranged from 1 = Not at all (e.g., wrong) to 5 = Extremely (e.g., wrong). We averaged the items (reverse-scored) permissibility, wrongness, and blame after finding satisfactory internal consistency (Cronbach's α = .86). Following our pre-registered plan, we analyzed punishment ratings separately because recent research suggests that punishment is considerably more sensitive to outcome than permissibility, wrongness, and blame (see Cushman, 2008; Kneer & Machery, 2019). The correlations of repeated measures were .54 for moral judgement and .35 for punishment.

Interpersonal Goals

To measure interpersonal goals, we used the German version of the Circumplex Scales of Intergroup Goals (CSIG; Locke, 2014). We used the CSIG instead of its equivalent on the interpersonal level, the Circumplex Scales of Interpersonal Values (CSIV; Locke, 2000), because the CSIG is considerably shorter (32 items) than the CSIV (64 items). Since our within-subjects design required participants to answer the questionnaire twice, we decided to adapt the short version of the CSIG to the interpersonal context (following Aydin et al., 2019), in order to prevent fatigue effects. The CSIG consists of eight four-item scales each representing one octant in the circumplex (see Figure 1). Upper and

lower octants along the vertical axis define high and low agency scores, whereas right and left octants along the horizontal axis define high and low communal scores. Each point within the circumplex represents a weighted mixture of agency and communion.

After each experimental vignette (neutral or bad outcome), participants rated the 32 items of the CSIG. Specifically, they were asked to imagine an interaction with either the victim (Mary) or the perpetrator (Beth)—depending on the experimental condition—and rate the respective goals from 1 = Not important to 5 = Very important: "Please imagine you interact with Mary. Complete the following items while thinking of the question: when I interact with Mary it is important to me *that…*" (e.g., "I show concern for her welfare," "I understand her point of view" [examples for communal items], "I am assertive," "I appear confident" [examples for agentic items]; for complete list of items see Appendix). We created an overall agency score by combining the octant scores as follows: agentic goals = .414 × (PA – HI + (.707 × [BC + NO – JK – FG])) (note that each octant has by convention a generic two-letter code). This score represents a participant's position on the vertical dimension in the circumplex shown in Figure 1. Likewise, we created an overall communal score by combining the octant scores as follows: communal goals = .414 × (LM – DE + (.707 × [JK + NO – BC – FG])) (Leary, 1957; Locke, 2014). This score represents a participant's position on the horizontal dimension
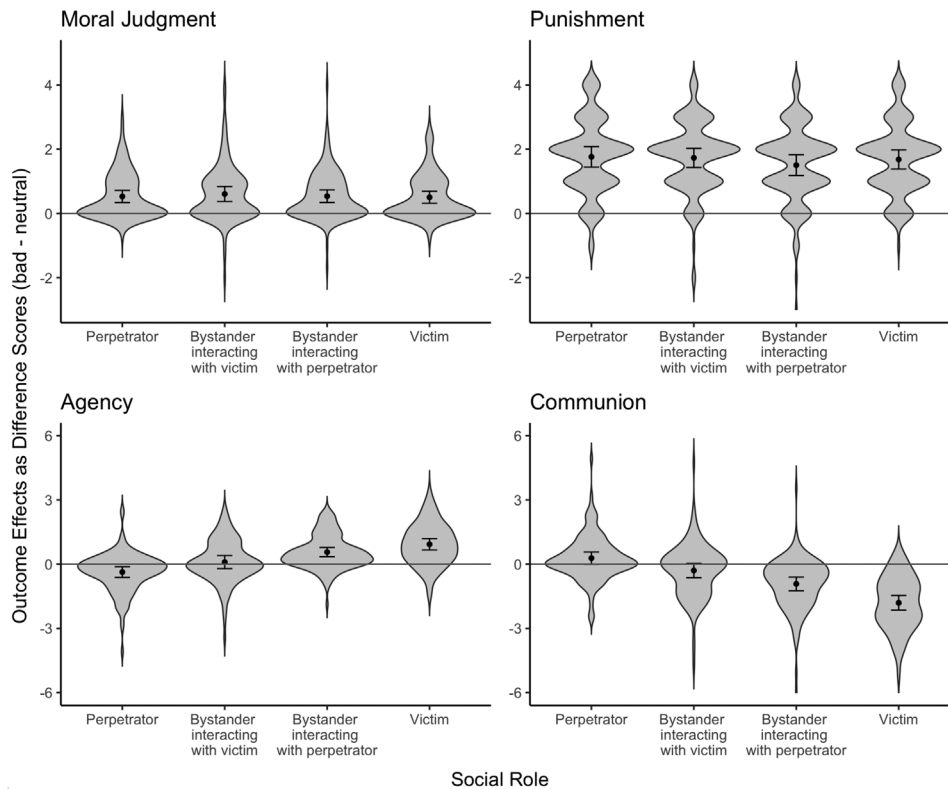
FIGURE 2 Violin plot of the difference scores (bad–neutral outcome) of perpetrators', bystanders', and victims' moral judgements (permissibility, wrongness, blame), punishment judgements, agency, and communion. Points represent means, error bars represent 95% confidence intervals, and the width of density plots represent the frequency of observations

in the circumplex shown in Figure 1. Cronbach's α were .85 for agency and .85 for communion (see Appendix for Cronbach's α of the octant scores). Correlations of repeated measures were r = .6 for agency and r = .44 for communion.

