Journal of Experimental Social Psychology 119 (2025) 104761

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Third-party punishment, vigilante justice, or karma? Understanding the dynamics of interpersonal and cosmic justice

Cindel J.M. White [a,*], Julia W. Van de Vondervoort [b]

[a] Department of Psychology, York University, Toronto, Canada
[b] Applied Behavioral Scientist, Toronto, Canada

ARTICLE INFO

Editor: Paul Piff

Keywords:
Third party punishment
Moral judgments
Culture
Karma

ABSTRACT

People around the world both engage in both interpersonal punishment and expect supernatural punishment of wrongdoers. That is, people will impose costs and withhold benefits from transgressors, and they expect bad things to happen to transgressors more often than to good people. Evolutionary theories have proposed that both interpersonal and supernatural justice beliefs result from similar motivations, cognitive mechanisms, and cultural evolutionary processes that bind human beings into cooperative groups. To explore these ideas, three preregistered studies (N = 3430) investigated situational factors and individual differences that shape reactions to interpersonal and supernatural justice. Perceived appropriateness of both interpersonal justice and supernatural justice depended on recipients' past moral actions, with more positive impressions when antisocial actions and bad outcomes befell previously antisocial victims. However, third-party interpersonal punishment was viewed far more negatively than interpersonal reprimands or supernatural punishments, especially when the potential punisher was unaware of the victim's past transgressions. Explicit belief in karma significantly moderated perceptions of harmful outcomes not caused by human agents, but karma belief was largely unrelated to perceptions of harm caused by humans. Together, results reveal distinct factors that predict judgments about interpersonal punishment and karmic punishments, and provide insight into the distinct dynamics of interpersonal and supernatural justice.

Fundamental human moral intuitions and motives for justice contribute to the widespread belief that bad things should happen to bad people (Baumard & Boyer, 2013). Performing immoral actions that harm other people should be met with negative consequences for the transgressor, but these negative consequences can come from a variety of different sources, that vary in how legitimate, warranted, and socially acceptable they seem to third-party observers. Certain transgressions are considered illegal and can be addressed via institutional punishment. But many other transgressions are policed by revenge from the victims or by third-party punishers, who were not directly harmed themselves. In these unofficial punishments, the transgressor is themselves harmed, loses benefits, is ostracized, or has their reputation damaged by the actions of other people (Fehr & Fischbacher, 2004). However, such interpersonal punishment is risky for human punishers, who can experience retaliation from the transgressor or experience reputational damage themselves, if their "punishments" are perceived as excessive, inappropriate, or just another instance of immoral behavior (Dhaliwal, Patil, & Cushman, 2021; Nikiforakis & Mitchell, 2014; Przepiorka & Liebe, 2016; Raihani & Bshary, 2015). To mitigate these risks of interpersonal punishment, most societies also contain cultural institutions that are capable of punishing wrongdoers, such as the belief that supernatural forces will punish misdeeds with accidents, illnesses, and other misfortunes, without requiring human agents of justice.

The current experiments demonstrate how observers' evaluations of bad things that happen to bad people differ depending on whether these bad outcomes come at the hands of knowing human punishers, people who unwittingly cause harm to bad people, or supernatural forces outside of human control. We first measure moral evaluations of targets who engage in antisocial behavior across different scenarios, to test how evaluations of human third-party punishment differ from evaluations of other sorts of antisocial interpersonal behavior. We compare evaluations of actions that are more prototypical of interpersonal punishment (i.e., harm knowingly done to a previously bad victim) to others that are more ambiguous (i.e., harm coincidentally done to a previously bad victim without knowledge or intention to punish), and to unambiguously antisocial behavior (i.e., harm to previously good victims). We find some

* Corresponding author at: Department of Psychology, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada.
E-mail address: cjmwhite@yorku.ca (C.J.M. White).

https://doi.org/10.1016/j.jesp.2025.104761
Received 5 January 2025; Received in revised form 6 May 2025; Accepted 6 May 2025
Available online 12 May 2025

evidence of a general motive to see bad things happen to bad people, in that observers view it as less immoral to harm a previously bad victim than a good victim, even if the perpetrator does not know that they are punishing a bad victim, consistent with a general desire to see bad things happen to bad people.

We then compare these evaluations of interpersonal behavior to evaluations of karmic punishments that do not require human involvement, to demonstrate that any form of interpersonal hostility is evaluated more negatively, due to general norms against causing harm towards other people. We also test whether these evaluations become more positive among those participants most likely to evaluate harm to bad victims as legitimate and expected, by recruiting participants who explicitly believe in the existence of karmic forces in the universe, through which any form of misfortune towards a bad person could be perceived as fair karmic retribution for past bad behavior. Together, these studies document important variation in the social evaluations of different forms of punishment, that reveal tensions between observers' desires for transgressors to face justice and desires for people to avoid antisocial behavior.

1. Motives encouraging interpersonal punishment

Many people endorse the view that the world is fundamentally just: People encounter fair procedures and they experience good and bad outcomes commensurate to the good and bad actions they do (Hafer & Bègue, 2005; Lerner, 1980). These justice motives can manifest as retributive motives to punish transgressors by making them suffer in return for the suffering they caused (Carlsmith & Darley, 2008; Feather & Sherman, 2002). These justice motives are not only applied to situations where human beings knowingly punish other human beings, but also can be applied to find justice in situations with more opaque causal processes. Research on immanent justice beliefs (Callan, Sutton, Harvey, & Dawtry, 2014) documents that observers will also perceive causal connections between people's past bad deeds and current experiences of misfortune that were not directly caused by human punishers. For example, Canadian students were more willing to report that a man's car accident was caused by cheating on his wife with another woman than by having a neutral and appropriate interaction with another woman (Callan, Ellard, & Nicol, 2006). Observers who want to maintain their view that the world is just can interpret misfortune as fair punishments for past misdeeds, even if there were no human beings who were aware of the past misdeeds and intentionally created the misfortune to punish them.

2. Challenges of interpersonal punishment

However, when punishment of transgressors comes at the hands of other human beings, a desire to see transgressors suffer for their misdeeds may conflict with other norms and beliefs about how humans should treat one another. Instances of punishment impose a cost upon transgressors, by causing them physical or emotional suffering or by withholding benefits or social exclusion – all of which are behaviors that would be labeled antisocial if they were directed towards a morally-good person. Therefore, much interpersonal punishment behavior can have risks for their punisher's reputation, even while it yields benefits by sanctioning and deterring immoral behavior.