3 | RESULTS

We conducted a series of 4 (social role [perpetrator vs. bystander interacting with victim vs. bystander interacting with perpetrator vs. victim]) × 2 (outcome [neutral vs. bad]) analyses of variance (ANOVA) on moral judgement, punishment, agency, and communion with repeated measures on the second factor. We used R (Version 3.4.3; R Core Team, 2017) and the R-packages *afex* (Version 0.20.2; Singmann, Bolker, Westfall, & Aust, 2016), *psy* (Version 1.1; Falissard, 2012), *emmeans* (Version 1.4.2; Lenth, 2019), *dplyr* (Version 0.8.3; Wickham, François, Henry, & Müller, 2018), *tidyr* (Version 1.0.0; Wickham & Henry, 2019), *taRifx* (Version 1.0.6.1; Friedman, 2018) and *ggplot2* (Wickham, 2016) for analyses and figures.

3.1 | Moral judgement

Replicating previous research, we found a significant main effect of outcome, $F(1, 572) = 313.92$, $p < .001$, $\eta_p^2 = .35$, 95% CI[3]

[0.29, 0.41], such that bad outcomes elicited harsher judgements ($M = 4.73$, $SD = 0.57$) than neutral outcomes ($M = 4.18$, $SD = 0.86$; the Appendix provides separate analyses of permissibility, wrongness, and blame). However, contrary to predictions, the interaction between outcome and social role was not significant for moral judgements, $F(3, 572) = 0.51$, $p = .673$, $\eta_p^2 = .00$, 95% CI [0, 0.01]. The similarity of outcome effects across social role conditions is apparent from the distributions of difference scores (moral judgement following bad outcome minus moral judgement following neutral outcome) on the upper left plot of Figure 2. The expansion of the density plot represents the frequency of observations. Means and 95% confidence intervals are shown within the violin-shaped plots. Inspection of the plot further shows that in all social role conditions, the violin's expansion is greatest at the value of zero, which means that the most frequent response pattern (43% of participants) was to make identical moral judgements across outcome conditions.

3.2 | Punishment

The interaction between outcome and social role was not significant for judgements of punishment either, $F(3, 572) = 1.62$, $p = .182$, $\eta_p^2 = .01$, 95% CI [0, 0.02]. There was an unsurprising main effect of outcome,

3 All confidence intervals are given on effect sizes.

TABLE 1 Means and standard deviations of study variables

Variable	Perpetrator (n = 139) Neutral	Bad	Bystander interacting with victim (n = 145) Neutral	Bad	Bystander interacting with perpetrator (n = 157) Neutral	Bad	Victim (n = 135) Neutral	Bad
Moral judgement	4.33 (0.77)	4.85 (0.44)	4.06 (0.91)	4.66 (0.60)	4.17 (0.83)	4.71 (0.56)	4.18 (0.93)	4.68 (0.65)
Punishment	2.88 (1.12)	4.65 (0.66)	2.70 (1.07)	4.43 (0.83)	2.89 (1.10)	4.39 (0.90)	2.70 (1.02)	4.39 (0.98)
Agency	0.09 (0.76)	−0.28 (0.84)	0.27 (0.80)	0.36 (1.14)	0.49 (0.88)	1.05 (1.08)	0.73 (0.81)	1.65 (1.09)
Communion	1.43 (1.02)	1.71 (1.08)	1.08 (1.15)	0.78 (1.53)	0.66 (1.15)	−0.26 (1.29)	1.03 (1.10)	−0.78 (1.29)

Note: Means and standard deviations (in parentheses) are given for perpetrators', bystanders', and victims' moral judgements, punishment judgements, agency, and communion after neutral and bad outcomes.

$F(1, 572) = 1278.13$, $p < .001$, $\eta_p^2 = .69$, 95% CI [0.65, 0.72], such that punishment in the bad outcome conditions ($M = 4.46$, $SD = 0.85$) exceeded punishment in the neutral outcome conditions ($M = 2.80$, $SD = 1.08$). The upper right plot of Figure 2 shows that the outcome effect was more pronounced for punishment than for moral judgement.

3.3 | Interpersonal goals

In contrast to the results for moral judgement and punishment, there were significant interactions between outcome and social role for both agentic, $F(3, 572) = 50.44$, $p < .001$, $\eta_p^2 = .21$, 95% CI [0.15, 0.26], and communal goals, $F(3, 572) = 82.05$, $p < .001$, $\eta_p^2 = .30$, 95% CI [0.24, 0.35]. As predicted, victims' increases in agency following the bad outcome were more pronounced than the increases in agency of bystanders interacting with the perpetrator, $t(572) = 3.34$, $p = .001$, $d = .42$, 95% CI [0.19, 0.66], and the decreases in agency of perpetrators, $t(572) = 4.9$, $p < .001$, $d = .6$, 95% CI [0.36, 0.84] (see Table 1). One unexpected result apparent from the lower left plot in Figure 2 is that the outcome effect among perpetrators (decreases in agency) was stronger than the (non-significant) outcome effect among bystanders interacting with victims, $t(572) = 2.53$, $p = .012$, $d = .28$, 95% CI [0.05, 0.51]. However, perpetrators' outcome effect was smaller than the outcome effect of bystanders interacting with perpetrators, though not significantly so $t(572) = −1.71$, $p = .087$, $d = .22$, 95% CI [−0.01, 0.45].

The lower right plot in Figure 2 illustrates the pattern of outcome effects on communion across social role conditions. Statistical comparisons confirm the impression that the violin plot representing communion is virtually a mirror image of the violin plot representing agency: As predicted, victims' decreases in communion were stronger than the decreases in communion of bystanders interacting with the perpetrator, $t(572) = 6.5$, $p < .001$, $d = .74$, 95% CI [0.51, 0.98], and the increases in communion of perpetrators, $t(572) = 10.93$, $p < .001$, $d = 1.35$, 95% CI [1.1, 1.61]. Similar to the results observed for agency, the condition of bystanders interacting with the victim did not conform to predictions, with its outcome effect being indistinguishable from the outcome effect of perpetrators, $t(572) = 0.15$, $p = .882$, $d = .02$, 95% CI [−0.21, 0.25]. Also unexpectedly, bystanders interacting with

the victim showed less communion following the bad outcome than following the neutral outcome (see Table 1). However, perpetrators' outcome effect was significantly smaller than the outcome effect of bystanders interacting with perpetrators, $t(572) = −4.78$, $p < .001$, $d = .59$, 95% CI [0.35, 0.82].