Theoretical and empirical models (e.g., Boyd, Gintis, Bowles, & Richerson, 2003; Fehr & Fischbacher, 2004) suggest that altruistic punishment by uninvolved third-parties is good for human sociality by deterring norm violations, and thereby fostering intragroup cooperation. Third-party punishment can also incur reputational benefits for the punisher, because it can signal that the punisher both disapproved of the norm violation and that they are willing to incur personal costs in order to sanction norm violators (Jordan & Rand, 2019; Redhead, Dhaliwal, & Cheng, 2021). However, empirical data offers mixed evidence of whether engaging in third-party punishment actually has reputational benefits, especially when experimental paradigms move from highly-constrained economic games to more ecologically-valid contexts with more options for how to respond to transgressions. For example, in economic games where third-party punishment is the only option for how participants can signal they disapprove of selfishness, choosing to punish rather than not punish has reputational benefits and can make the punisher seem more trustworthy (e.g., Barclay, 2006; dos Santos, Rankin, & Wedekind, 2013; Jordan, Hoffman, Bloom, & Rand, 2016; Jordan & Rand, 2019; Nelissen, 2008). However, real world scenarios offer many alternative options for how to respond to a transgression, such as gossiping about or calling out the transgressor's bad behavior, avoiding them, or compensating the victim, all without directly harming the transgressor. These indirect forms of norm sanctioning are more common than direct third-party punishment in naturalistic settings (Balafoutas & Nikiforakis, 2012; Balafoutas, Nikiforakis, & Rockenbach, 2014; Pedersen et al., 2020; Pedersen, McAuliffe, & McCullough, 2018), and direct forms of punishment can often generate negative reputations (Atwater, Waldman, Carey, & Cartier, 2001; Vaccarino & Kawakami, 2021). Likewise, in modified economic games where participants have the options to punish the transgressor or compensate the victim, participants are more likely to choose helping the victim (Batistoni, Barclay, & Raihani, 2022; Chavez & Bicchieri, 2013; Heffner & FeldmanHall, 2019; Leliveld, van Dijk, & van Beest, 2012; Lotz, Okimoto, Schlösser, & Fetchenhauer, 2011), and the helping response yields a much larger reputational benefit than choosing only to punish the transgressor (Dhaliwal et al., 2021; Ferguson et al., 2019; Jordan et al., 2016; Lee & Warneken, 2020; Li et al., 2018; Nikiforakis & Mitchell, 2014; Przepiorka & Liebe, 2016; Raihani & Bshary, 2015). These studies imply that even though observers view moral transgressions as often deserving of punishment, human punishers are not necessarily evaluated positively for carrying out these punishments.

Our studies test two factors that would likely moderate evaluations of third-party punishers. First, the punisher must have knowledge of the immoral behavior that they are punishing. Causing harm to someone without knowing if they are a good person or bad person will probably not be interpreted as appropriate by outside observers, even if coincidental harm to a bad person might satisfy other desires to see bad people suffer. Second, punishers may get a stronger reputational boost when punishment comes at a cost to themselves, rather than when punishment is non-costly or actually benefits themselves (Rai, 2022), because costly punishment signals that the punisher is motivated by genuine disapproval and altruistic motives, rather than being driven by self-interest in gaining personal benefits by harming bad people. Exploring how these factors interact with belief in supernatural punishment clarifies the dynamics of these systems.

3. Supernatural punishment

Belief in supernatural punishment can satisfy personal justice motives and provide socially-beneficial norm enforcement, without facing the same limitations that human punishers face. Many people believe that gods and non-theistic cosmic forces are watching human moral behavior, and intervene in the world to reward and punish human behavior, through processes that extend beyond the limitation of mundane human agents of justice (Baumard & Boyer, 2013; Norenzayan et al., 2016). Whereas human punishers may be driven by non-altruistic motives (such as personal gain by imposing costs on the transgressor) and may lack knowledge of their victim's true character, supernatural punishment of wrongdoers is likely to be perceived as more legitimate and appropriate because supernatural beings are often perceived as omniscient and benevolent (Johnson, Okun, Cohen, Sharp, & Hook, 2019; White & Norenzayan, 2022). Belief in the existence of supernatural entities, like karma, that punish immoral behavior varies substantially across individuals and systematically across members of different cultural groups. Karmic beliefs are much higher in groups with a long cultural history of religious teachings that reinforce these beliefs, such as

teachings about karma in Hindu and Buddhist cosmologies. People who hold these explicit beliefs about karmic causality should be the most likely to infer causal connections between someone's past bad behavior and their current misfortunes, even when one's current misfortunes cannot be explained through more naturalistic causal processes (e.g., Taylor, Clutterbuck, Player, Shah, & Callan, 2022; White, Norenzayan, & Schaller, 2019). However, it is unclear how karmic beliefs would impact evaluations of interpersonal punishment.

One possibility is that people who believe in karma as a general cosmic principle may view human third-party punishment as more appropriate, because human punishers may be perceived as merely the agents enacting karmically-deserved punishments. Religious and philosophical texts describe a variety of plausible pathways through which karmic retribution can come about, including by impersonal misfortune such as accidents and illnesses, intervention by powerful deities, and at the hands of other people whose karmic pathways align in legitimate and deserved punisher-punishee relationships (Bronkhorst, 2011). Someone's antisocial behavior towards a previously-bad victim may be viewed as the deserved karmic punishment, even if the human punisher is unaware of their victim's past bad behavior. Therefore, karma believers may be lenient when antisocial behavior is directed towards bad victims.

Alternatively, if people believe that karmic forces will punish wrongdoers, they might be even less inclined to support interpersonal punishment. Karma believers may expect that they can offload their punishment responsibilities to supernatural forces without needing to engage in potentially-costly behavior themselves (e.g., documented as offloading punishment to God, Laurin, Shariff, Henrich, & Kay, 2012, or karma, Zhou, Baimel, & White, 2024). Karma believers may even be more likely to believe that they should avoid antisocial punishment behaviors because this may encourage karmic punishment for themselves as recompense for causing harm (Goyal & Miller, 2023; White & Norenzayan, 2022). These effects would be consistent with cultural evolutionary theorizing that belief in supernatural punishers, like karma, are widespread because they are an efficient way to enforce cooperative norms without requiring human beings to suffer the potential costs (reputational or material) from engaging in punishment themselves (Norenzayan et al., 2016; Schloss & Murray, 2011). Our studies tested how believing in karma is associated with these different views of human punishment, if there is any association at all.

4. Overview of current studies

We conducted three studies in which participants evaluate and provide explanations for vignettes describing negative, unfortunate, or otherwise bad experiences (e.g., getting in a mild car accident, losing a phone). These scenarios systematically manipulate (a) whether the victim previously committed a good deed or bad deed (Studies 1 and 2), and (b) whether the antisocial target is knowledgeable or does not know about the victim's past behavior (Studies 1–3). Study 3 also investigated whether evaluations of harm to bad victims differed depending on whether the antisocial behavior (i.e., the potential punishment) is costly, costless, or merely involves reprimand.

We also asked participants to evaluate and provide explanations for vignettes describing misfortune that is not caused by human action (Studies 1–2), to compare how evaluations of interpersonal antisocial behavior compare to evaluations of impersonal karmic punishment. We further measured individual differences in explicit belief in karma (Studies 1–3), and tested effects across samples from varied religious backgrounds, known to systematically differ in their karmic beliefs (Studies 2–3). These comparisons reveal whether participants evaluate interpersonal punishment differently depending on whether they believe that supernatural forces operate such that bad things eventually happen to bad people. Together, these studies reveal systematic differences between perceptions of interpersonal and impersonal punishment.

STUDY 1

To begin investigating evaluations of third-party punishment, we conducted two separate preregistered studies in which participants read about target characters who responded either prosocially or antisocially to a victim who had previously behaved prosocially or antisocially to another person. Vignettes manipulated whether the target had knowledge of the victim's prior behavior, and we measured individual differences in belief in karma. Although our main predictions pertained to evaluations of antisocial behavior, we also included evaluations of prosocial behavior in this study, and did not have clear predictions about whether knowledge would matter when judging prosocial behavior. Prosocial behavior may seem more justified when in response to the recipient's past good deeds, or be perceived as more selfless when directed towards previously-antisocial recipients who don't deserve one's good will. Alternatively, good deeds may be evaluated positively regardless of who they are directed towards.

Methods were nearly identical across both studies, and similar results are present when both samples are analyzed separately, therefore below we present the results from both samples aggregated together (see Supplementary Materials for full details of the separate preregistered analyses). Preregistered methods and analysis plans for this study are available on the Open Science Framework (Sample 1A: https://osf.io/wp6gj/ and Sample 1B: https://osf.io/6r5pe/), and all data and analysis code is available at https://osf.io/85639/. For all studies we report all measures, manipulations, and data/participant exclusions, here or the Supplementary Material.

5. Methods

5.1. Participants

For two studies, participants were recruited from American workers on Amazon's Mechanical Turk (MTurk) website, in exchange for a small monetary payment. Before completing the survey, interested participants completed a short demographic questionnaire, which we used to recruit a sample that reflect the broader American population's distribution of gender, ethnicity, educational attainment, income, religious affiliation, and political affiliation (see preregistration for planned sample demographics). Those who exceeded quota requirements were excluded from participating.

For Sample A, we aimed recruit a sample of 500 participants, based on simulations (using the SIMR package in R, Green & MacLeod, 2016) of the required sample size to detect the key effects as estimated from a pilot study. Sample B aimed to recruit a sample of 700 participants, to ensure we had at least 90 % power to detect the main effect of the actor's behavior and recipient's behavior, as estimated from Sample A. We also excluded participants from the final sample if they failed either of the two attention check questions placed within the survey. New participants were recruited to replace any participants excluded based on these criteria, until the desired sample size had been reached.

The final sample across both data collections included 1222 participants from the United States (Sample A $n = 506$, Sample B $n = 716$), including 59 % women, 41 % men; 78 % White, 8 % Black, 6 % Hispanic/Latino, 4 % Asian, 4 % multiple or other ethnicities; 69 % Christian, 4 % Hindu, Buddhist, or Sikh, 21 % non-religious, atheist or agnostic, 6 % other religious affiliations; aged 18–74 ($M = 38$ years). This sample size provided 80 % power (alpha = 0.05) to detect small effects ($d = 0.08$ or greater in paired-samples t-test between conditions and correlations of $r = 0.08$ or greater).

5.2. Materials and procedure

All key measures (described below) were completed as part of a larger survey that included evaluation of a variety of vignettes, questionnaires assessing participants' explicit supernatural beliefs, beliefs about justice, and religious demographic measures. Exact wording of all materials is available in the preregistration documents.

5.2.1. Interpersonal behavior

5.2.1.1. Vignettes describing behavior. Participants read 4 vignettes in which an individual (the victim character) performs a prosocial or antisocial action, such as returning someone's wallet or stealing someone's wallet. Later on, this character becomes the victim of the focal target character's prosocial or antisocial action, such that the target either helped the victim character or hindered them (e.g., returned their phone or stole their phone). The focal target character was always described as a third party (i.e., not the recipient of the victim's past behavior). The content of each vignette came from one of five possible scenarios, and included transgressions such as parking in a handicap spot without a permit, cutting someone off in traffic, stealing a wallet/phone, refusing to help someone carrying groceries/books, refusing to lend small amount of money, and breaking a promise to help a friend (see Supplementary Materials and preregistration documents for all scenarios). Each vignette included different characters, tailored to manipulate the relevant variables (see Table 1 for an example). The combination of scenario types and manipulated variables were counterbalanced across participants, and the order in which vignettes are viewed were randomized across participants.

Participants were randomly assigned to read about vignettes where the ***target's actions*** were either prosocial or were antisocial (e.g., helping vs. refusing to help), manipulated between-subjects. Vignettes manipulated within-subjects whether the ***victim's past behavior*** had been prosocial or antisocial (e.g., sharing vs. refusing to share), and whether the target was ***knowledgeable*** or lacked knowledge of the victim's past behavior. As a result, participants read two valent-congruent outcomes in which the target's behavior could be interpreted as reward or punishment for the victim's past behavior, such as stealing a cell phone from someone who previously stole a stranger's wallet. They also read two vignettes with the target's behavior incongruent with the past behavior, such as stealing a cell phone from someone who previously found and returned a stranger's wallet. We chose to manipulate victim's past behavior and target knowledge within-subjects in order to increase statistical power to detect potentially-small effects (given pilot test results showing small effects of these variables; a similar pattern of results is found if only the first trial participants viewed is analyzed in between-subjects comparisons, see Supplementary Materials). We manipulated target action between-subjects because we anticipated a large difference in evaluations and suspected somewhat different psychological principles could be relevant when evaluating antisocial and prosocial actions.

In Sample A, knowledge was manipulated by the target being present to see the victim's past behavior or being absent and unable to witness the behavior. To reinforce this knowledge manipulation, Sample B also included descriptions of the target's emotional reaction to witnessing the victim's behavior (e.g., being mad at antisocial behavior and pleased by prosocial behavior).

After reading each vignette, participants provided a written description of "Why did [Target] do that?" These responses were coded by naïve research assistants for whether they refer to aspects of the target's and victim's character traits, and full results of these explanations are available in the Supplementary Materials.

Table 1
Sample vignettes, Study 1.