4 | SUMMARY OF RESULTS

The outcome effect on moral judgements and punishment judgements was not significantly different across social roles. In light of the fact that the sensitivity of our design allowed us to detect effects as small as $\eta_p^2 = 0.02$, this suggests that differences due to social role are negligible. The lack of interaction is also unlikely to be due to a ceiling effect, considering that victims' moral judgements following bad outcomes—predicted to be more severe—was actually lower than perpetrators' moral judgements.

However, as predicted, social role moderated the outcome effect on interpersonal goals: compared with perpetrators, victims showed stronger outcome effects on agency and communion. Furthermore, we found that victims' outcome effects were larger than bystanders' outcome effects when interacting with perpetrators.

The results for the condition of bystanders interacting with victims were not as expected. First, outcome effects were not stronger than in the perpetrator condition. Second, bystanders interacting with victims pursued less communion following the bad outcome than following the neutral outcome. These unexpected results might have been due to the subtlety of the instructions. We first asked participants to judge the perpetrator's behaviour (e.g., "How wrong was it for Beth to leave Mary's son alone in the above scenario?"), and we then asked them to indicate their interpersonal goals towards the victim (e.g., "When I interact with Mary, it is important to me that I am friendly"). The name of the parent ("Mary") was the only cue that participants should imagine an interaction with the victim. Some participants may have overlooked this information, and instead imagined further interactions with the perpetrator (as they also had to assess the behaviour of the perpetrator in moral regards just beforehand). This might have compromised the construct validity of this condition. Therefore, we omit the bystander interacting with victim condition for subsequent analyses.

TABLE 2 Correlations of outcome effects on moral judgement and punishment with outcome effects on agency and communion

| Social Role | Moral judgement | | Punishment | |
	Agency	Communion	Agency	Communion
Victim	.08	.03	.12	−.29[***]
Bystander[†]	.18[*]	−.16[*]	.15	−.19[*]
Perpetrator	.07	.07	.03	.03

Note: [*] *p* < .05, [***] *p* < .001. [†]Bystander interacting with perpetrator.

5 | ROBUSTNESS CHECKS USING BAYESIAN REGRESSION MODELING

Might our non-significant result regarding moral judgements and punishment judgements be due to violated statistical assumptions? We conducted Bayesian analyses for ordinal variables because we could not be certain that our measures of moral judgements satisfied the assumptions of interval scaling (cf. Liddell & Kruschke, 2018). Bayesian modelling allows for specifying a potentially more appropriate distribution of the dependent variable.

Using the R-Package *brms* (Bürkner, 2017), we fitted Bayesian Generalised Multilevel Models by regressing each item (permissibility, wrongness, blame, and punishment) separately on social role and outcome, using a cumulative link function, minimally informative priors, and random intercepts for participants. The results were highly similar to the results obtained with ANOVA modelling. Bad outcomes were judged more harshly than neutral outcomes: the 95% Credibility Intervals did not include 0 (for full report of results see Appendix). With regard to the interaction between outcome and social role, in line with ANOVA models, these analyses generally (in 11 out of 12 analyses) confirmed the absence of evidence for a moderating effect of social role. Treating moral judgements as ordinal data, as the Bayesian Generalised Multilevel Model allowed us to do, is statistically sounder than the preregistered metric ANOVA reported above. These analyses increase our confidence in the overall conclusion that divergent moral judgements following bad and neutral outcomes are not moderated by social role.

6 | ARE JUDGEMENTS AND GOALS RELATED?

The evidence suggests that the social role people assume during a transgression moderates the size of outcome effects on interpersonal goals, but not the size of outcome effects on moral judgements. This divergence is at variance with our theoretical assumption that the harshness of moral judgements determines the strength of subsequent role-specific interpersonal goals. Given that both moral judgements and interpersonal goals were measured but not manipulated, we do not know whether moral judgements are causes, consequences, or mere correlates of interpersonal goals. As an indirect test of within-subject mediation (Judd, Kenny, & McClelland, 2001), we performed correlation analyses. Depicted in Table 2, the correlations of the outcome effects on judgements and goals were small and mostly not sig-

nificantly different from zero, yielding no evidence for mediation or reverse mediation. In other words, the extent to which victims and perpetrators adjust the level of harshness of their moral judgement to the severity of the outcome does not predict the extent to which they adjust their goals for interacting with each other. The only exceptions were that (a) bystanders' outcome effects on moral judgements and agentic as well as communal goals were correlated; and (b) victims' and bystanders' outcome effects on punishment and communal goals were correlated, such that outcome-dependent increases in the willingness to punish the perpetrator were accompanied by decreases in communal goals.

Considering neutral and bad outcomes separately (see Table 3), however, we find systematic correlations between moral judgements and interpersonal goals speaking to individual differences in the extremity of judgements independent of outcome severity. The more harshly perpetrators judged their own actions, the stronger the communal goals they wanted to pursue. By contrast, victims' and bystanders' moral judgements and calls for punishment were positively associated with their agentic goals and negatively with their communal goals. The more harshly they judged a transgression, the more agentically and the less communally they wanted to act. This pattern fits the predictions of the needs-based model (Shnabel & Nadler, 2015). The degree to which parties view an action as immoral corresponds to the intensity of role-specific agentic and communal goals.