Stage 1: Victim's past behavior	
Prosocial Victim	In a cafe, several people are waiting in line to buy coffee. A man in line accidentally drops his wallet on the floor and goes to exit the cafe without noticing. A woman sees the wallet on the floor, picks it up, and returns it to the man who had dropped it.
Antisocial Victim	In a cafe, several people are waiting in line to buy coffee. A man accidentally drops his wallet on the floor and goes to exit the cafe without noticing. A woman sees the wallet on the floor, picks it up, and puts it into her purse.
Stage 2: Target's behavior	
Prosocial Target - Known	A second man, named James, is also waiting in line and sees the woman take [return] the wallet. James is upset that the woman took the wallet. [James is glad that the woman returned the wallet.] When the woman is ordering coffee, she puts her cell phone onto the counter. When she leaves the counter, she forgets to pick up her phone and walks away. James takes the phone and returns it to her before she leaves the cafe.
Prosocial Target - Unknown	After the wallet was taken [returned], a second man, named James, enters the cafe and joins the line. When the woman is ordering coffee, she puts her cell phone onto the counter. When she leaves the counter, she forgets to pick up her phone and walks away. James takes the phone and returns it to her before she leaves the cafe.
Antisocial Target - Known	A second man, named James, is also waiting in line and sees the woman take [return] the wallet. James is upset that the woman took the wallet. [James is glad that the woman returned the wallet.] While the woman is ordering coffee, she puts her cell phone onto the counter. When she leaves the counter, she forgets to pick up her phone and walks away. James takes the phone and puts it into his pocket.
Antisocial Target - Unknown	After the wallet was taken [returned], a second man, named James, enters the cafe and joins the line. While the woman is ordering coffee, she puts her cell phone onto the counter. When she leaves the counter, she forgets to pick up her phone and walks away. James takes the phone and puts it into his pocket.
Impersonal	Later, when the woman is ordering coffee, her phone slips out of her pocket onto the floor. She doesn't notice and she exits the cafe without her phone.

Note. Vignettes were comprised of each possible combination of victim's past behavior (Stage 1) and the target's behavior or impersonal bad experience (Stage 2). In this set of vignettes, participants were asked to provide moral evaluations of James' behavior in these scenarios.

5.2.1.2. Evaluations of interpersonal behavior. Four closed-ended questions assessed moral evaluations of the target's actions, including whether they are a bad person or a good person (1 = *extremely bad,* 7 = *extremely good*), should be blamed or praised (1 = *blamed a lot*, 7 = *praised a lot*), whether their behavior was unacceptable or acceptable (1 = *completely unacceptable*, 7 = *completely acceptable*), and bad or good (1 = *extremely bad*, 7 = *extremely good*). Items were averaged into a composite score of moral judgment ($\alpha = 0.97$), with scores <4 indicating negative moral evaluations and > 4 indicating positive moral evaluations.

5.2.2. Impersonal events

5.2.2.1. Vignette describing the event. After evaluating four trials of interpersonal behavior, participants evaluated one event in which a negative outcome happened without the intervention of any human agent. These vignetted also manipulated whether the ***victim's past behavior*** was prosocial or antisocial, before experiencing the negative outcome. These events contained a similar structure and outcome as the interpersonal vignettes, but the bad outcome merely happened by chance, rather than through human agency. For example, rather than someone sealing their wallet, in this condition the victim merely loses her phone when it falls out of her pocket without her noticing. The content of this vignette came from whichever of the five scenarios was not viewed previously.

5.2.2.2. Open-ended explanations of the event. After reading the vignette, participants provided an open-ended written description of why the event occurred. These responses were coded by naïve research assistants for whether their responses mentioned (a) the target's moral character traits (e.g., they are mean or kind), (b) the target's recklessness or carelessness, (c) whether it was an accident, (d) physical forces (e.g., gravity, a gust of wind), or (e) supernatural forces, including

specifically whether they referred to karma as a cause of the outcome.

5.2.2.3. Evaluations of event. Participants also reported in closed-ended questions whether this negative outcome was acceptable (1 = *completely unacceptable*, 7 = *completely acceptable*) and bad or good (1 = *extremely bad*, 7 = *extremely good*), which were averaged into a composite measure of ***evaluations*** of the impersonal event. Participants were also asked directly whether the victim was ***responsible*** for the bad outcome and whether the bad outcome was ***caused*** by the victim's previous behavior (1 = *not at all*, 7 = *completely*).

5.2.3. Supernatural beliefs and demographics

After evaluating the key vignettes, participants completed a set of other supernatural beliefs and demographic questionnaires. These included a 16-item measure of belief in karma (e.g., "Karma is a force that influences the events that happen in my life," "When people are met with misfortune, they have brought it upon themselves by previous behaviour in their life", α = 0.93, White et al., 2019).

6. Results

6.1. General analysis strategy

All analyses were performed as multi-level models that include random intercepts for each participant (to account for the repeated-measures nature of the data) and crossed random effects based on the scenario content of the vignette (to account for variation in stimuli). The final model predicted judgments from all interactions between dummy-coded condition indicators, in a 2 (Valence of Action: 0 = Antisocial, 1 = Prosocial) × 2 (Valence of victim's previous behavior: 0 = Antisocial, 1 = Prosocial) × 2 (Knowledge: 0 = Unknown, 1 = Known) fully factorial design. We report below analyses collapsed across data from both samples, to increase power, but report separate results for each sample in the Supplementary Materials.

6.2. Evaluations of interpersonal behavior

The first analysis tested how evaluations of the actions varied across conditions, depicted in Fig. 1 and Table 2. As predicted, there was a main effect of the target's action, such that prosocial actions were consistently evaluated positively and antisocial actions evaluated negatively, $b > 3.16$, $ps < 0.001$. Further, in the case of antisocial actions, there was a main effect of the victim's past behavior: The target's antisocial behavior was evaluated more negatively when directed towards a previously-prosocial recipient, and more positively when directed towards a previously-antisocial recipient, regardless of whether the target

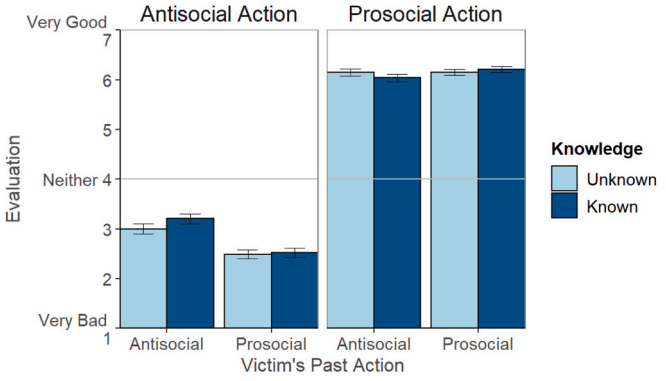

Fig. 1. Mean evaluations (with 95 % confidence intervals) of antisocial and prosocial actions, when directed towards victims who were previously antisocial or prosocial themselves, according to knowledge of the victim's past behavior, Study 1.

Table 2
Results of full multilevel regression model predicting evaluations of interpersonal behavior from the valence of the target's action (0 = Antisocial, 1 = Prosocial), valence of victim's previous behavior (0 = Antisocial, 1 = Prosocial), and the target's knowledge of the victim's behavior (0 = Unknown, 1 = Known), Study 1 combined across both samples.

	Evaluation		
Predictors	*b*	*95 % CI*	*p*
Intercept	3.00	2.67, 3.33	<0.001
Target's action valence	3.16	3.04, 3.27	<0.001
Victim's action valence	−0.50	−0.60, −0.40	<0.001
Knowledge	0.21	0.11, 0.32	<0.001
Target * Victim	0.50	0.35, 0.64	<0.001
Target * Knowledge	−0.33	−0.47, −0.18	<0.001
Victim * Knowledge	−0.18	−0.32, −0.03	0.017
Target * Victim * Knowledge	0.36	0.15, 0.56	0.001
Random Effects			
σ^2	0.84		
$\tau_{00\ Participant}$	0.19		
$\tau_{00\ vignette}$	0.13		
Marginal R^2 / Conditional R^2	0.709 / 0.790		

was knowledgeable, $b = -0.68$ [−0.78, −0.58], or lacked knowledge of the victim's past behavior, $b = -0.50$ [−0.60, −0.40], $ps < 0.001$. There was also a three-way interaction between target's behavior, victim's behavior, and knowledge, $b = 0.36$ [0.15, 0.56], $p = .001$, such that knowledge only had an effect when antisocial action was directed towards a previously-antisocial victim. In this case, knowing about the past antisocial behavior predicted more positive evaluations, $b = 0.21$ [0.11, 0.32], $p < .001$. Knowledge also led to slightly more negative evaluations of prosocial behavior towards a previously-antisocial recipient, $b = -0.11$ [−0.22, −0.01], $p = .028$. Knowing (vs. not knowing) about a victim's previous prosocial behavior had no effect on evaluations, $ps > 0.20$.

The pattern of results was similar when we conducted separate analyses for evaluations of the action itself and evaluations of the target person who performed the action. Results were also similar when we separately analyzed Sample A and Sample B (as initially preregistered, see Supplementary Materials), except that the three-way interaction between target behavior, victim behavior, and knowledge, and the two-way victim behavior by knowledge interaction, were only statistically significant in Sample A and in the combined analysis. These interactions were not significant in Sample B (although effects are in the same direction), but the general pattern of evaluations was similar, and in both samples there was a significant effect of knowledge leading to more positive evaluations when antisocial actions were done to antisocial victims (Sample A: $b = 0.19$ [0.03, 0.35], $p = .023$, Sample B: $b = 0.22$ [0.09, 0.36], $p = .001$).

6.3. Evaluations of impersonal events

A separate set of analyses investigated how evaluations of impersonal events (that lack human causes of misfortune) compared to evaluations of similar events caused by humans. As predicted, impersonal events were evaluated more positively, $m = 4.16$, 95 % CI [4.07, 4.26], than analogous events caused by humans either knowingly, $m = 3.20$, 95 % CI [3.10, 3.30], $b = -0.96$ [−1.10, −0.82], or unknowingly, $m = 2.99$ [2.89, 3.10], $b = -1.18$ [−1.31, −1.04], $ps < 0.001$.

Negative impersonal events were also evaluated differently when they happened to a previously prosocial versus previously antisocial victim (see Fig. 2), with impersonal misfortune directed towards antisocial victims judged to be more causally related to their past behavior, $t(1112.4) = 3.07$, $p = .002$, $d = 0.18$ [0.06, 0.03], something the victims are more responsible for, $t(1114.2) = 9.74$, $p < .001$, $d = 0.58$ [0.45, 0.70]], than when the same events befall prosocial victims. Misfortune

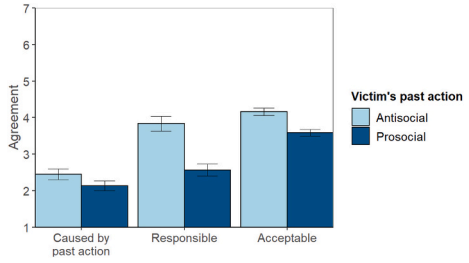

Fig. 2. Perceptions of impersonal misfortune based on the victim's past behavior, Study 1.

for antisocial victims was subsequently rated as more acceptable, $t(1219.6) = 8.57$, $p < .001$, $d = 0.49$ [0.38, 0.60], close to the neutral scale midpoint of neither acceptable nor unacceptable, whereas the same events befalling prosocial victims were rated as slightly unacceptable. This shows that victim's past behavior impacts evaluations of both human-caused misfortune (i.e., antisocial behavior) as well as incidental misfortune that lacks human agency.

6.4. Explanations for impersonal events

Open-ended explanations for impersonal events also showed variation based on the victim's past behavior (see Fig. 3). As predicted, participants spontaneously mentioned "karma" in their explanations more often for events when misfortune befell a previously-antisocial victim (i.e., in cases of moral congruity) than a previously-prosocial victim, $OR = 0.24$ [0.16, 0.35], $p < .001$, but these conditions did not significantly differ in the likelihood of explicit referring to the victim's moral character traits, $OR = 0.87$ [0.60, 1.25], $p = .45$, or non-moral traits such as recklessness, $OR = 1.19$ [0.87, 1.62], $p = .29$. Participants were also significantly more likely to refer to mundane physical causes of misfortune when they befell a prosocial rather than antisocial victim, $OR = 1.58$ [1.25, 2.00], $p < .001$, such as saying that "a gust of wind blew it" or "her pocket wasn't deep enough to secure the phone adequately".