Taken together, the findings suggest that (pre-existing) individual differences in moral judgement are systematically related to interpersonal goals, but that the moral judgement differences produced by the manipulation of outcome severity are for the most part unrelated to differences in interpersonal goals between neutral and bad outcomes.

7 | SUMMARY OF ADDITIONAL STUDIES

We briefly summarise two additional studies that we chose not to include in this report due to their redundancy (for a full description of methods and results the reader is referred to the Appendix). In Study A.1 (*N* = 107), we contrasted only perpetrators and victims using a different moral scenario previously employed in research on outcome effects (Cushman, 2008). In this scenario, we asked participants to imagine being in a sculpture class, welding pieces of metal with a partner. The perpetrator welds together two pieces of metal, one of which their partner (the victim) is holding in his hands. Depending on the outcome condition, the partner happened to let go of the metal (and is not

TABLE 3 Correlations of moral judgement and punishment with agency and communion within outcome conditions

	Moral judgement and agency		Moral judgement and communion		Punishment and agency		Punishment and communion	
	Neutral outcome	Bad outcome	Neutral outcome	Bad outcome	Neutral outcome	Bad outcome	Neutral outcome	Bad outcome
Victim	.31***	.35***	−.17*	−.43***	.17	.35***	−.33***	−.5***
Bystander[†]	.32***	.31***	−.1	−.2	.29***	.33***	−.25**	−.27***
Perpetrator	.02	−.04	.24**	.27**	−.02	−.05	−.05	.21*

Note: * $p < .05$, ** $p < .01$, *** $p < .001$. [†] Bystander interacting with perpetrator.

burned at all) or did not let go (and the partner's hand is burned). This study revealed the predicted interaction for moral judgement and communion but not for punishment and agency.

The Study A.2 ($N = 267$) followed the same procedure as the main study presented above except that it did not include a bystander interacting with the victim condition and the order of bad and neutral outcomes was varied. When the bad outcome was presented first, outcome effects on moral judgement and interpersonal goals were negligible. By contrast, when participants judged the neutral outcome first, outcome effects on moral judgement emerged for all social roles. A fixed order of vignettes, with the neutral outcome vignette preceding the bad outcome vignette, therefore, seemed to be a methodological prerequisite for examining moderators of the outcome effect. Focusing on the subset of participants who received the neutral outcome vignette first ($N = 125$), we found that the outcome effects on moral judgements and punishment judgements were not significantly different across social roles. However, consistent with our hypothesis about interpersonal goals, social role moderated outcome effects on agency and communion: victims showed the strongest outcome effects and perpetrators showed the weakest outcome effects, with bystanders falling in between.

Including these two additional studies in the Bayesian robustness checks of moral judgement items and mediation analyses of judgements and goals does not change conclusions. The only discrepancy was that Bayesian regression analyses revealed evidence for the absence of an interaction between outcome and social role on moral judgement also for Study A.1, where the ANOVA yielded a significant interaction. Overall, the results of these two additional studies (a) increase our confidence that the moderation of social role on the outcome effect may not exist for moral judgement, and (b) corroborate our predictions about the interaction of social role and outcome on interpersonal goals.

8 | DISCUSSION

The present research replicates and extends previous research on outcome effects by considering the different social roles of judges and measuring their goals for interactions following a transgression. Bearing in mind that real transgressions are more complex and may provoke more intense responses than the idiosyncratic scenarios we used, we believe that the minimal scenarios, together with the experimental

method, allowed us to make a set of strong contributions with theoretical implications for previously unrelated literatures. We first discuss how our results add to the literature on moral judgement, before we turn to our key finding on moral action (i.e., social role moderated the outcome effect on interpersonal goals) and its implication for reconciliation processes.

8.1 | Implications for moral judgement

We replicated the outcome effect for moral judgement among bystanders (e.g., Cushman, 2008; Cushman et al., 2009; Gino et al., 2009; Lench et al., 2015; Young et al., 2010) and extended it to victims and perpetrators. Participants judged actions resulting in a bad outcome more harshly than actions resulting in a neutral outcome even when the bad outcome was a matter of bad luck. However, contrary to predictions, social roles did not moderate these effects.

We confirmed previous reports of large proportions of participants making identical moral judgements for neutral and bad outcomes (43%; Kneer & Machery, 2019; Schwitzgebel & Cushman, 2012). However, our proportions are lower than the ones reported by Schwitzgebel and Cushman (2012) or Kneer and Machery (2019). The latter report a that roughly 80% of participants ascribe the same measure of wrongness, permissibility and blame to the lucky and the unlucky agents. The philosophical *Puzzle of Moral Luck*, they suggest, is thus misconceived: the puzzle refers to a within-subjects intuition or a comparative evaluation of actions and agents, but robust outcome effects can only be obtained in between-subjects designs. In our experiment, the proportion of participants who judge the agents identically across outcomes, although still substantial, is lower. We would like to suggest that this difference is the consequence of two design features: first, rather than inviting a comparative assessment by having participants read both scenarios before making their judgements, participants read one scenario, judged the agent, filled in a lengthy questionnaire concerning interpersonal goals with more than two dozen questions, and were then presented with the second scenario. The contrastive aspect of the within-subjects design was thus considerably reduced. Second, based on evidence that outcome effects are stronger when the neutral outcome is presented first (see Study A.2 in the Appendix), we purposefully chose not to counterbalance the order of outcomes. In short, the present results are consistent with previous research and cannot support a revival of the philosophical *Puzzle of Moral Luck*.

As will be recalled, we followed Cushman (2008) by using four distinct variants of moral judgement: permissibility, wrongness, blame, and deserved punishment. Cushman's model separates moral judgement into two kinds. They differ in their relative weighting of information pertaining to the agent's mental states on the one hand and factors regarding the causal action sequence (including outcome) on the other. On Cushman's view, permissibility and wrongness judgements (Type$_1$) are predominantly—though not exclusively—influenced by mental factors, whereas judgements of blame and deserved punishment (Type$_2$) are strongly sensitive to causal factors such as severity of outcome. However, recent studies suggest that blame patterns with permissibility and wrongness rather than with deserved punishment, in so far as the latter is more sensitive to causal factors than the other three types of moral judgement (for a discussion see Kneer & Machery, 2019). We too found the effect size of outcome on blame to be similar to those for outcome on permissibility and wrongness (they ranged between $\eta_p^2 = .23$ and .31 for permissibility, wrongness and blame). The effect size of outcome on punishment, by contrast, was about twice as pronounced ($\eta_p^2 = .69$, see Appendix for detailed results). Note, however, that the question concerning the status of blame is far from settled and pursued in more detail in ongoing research by Prochownik and Cushman (2020).

8.2 | Implications for moral action

Our key finding was that victims and perpetrators differed in their outcome effects on interpersonal goals regarding their social interaction after the transgression. Consistent with the needs-based model (Shnabel & Nadler, 2015), victims had more agentic and more uncommunal goals when transgressions resulted in harm rather than no harm, whereas perpetrators had more communal and unagentic goals. Consistent with the prediction that victims would weight outcome information more heavily than perpetrators (cf. Pronin, 2009), the outcome effects on role-specific goals were stronger among victims than among perpetrators, with outcome effects among bystanders falling in between.

Within-subjects mediation analyses suggested that the outcome effects on moral judgement and interpersonal goals were for the most part unrelated. This raises questions about strict cognitivism, according to which moral judgement (and not, for instance, non-conscious affective responses) drives moral action (here interpersonal goals). However, from the point of view of cognitivism, it is a welcome finding that moral judgement is insensitive to social role: whether or not an action is wrong, and whether or not an agent is blameworthy, does not depend on whether the judging party finds themselves in the role of victim, perpetrator or bystander. All it takes is an explanation for the decoupling of the outcome sensitivity with regard to judgements and action dispositions.

Recall that the *Puzzle of Moral Luck* arises due to the fact that, in line with the Control Principle inspired by Kant, we do not want to blame people for factors that go beyond their control. On the other hand,

and in line with Consequentialism, we do seem to register a normative difference across the lucky and the unlucky agent. According to one intriguing attempt to reconcile the conflicting intuitions, the perceived normative difference across agents is not *moral* in nature (and hence cannot clash with the Control Principle). Could something along these lines explain our surprising findings?

In his original discussion of moral luck, Williams (1981) argues that in cases where an agent is the causal origin of a bad consequence yet entirely free from moral blame, they will feel a type of regret that differs from both ordinary (non-moral) regret on the one hand and (morally inflected) remorse or guilt on the other. What is distinctive of this kind of situation is that, in contrast to a mere bystander, the agent "might have acted otherwise, and the focus of the regret is on that possibility, the thought being informed [in part by first-personal conceptions of how one might have acted otherwise]" (Williams, 1981, p. 27).

Agent-regret, Williams emphasises, is not confined to the mental level but engenders distinctive behavioural dispositions: those who blamelessly harm another tend to be disposed to make amends. What is more, rather than being morally *obliged*, they have—as philosophers would say—special normative *powers* to *take* responsibility for the outcome, much like parents do for their children (Enoch, 2012). The kind of responsibility the blameless agents can decide and might be expected to shoulder, Enoch suggests, parallels the kind of responsibility we assume in making a promise: it does not arise from a moral debt towards others, but is generated by a voluntary commitment.

If the specific social role (agent vs. bystander) leaves a morality-independent normative footprint, then it will not stop short of the victim: the victim can reasonably expect the agent to manifest agent-regret, to take responsibility for the outcome, and to provide restitution of some sort.

While Williams and Enoch predict distinctive, role-specific interpersonal goals, these are created directly by the causal structure, and they are independent of the moral considerations invoked by the needs-based model. Agent-regret would thus provide the kind of explanation needed in defence of (the possibility of) cognitivism: moral judgements are not, and should not, be role-dependent. Moral action, rather than being a direct consequence of an assessment of the moral debt that the perpetrator owes to the victim (Shnabel & Nadler, 2015), might result from normative expectations regarding the voluntary commitment the perpetrator should make to redress the harm.

One might think that agent regret and the normative powers it engenders apply first and foremost to agents (i.e., perpetrators). It is perpetrators who feel agent regret, and whose normative powers are increased when the consequences of their actions are dire. This might seem at odds with the fact that we observed the largest outcome effects not among perpetrators, but among victims. However, it is important to note that normative expectations drive the quality of the behaviour (i.e., instead of the moral judgements), but not the extent to which outcome information is utilised. As we predicted and found, victims pay more attention to outcome information when setting their interpersonal goals than bystanders or perpetrators.

8.3 | Implications for reconciliation processes

Our results regarding the independence of moral judgements and interpersonal goals have interesting implications for the explanation of the different needs of perpetrators and victims as postulated by the needs-based model. Building on the assumption that moral debt is what drives interpersonal behaviour, manipulation checks used in previous research typically assessed the extent to which perpetrators acknowledged their offences (e.g., Shnabel & Nadler, 2008) or privileged groups conceded that their privileges are illegitimate (e.g., Aydin et al., 2019). Apparently, the predictions of the needs-based model also hold when agents accidentally and thus blamelessly harm others, in which case they cannot be considered perpetrators proper. This is an intriguing finding which deserves further study.