6.5. Belief in karma

Our final analysis explored whether individual differences in belief in karma predicted evaluations of events. As depicted in Table 3, belief in karma was strongly positively correlated with viewing impersonal events as caused by victim's past behavior, consistent with the karmic logic that bad things befall those who commit bad deeds. However, belief in karma was not significantly associated with positive evaluations of impersonal misfortune, or with evaluations of interpersonal antisocial behavior unknowingly directed towards previously antisocial victims. Karma predicted slightly more positive evaluations of antisocial behavior when *knowingly* directed towards an antisocial actor, and more positive evaluations of prosocial behavior. However, the condemnation

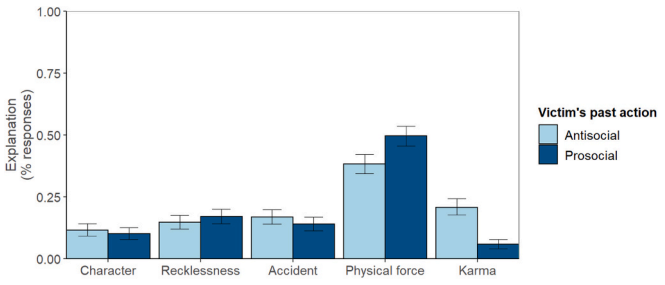

Fig. 3. Presence of references to different explanations for impersonal misfortune, Study 1. Responses were also coded for mentions of God, Fate, and Luck, but these were mentioned at very low rates (<2 %) across both conditions.

of targets who unknowingly harm others was present regardless of belief in karma. In other words, believing in karma does not lead people to license antisocial behavior, even when they might infer that the victims may deserve it.

STUDY 2

Study 1 found evidence that evaluations of interpersonal third-party antisocial behavior are shaped by whether the outcome is congruent with the victim's past behavior – with antisocial behavior being more appropriate when the victim was previously antisocial – and whether the antisocial target is knowledgeable of the past transgression – with knowledgeable punishment being more acceptable. However, there remains a general tendency to disapprove of antisocial behavior at the hands of human third-parties. Negative outcomes are viewed as much more appropriate when they merely occur by accident, without the intervention of human beings, although even these ratings have an average around scale midpoint (neither good nor bad), rather than being judged overtly positive by most participants. Participants' explicit belief in karma – as a supernatural force that should ensure that bad things will happen to bad people, even without the intervention of human punishers – did not predict greater approval of interpersonal antisocial behavior (even when it could most appropriately be considered as justified punishment). However, these first samples were drawn from the general US population, who were predominantly Christian or non-religious, and therefore may not contain sufficient numbers of participants who strongly endorse and regularly think about karma, in order to have any noticeable impact on judgments. Therefore, Studies 2 and 3 targeted samples that equally represent members of karmic religious traditions (e.g., Hinduism and Buddhism) alongside Christian and Atheist samples, to test for variation in judgments across members of each religious denomination. We also dropped the prosocial target from the following studies, to focus more narrowly on evaluations of third-party punishment.

7. Methods

Prior to completing this study all methods and analysis plans were preregistered at https://osf.io/wtc58/.

7.1. Participants

Participants were undergraduates recruited from Human Subjects Pool at a large North American university, in return for partial course credit. We aimed to recruit participants with different levels of belief in karma, and due to the close connection between belief in karma and different religious traditions (White & Norenzayan, 2019) we recruited participants from three different groups based on their self-reported religious affiliations in a prescreening survey. We recruited one group of Hindus, Buddhists, and Sikhs (i.e., individuals associated with religions that endorse karmic beliefs), one group of Christians (i.e., individuals associated with a religion, but one that does not traditionally include karmic beliefs), and one group of Atheists (i.e., individuals unassociated with any religion or supernatural beliefs). We aimed to recruit a sample of 350 participants from each group, comparable to the sample size in the antisocial target conditions in Study 1 Sample B. We also excluded any participants who failed either of two attention checks placed within the survey ($n = 68$).

The final sample consisted of 392 Christians, 398 atheists, and 349 members of karmic religions (73 Hindus, 132 Buddhists, 116 Sikhs, 28 identified as non-religious within the survey). Overall this sample included 76 % women, 24 % men; 23 % White, 2 % Black, 39 % East Asian, 23 % South Asian, 13 % multiple or other ethnicities; aged 18–53 ($M = 20$ years). This sample size provided 80 % power (alpha = 0.05) to detect small effects between conditions in the whole sample ($d = 0.08$ or greater in paired-samples t-test between conditions and correlations of $r = 0.08$ or greater) and within each religious group when analyzed separately ($d = 0.15$ or greater in paired-samples t-test between

Table 3
Bivariate correlations between belief in karma, evaluations of impersonal negative events, and evaluations of interpersonal antisocial behavior (lower diagonal) and prosocial behavior (upper diagonal), Study 1. *$p < .05$, **$p < .01$, **$p < .001$.

	Belief in Karma	Impersonal: Caused by past	Impersonal: Responsible	Impersonal: Eval.	Good, known: Eval.	Bad, known: Eval:	Good, unknown: Eval.	Bad, unknown: Eval.
Belief in Karma		0.05	0.07	−0.03	0.12**	0.05	0.16***	0.13***
Impersonal: Caused by past	0.46***		0.50***	0.06	−0.16**	−0.27***	0.10	0.03
Impersonal: Responsible	0.07	0.32***		−0.02	−0.14*	−0.23***	0.06	−0.04
Impersonal: Evaluation	0.06	0.17***	0.01		−0.09	−0.08	−0.09	−0.12**
Good victim, known: Evaluation	−0.02	−0.06	−0.15**	0.01		0.43***	0.58***	0.52***
Bad victim, known: Evaluation	0.10*	0.17***	0.29***	0.09	−0.03		0.37***	0.37***
Good victim, unknown: Evaluation	−0.01	−0.01	0.00	0.09	0.11**	−0.09*		0.52***
Bad victim, unknown: Evaluation	0.06	−0.02	−0.21***	0.02	0.07	−0.01	−0.04	

conditions and correlations of $r = 0.15$ or greater).

7.2. Materials and procedure

7.2.1. Evaluations of interpersonal behavior

The vignettes and focal measures analyzed below were identical to those in the antisocial target conditions of Study 1 (Sample B). Participants read 4 vignettes in which the target character performs an antisocial action. These vignettes (randomly drawn from a pool of five scenarios) manipulated within-subjects whether the victim was previously engaged in antisocial or prosocial behavior, and whether the target was knowledgeable of the victim's past behavior.

After reading the description, participant provided an open-ended explanation for why the target performed the antisocial action, which research assistants coded for reference to the victim's character, the target's character, and whether the explanation implied knowledge of the event. Then, participants provided the 4-item moral evaluation of the action ($\alpha = 0.90$), which was coded with lower values indicating more negative evaluations and higher values indicating positive evaluations.

7.2.2. Evaluations of impersonal events

Participants also read one description of a negative event that occurred without the intervention of any human being. Following the methods of Study 1, participants provided an open-ended explanation for why the negative event occurred, which was coded by research assistants. Participants then evaluated whether the event was acceptable and good, whether the victim was responsible for the event, and whether the event was caused by the victim's past behavior.

After evaluating the interpersonal and impersonal events, participants completed a larger set of measures of supernatural beliefs and social judgments (see preregistration for full details), including levels of explicit belief in karma and their demographic characteristics.