Reconciliation is fundamentally a dyadic process as is evident in the apology-forgiveness cycle where apology and forgiveness presuppose each other (Tavuchis, 1991). To the extent that this dyadic exchange involves sensitive calibration of responses, our results regarding the mismatch of moral judgements severity and interpersonal goals also have important implications for successful reconciliation. Apologies are more likely to be accepted and lead to forgiveness if they are matched in length and content to the severity of the transgression (Bennett & Earwaker, 1994; Kirchhoff, Wagner, & Strack, 2012). However, the lack of connection between moral judgement and moral action causes concern about such a fit: parties adapt their interpersonal goals in response to outcome severity differently. No matter whether actions resulted in harm or not, perpetrators always showed communal goals towards the victim, whereas victims only increased their agentic goals towards the perpetrator after harmful outcomes. Although it is yet to be shown that agentic goals relate to the desire to receive apologies and communal goals relate to the willingness to offer apologies, one could speculate that perpetrators' willingness to apologise is greater than victims desire to receive apologies after neutral outcomes. In fact, Leunissen, De Cremer, Folmer, and Van Dijke (2013) reported a similar finding, namely that perpetrators' willingness to offer an apology exceeded victims' desire to receive an apology after unintended (but not intended) transgressions. Thus, our finding that victims' interpersonal goals are more contingent on outcome severity than on perpetrators' goals suggests a potential miscalibration between the type of apology expected by the victim and the one the perpetrator is offering.

Besides the role of outcome severity for apologies, research shows that outcome severity conditions the victim's willingness to forgive the perpetrator. The more severe a transgression's outcome, the more time has to pass in order for victims to forgive their perpetrators (McCullough, Luna, Berry, Tabak, & Bono, 2010). As McCullough et al. (2003) state, consequences of severe (as compared to minor) transgressions are more profound and enduring and are, therefore, "likely to serve as cues to transgression recipients that it is still in their best interests to take self-protective stances toward the transgressor" (p. 543). Our finding that victims show increased agentic goals after more severe outcomes is in line with this theorising. It is, thus, plausible that the different effects of outcome severity on victims' and perpetrators' interpersonal goals extend to victims' willingness to forgive or perpetrators'

expectations to be forgiven. Perpetrators may underestimate the time that has to pass until victims forgive their perpetrators when the transgression was harmful (rather than harmless).

8.4 | Limitations and future research

Some limitations and their implications for future research should be noted. First, we investigated our research questions solely in mixed designs, always varying outcome within-subject and social role between-subjects. Although this study design was derived from insights of prior research, future research is needed to examine the external validity of our results using different designs. For instance, since the strength of outcome effects on moral judgements hinges on whether outcome is varied within-subject or between-subjects, it is possible that the interaction of social role and outcome yields different results as well.

Second, the scenarios we used to investigate the interaction of social role and outcome on moral judgement and interpersonal roles were rather specific. Although we replicated our findings in two different scenarios (bathtub scenario as in the main study and sculpture class scenario as in Study A.1), we cannot be certain about the results' generalisability across social contexts. Negligent actions, such as not carefully sticking to COVID hygiene rules or drunk driving, which could equally result in harmless or harmful consequences, are frequent and bear potential for misunderstandings. It is, therefore, important to better understand how victims' and perpetrators' moral judgements and interpersonal goals diverge after such actions. While the present research offers a starting point, future research is needed to assess the generalisability of findings across different social contexts.

Finally, outcome effects do not stop short of interpersonal situations, but are also common in intergroup settings. For instance, in 2015, the European countries Norway and Hungary had equally strict migration laws, but Hungary happened to be located in an area where many refugees passed, whereas Norway was not. While many people disapproved of the way the Hungarian administration dealt with refugees, Norway was off the hook. It is plausible that the differential sensitivity to outcome effects of victims and perpetrators will also be observed at the intergroup level. To give an example, the literature on competitive victimhood shows that disagreement about intergroup conflicts is prevalent: conflicting groups accentuate the severity of intergroup transgressions in a way that allows them to present their ingroup as the real victims (SimanTov-Nachlieli & Shnabel, 2014). Would judgements about the severity of intergroup transgressions also diverge from intergroup goals pertaining to agency and communion? Research on these goals associated with victim and perpetrator groups has so far not paid much attention to the role of outcome severity. This would be a natural starting point for future research.

9 | CONCLUSION

Our results suggest that when victims and perpetrators reflect on a transgression, they tend to agree with each other (and with bystanders)

in their judgements of the permissibility and wrongness of the action, their (self-)blame, and the degree of appropriate punishment. For some participants, these moral judgements were unaffected by the severity of the outcome, while for others, judgements were harsher when outcomes were more severe, driving the overall outcome effects that we observed. Largely independent of moral judgement, however, victims and perpetrators were differentially sensitive to outcome information when they reported their interpersonal goals regarding agentic and communal action towards each other. Our discussion points to the fascinating possibility that moral action can, at least in parts, be explained by role-specific non-moral normative demands familiar from the philosophical literature on agent-regret (Enoch, 2012; Williams, 1981) paired with selective information-processing predicted by the introspection illusion (Pronin, 2009).

CONFLICT OF INTEREST

The authors report no conflicts of interest.

ETHICS APPROVAL

Authors confirm that this article adheres to ethical guidelines specified in the APA Code of Conduct as well as the authors' national ethics guidelines.

DATA AVAILABILITY STATEMENT

All data is available at https://osf.io/t7e2k/?view_only=4eb9292 ce08048c99705072f0301308e.