8. Results

8.1. Evaluations of interpersonal behavior

Using the same multilevel modeling strategy as in Study 1, we predicted moral evaluations from valence of victim's previous behavior, knowledge of the victim's previous transgressions, and the interaction between them, and random intercepts by participant and vignette scenario, separately for each of the three religious groups (Christians, Atheists, and participants from karmic religions; supplementary analyses testing all groups together did not find any statistically significant interactions with religious group, implying similar patterns across groups, although the present dataset may be underpowered to fully test this claim). Results are displayed in Fig. 4 and Table 4. As predicted, in all three religious groups there was a main effect of the victim's past behavior, such that antisocial actions are evaluated more negatively when directed towards a previously-prosocial recipient. There was also the predicted interaction between knowledge and valence of the victim's past behavior, among Karmic and Atheist participants, such that knowledge of the victim's past behavior led to more positive evaluations when the victim was previously antisocial, but knowledge about a victim's past good behavior did not significantly affect judgments. Christian participants showed the same pattern, although the interaction between victim's behavior and knowledge was not significant.

8.2. Evaluations of impersonal events

We next used a multi-level model to predict how evaluations of impersonal/karmic events (that lack human causes of misfortune) compare to evaluations of human-caused events. As predicted, and displayed in Fig. 5 (see Table S15 for full results), karmic events were evaluated as much more positive than analogous bad experiences caused by humans (either knowingly or unknowingly), across members of each religious group. Impersonal events were rated as neutral on average, rather than overtly negative (as human antisocial behavior was) or positive (as might be expected from believers applying a strong karmic framework to this situation). This consistency across groups in vignette judgments existed alongside substantial cultural differences in explicit belief in karma, which was much higher among Hindu/Buddhist/Sikh participants, $m = 3.33$ ($sd = 0.71$), than Christian, $m = 2.56$ ($sd = 0.77$), t

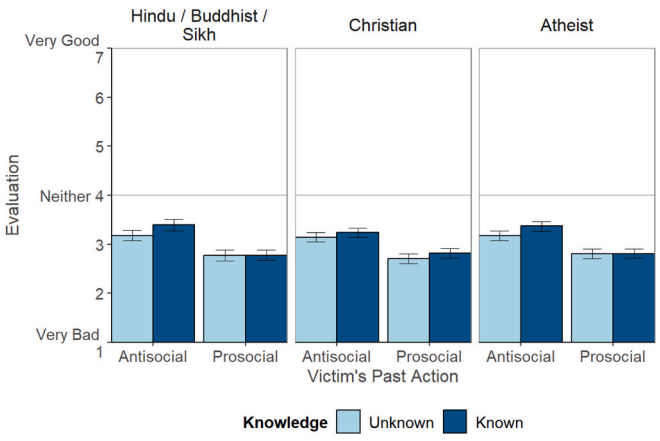

Fig. 4. Mean evaluations (with 95 % confidence intervals) of antisocial actions when directed towards victims who were previously antisocial or prosocial themselves, according to knowledge of the victim's past behavior, Study 2.

Table 4
Results of full multilevel regression model predicting evaluations of interpersonal behavior from valence of victim's previous behavior (0 = Antisocial, 1 = Prosocial), and the target's knowledge of the victim's behavior (0 = Unknown, 1 = Known), Study 2, separately for each religious group.

	Hindu / Buddhist / Sikh			Christian			Atheist		
Predictors	*b*	*CI*	*p*	*b*	*CI*	*p*	*b*	*CI*	*p*
Intercept	3.20	2.64, 3.76	<0.001	3.12	2.61, 3.63	<0.001	3.17	2.69, 3.65	<0.001
Victim valence	−0.41	−0.54, −0.28	<0.001	−0.39	−0.50, −0.28	<0.001	−0.36	−0.48, −0.25	<0.001
Knowledge	0.19	0.06, 0.32	0.003	0.11	0.00, 0.23	0.046	0.19	0.07, 0.31	0.001
Valence * Knowledge	−0.23	−0.41, −0.05	0.014	−0.04	−0.20, 0.12	0.647	−0.19	−0.35, −0.02	0.026
Random Effects									
σ^2	0.75			0.64			0.71		
$\tau_{00\ \text{Participant ID}}$	0.07			0.09			0.03		
$\tau_{00\ \text{vignette}}$	0.40			0.33			0.29		
Marginal R^2 / Conditional R^2	0.057 / 0.422			0.040 / 0.423			0.052 / 0.347		

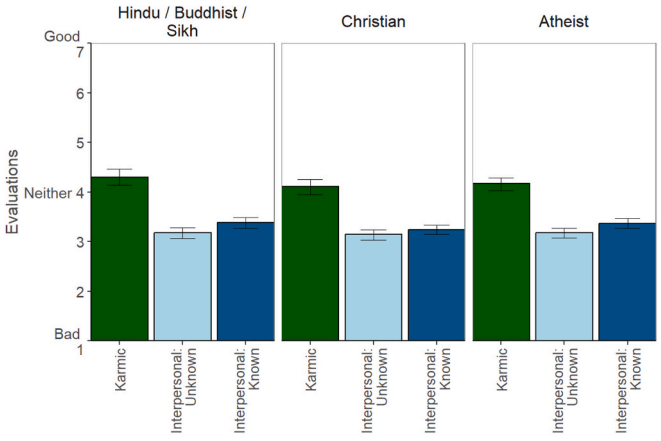

Fig. 5. Mean evaluations (with 95 % confidence intervals) of bad outcomes caused by impersonal karmic sources and interpersonal antisocial behavior, Study 2.

(738) = 14.14, $p < .001$, or Atheist participants, $m = 2.19$ ($sd = 0.86$), t (742) = 19.81, $p < .001$, and higher among Christians than Atheists, t (782) = 6.32, $p < .001$.

We next examined how evaluations of impersonal experiences differed based on the victim's past behavior, first separately within each religious group. As predicted, within each religious group, bad experiences were evaluated as more positively, the victim was more responsible, and their circumstances were caused by past behavior more when the victim had previously performed a bad action rather than a good action (see Table S16 and Fig. S4).

8.3. Belief in karma: Evaluations of impersonal harm

We then examined all data together in a regression model that also included participants' level of belief in karma and the interaction between karma belief and the victim effect. This size of these differences between antisocial victims and prosocial victims were significantly moderated by participants' strength of belief in karma (Table 5). As depicted in Fig. 6, the greatest differences between judgments of impersonal events were among participants who explicitly believed in karma, while participants below the scale midpoint in belief did not judge antisocial victims differently from prosocial victims. Also, explicit belief in karma only predicted increased responsibility judgments when bad outcomes befell an antisocial victim; belief in karma was unassociated with responsibility judgments when outcomes befell a prosocial victim. Likewise, in participants' open-ended explanations for why the karma-caused event happened, karma was most likely to be explicitly mentioned when the bad experienced happened to an antisocial victim, and rarely mentioned for prosocial victims, $OR = 0.05$ [0.02, 0.11], $p <$ .001 (see Supplementary Materials for other coding of open-ended responses).