ORCID

Lisa Katharina Frisch https://orcid.org/0000-0003-4451-9587
Markus Kneer https://orcid.org/0000-0002-4223-0715
Joachim Israel Krueger https://orcid.org/0000-0001-9607-1695
Johannes Ullrich https://orcid.org/0000-0002-0471-7004

REFERENCES

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*, 751–763. https://doi.org/10.1037/0022-3514.93.5.751

Adams, G. S. (2016). Asymmetries between victims' and transgressors' perspectives following interpersonal transgressions. *Social and Personality Psychology Compass*, *10*(12), 722–735. https://doi.org/10.1111/spc3.12291

Adams, G. S., & Inesi, M. E. (2016). Impediments to forgiveness: Victim and transgressor attributions of intent and guilt. *Journal of Personality and Social Psychology*, *111*(6), 866–881. https://doi.org/10.1037/pspi0000070

Aydin, A., Ullrich, J., Siem, B., Locke, K., & Shnabel, N. (2019). The effect of social class on agency and communion: Reconciling identity-based and rank-based perspectives. *Social Psychological and Personality Science*, *10*(6), 735–745. https://doi.org/10.1177/1948550618785162

Bakan, D. (1966). *The duality of human existence. An essay on psychology and religion.* Rand Mcnally.

Baumeister, R. F. (1999). *Evil: Inside human violence and cruelty.* Macmillan.

Baumeister, R. F., Stillwell, A., & Wotman, S. R. (1990). Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger. *Journal of Personality and Social Psychology*, *59*(5), 994–1005. https://doi.org/10.1037/0022-3514.59.5.994

Bennett, M., & Earwaker, D. (1994). Victims' responses to apologies: The effects of offender responsibility and offense severity. *The Journal of Social Psychology*, *134*(4), 457–464. https://doi.org/10.1080/00224545.1994.9712196

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgement. *Cognition*, *108*(2), 353–380. https://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *Plos One*, *4*(8), e6699. https://doi.org/10.1371/journal.pone.0006699.t001

Enoch, D., (2012). Being responsible, taking responsibility, and penumbral agency. In Heuer, U., & Lang, G., *Luck, value, and commitment: Themes from the ethics of Bernard Williams* (pp. 95–132). Oxford University Press.

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self-or social prediction?. *Journal of Personality and Social Psychology*, *79*(6), 861. https://doi.org/10.1037/0022-3514.79.6.861

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Falissard, B. (2012). Psy: Various procedures used in psychometry. Retrieved from https://CRAN.R-project.org/package=psy

Feltz, A., M. Harris, & Perez, A. (2012). Perspective in intentional action attribution. *Philosophical Psychology*, *25*(5), 673–687. https://doi.org/10.1080/09515089.2011.627540

Fincham, F. (1982). Moral judgment and the development of causal schemes. *European Journal of Social Psychology*, *12*(1), 47–61. https://doi.org/10.1002/ejsp.2420120104

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from the perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902. https://doi.org/10.1037/pspa0000163

Friedman, A. B. (2018). *taRifx: Collection of utility and convenience functions.* Retrieved from https://CRAN.R-project.org/package=taRifx

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *75*(3), 617–638. https://doi.org/10.1037/0022-3514.75.3.617

Gino, F., Moore, D. A., & Bazerman, M. H. (2009). No harm, no foul: The outcome bias in ethical judgements. *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 Behaviour and Human Decision Processes*, *111*(2), 93–101. https://doi.org/10.1016/j.obhdp.2009.11.(?PMU?)001

Gold, N., Pulford, B. D., & Colman, A. M. (2015). Do as I say, don't do as I do: Differences in moral judgements do not translate into differences in decisions in real-life trolley problems. *Journal of Economic Psychology*, *47*, 50–61. https://doi.org/10.1016/j.joep.2015.01.001

Hartman, R. J. (2017). In *defense of moral luck: Why luck often affects praiseworthiness and blameworthiness.* Routledge.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887. https://doi.org/10.1037/0022-3514.75.4.887

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the cause of behavior. In E. E. Jones, D. E. Kanouse, H. H.

Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). General Learning Press.

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, 6(2), 115–134. https://doi.org/10.1037/1082-989X.6.2.115

Kant, I. (1998/1785). Groundwork of the metaphysics of morals. *Translated by Mary Gregor*. Cambridge University Press. (Originally published 1785.)

Kirchhoff, J., Wagner, U., & Strack, M. (2012). Apologies: Words of magic? The role of verbal components, anger reduction, and offence severity. *Peace and Conflict: Journal of Peace Psychology*, 18(2), 109–130. https://doi.org/10.1037/a0028092

Kneer, M. (2018). Perspective and epistemic state ascriptions. *Review of Philosophy and Psychology*, 9(2), 313–341. https://doi.org/10.1007/s13164-017-0361-4

Kneer, M., & Machery, E. (2019). No luck for moral luck. *Cognition*, 182, 331–348. https://doi.org/10.1016/j.cognition.2018.09.003

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(3), 190–194. https://doi.org/10.1093/analys/63.3.190

Leary, T. (1957). *Interpersonal diagnosis of personality*. Ronald Press.

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. https://doi.org/10.1111/bjop.12072

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. Retrieved from https://CRAN.R-project.org/package=emmeans

Leunissen, J. M., De Cremer, D., Folmer, C. P. R., & Van Dijke, M. (2013). The apology mismatch: Asymmetries between victim's need for apologies and perpetrator's willingness to apologize. *Journal of Experimental Social Psychology*, 49(3), 315–324. https://doi.org/10.1016/j.jesp.2012.12.005

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong?. *Journal of Experimental Social Psychology*, 79, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

Locke, K. D. (2000). Circumplex scales of interpersonal values: Reliability, validity, and applicability to interpersonal problems and personality disorders. *Journal of Personality Assessment*, 75(2), 249–267. https://doi.org/10.1207/S15327752JPA7502_6

Locke, K. (2014). Circumplex scales of intergroup goals: An interpersonal circle model of goals for interactions between groups. *Personality and Social Psychology Bulletin*, 40(4), 433–449. https://doi.org/10.1177/0146167213514280

Locke, K. D. (2015). Agentic and communal social motives. *Social and Personality Psychology Compass*, 9(10), 525–538. https://doi.org/10.1111/spc3.12201

Malle, B. F. (2005). Folk theory of mind: Conceptual foundations of human social cognition. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 225–255). Oxford University Press.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, 25(2), 147–186. https://doi.org/10.1080/1047840X.2014.877340

Malle, B. F., & Pearce, G. E. (2001). Attention to behavioural events during interaction: Two actor–observer gaps and three attempts to close them. *Journal of Personality and Social Psychology*, 81(2), 278. https://doi.org/10.1037/0022-3514.81.2.278

Martin, J. W., & Cushman, F. (2015). The adaptive logic of moral luck. In J. Sytsma & W. Buckwalter (Eds.), *The Blackwell companion to experimental philosophy* (pp. 190–202). Wiley-Blackwell.

McCullough, M. E., Luna, L. R., Berry, J. W., Tabak, B. A., & Bono, G. (2010). On the form and function of forgiving: Modeling the time-forgiveness relationship and testing the valuable relationships hypothesis. *Emotion (Washington, D.C.)*, 10(3), 358–376. https://doi.org/10.1037/a0019349

McCullough, M. E., Fincham, F. D., & Tsang, J. A. (2003). Forgiveness, forbearance, and time: The temporal unfolding of transgression-related interpersonal motivations. *Journal of Personality and Social Psychology*, 84(3), 540–557. https://doi.org/10.1037//0022-3514.84.3.540

Nagel, T. (1979). *Moral luck: Mortal questions*. New York: Cambridge University Press.

Nelkin, D. K. (2019). Moral Luck. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019 ed.). Retrieved from https://plato.stanford.edu/archives/sum2019/entries/moral-luck/

Noor, M., Shnabel, N., Halabi, S., & Nadler, A. (2012). When suffering begets suffering: The psychology of competitive victimhood between adversarial groups in violent conflicts. *Personality and Social Psychology Review*, 16, 351–374. https://doi.org/10.1177/1088868312440048

Prochownik & Cushman (2020). Replication of Kneer and Machery (2019): Exploring the sensitivity of judgments of blame vs. blameworthiness to outcomes in moral luck scenarios. Unpublished Manuscript.

Pronin, E. (2009). The introspection illusion. *Advances in Experimental Social Psychology*, 41, 1–67. https://doi.org/10.1016/S0065-2601(08)00401-2

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, 43(4), 565–578. https://doi.org/10.1016/j.jesp.2006.05.011

Pronin, E., Wegner, D. M., McCarthy, K., & Rodriguez, S. (2006). Everyday magical powers: The role of apparent mental causation in the overestimation of personal influence. *Journal of Personality and Social Psychology*, 91, 218–231. https://doi.org/10.1037/0022-3514.91.2.218

R. C., ore Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, 86, 61–79. https://doi.org/10.1037/0033-295X.86.1.61

Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, 18(5), 803–817. https://doi.org/10.1162/jocn.2006.18.5.803

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, 22(1), 32–70. https://doi.org/10.1177/1088868317698288

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. https://doi.org/10.1111/j.1468-0017.2012.01438.x

Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: Satisfying the differential emotional needs of victim and perpetrator as a key to promoting reconciliation. *Journal of Personality and Social Psychology*, 94(1), 116–132. https://doi.org/10.1037/0022-3514.94.1.116

Shnabel, N., & Nadler, A. (2015). The role of agency and morality in reconciliation processes: The perspective of the needs-based model. *Current Directions in Psychological Science*, 24(6), 477–483. https://doi.org/10.1037/0022-3514.94.1.116

Siem, B., Von Oettingen, M., Mummendey, A., & Nadler, A. (2013). When status differences are illegitimate, groups' needs diverge: Testing the needs-based model of reconciliation in contexts of status inequality. *European Journal of Social Psychology*, 43(2), 137–148. https://doi.org/10.1002/ejsp.1929

SimanTov-Nachlieli, I., & Shnabel, N. (2014). Feeling both victim and perpetrator: Investigating duality within the needs-based model. *Personality and Social Psychology Bulletin*, 40(3), 301–314. https://doi.org/10.1177/0146167213510746

SimanTov-Nachlieli, I., Shnabel, N., Aydin, A. L., & Ullrich, J. (2018). Agents of prosociality: Agency affirmation promotes mutual prosocial tendencies and behaviour among conflicting groups. *Political Psychology*, 39, 445–463. https://doi.org/10.1111/pops.12418

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *Afex: Analysis of factorial experiments*. Retrieved from https://CRAN.R-project.org/package=afex

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, 4, 1–8. https://doi.org/10.3389/fpsyg.2013.00250

Tavuchis, N. (1991). *Mea culpa: A sociology of apology and reconciliation.* Stanford University Press.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37,* 395–412. https://doi.org/10.1037/0022-3514.37.3.395

Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behaviour. In W. Grove & D. Ciccetti (Eds.), *Thinking clearly about psychology: Essays in honour of Paul Everett Meehl* (pp. 89–113). University of Minnesota Press.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag.

Wickham, H., François, R., Henry, L. & Müller, K. (2018). dplyr: A grammar of data manipulation. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H. & Henry, L. (2019). tidyr: Tidy messy data. Retrieved from https://CRAN.R-project.org/package=tidyr

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

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. https://doi.org/10.1007/s13164-010-0027-y

Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass, 7*(8), 585–604. https://doi.org/10.1111/spc3.12044

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.