8.4. Belief in karma: Evaluations of interpersonal harm

Our final analyses explored whether individual differences in belief in karma predicted evaluations of interpersonal events, in addition to impersonal events. In an overall model, there was no main effect of belief in karma, $b = 0.03$ [−0.02, 0.08], $p = .19$, nor any significant interactions between belief in karma and other predictors (knowledge and victim's past behavior), $ps > 0.15$. Bivariate correlations between belief in karma and evaluations of different forms of antisocial behavior only showed one significant association, with less positive evaluations when good victims were harmed, $r = -0.08$ [−0.14, −0.03] (see Supplementary Table S19). This is different from the type of evaluation that was significantly correlated with belief in karma in Study 1. This suggests that the association between belief in karma and evaluations of interpersonal antisocial behavior and punishment is weak and inconsistent.

STUDY 3

One consistent finding in Studies 1–2 was that all forms of antisocial behavior were evaluated quite negatively. Even when directed knowingly towards victims who deserve it, these factors only led participants to perceive the "punishments" as *less negative* than antisocial behavior directed towards undeserving victims, but never as positive. We therefore designed a new set of scenarios that incorporated additional factors

Table 5
Results of multilevel regression model judgments about impersonal events from participants' level of belief in karma, valence of victim's bad behavior, and their interaction, Study 2, aggregated across data from all religious groups.

	Caused by past action			Responsible			Evaluation of impersonal event		
Predictors	*b*	*95 % CI*	*p*	*b*	*95 % CI*	*p*	*b*	*95 % CI*	*p*
Intercept	2.31	2.19, 2.42	<0.001	3.32	3.15, 3.49	<0.001	4.19	4.11, 4.28	<0.001
Belief in karma	0.80	0.69, 0.92	<0.001	0.38	0.22, 0.55	<0.001	0.08	0.00, 0.16	0.046
Victim valence	−0.69	−0.85, −0.52	<0.001	−0.63	−0.87, −0.38	<0.001	−0.40	−0.52, −0.28	<0.001
Karma * Victim valence	−0.64	−0.81, −0.48	<0.001	−0.40	−0.65, −0.16	0.001	−0.18	−0.29, −0.06	0.004
R^2 / R^2 adjusted	0.195 / 0.193			0.040 / 0.037			0.044 / 0.041		

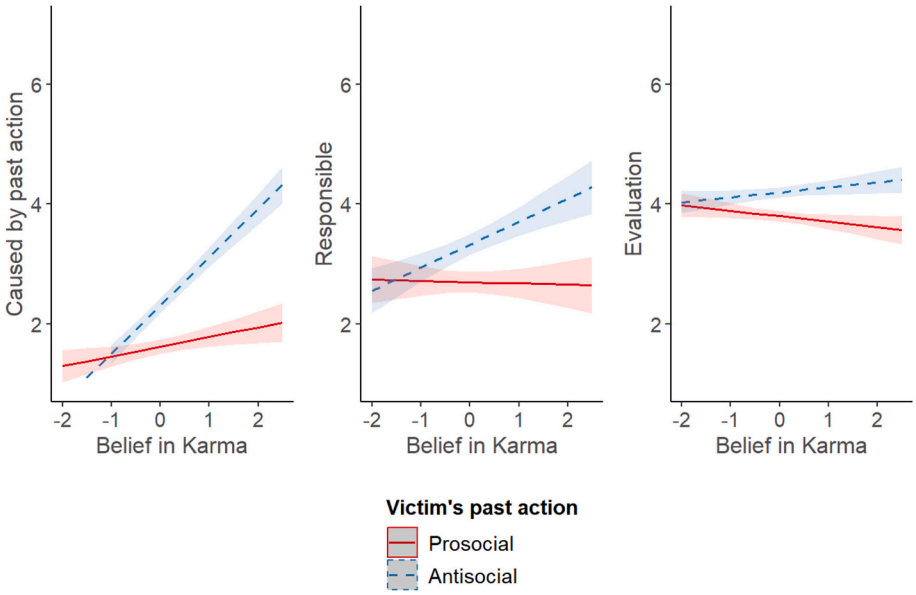

Fig. 6. Interaction between participants' level of belief in karma and victim's past behavior, predicting judgments that the impersonal event was (a) caused by past actions, (b) the victim's responsibility, and (c) positively evaluated.

that previous literature has identified as relevant to judgments of punishment: whether the punishments are costly or non-costly to the punisher, and whether they merely involve reprimanding the victim for their behavior rather than imposing costly penalties on the victim. Punishment knowingly directed towards an antisocial victim may be viewed as more virtuous when it comes with costs for the punisher, rather than when punishments are costless or even generate benefits for the punisher. We investigated whether evaluations of antisocial behavior differ according to these factors and whether belief in karma predicts evaluations of these new scenarios.

9. Methods

Prior to completing this study all methods and analysis plans were preregistered at https://osf.io/u37mf/.

9.1. Participants

As in Study 2, undergraduate students were recruited from a large North American university, in return for partial course credit, from three groups based on religious denomination: Hindus/Buddhists/Sikhs, Christians, and Atheists. We aimed to recruit a sample of 350 participants from each group, after excluding participants who failed attention checks placed within the survey ($n = 47$). As preregistered, we stopped data collection when either 350 participants had completed the survey or after an entire year had passed since initiation of data collection (data was collected between Fall 2020 and Summer 2021).

The final sample consisted of 412 Christians, 368 atheists, and 289 members of karmic religions (87 Hindus, 88 Buddhists, 84 Sikhs, 30 identified as non-religious within the survey). Overall this sample included 75 % women, 25 % men; 22 % White, 39 % East Asian, 22 % South Asian, 17 % multiple or other ethnicities; aged 18–49 ($M = 20$ years). As in Study 2, these religious groups differed significantly in their belief in karma, with belief being higher among members of karmic religions, $m = 3.41$ ($sd = 0.75$), than among Christians, $m = 2.62$ ($sd = 0.74$), or Atheists, $m = 2.41$ (0.83), $ps < 0.001$. This sample size provided 80 % power (alpha = 0.05) to detect small effects between conditions in the whole sample ($d = 0.09$ or greater in paired-samples t-test between conditions and correlations of $r = 0.09$ or greater) and within each religious group when analyzed separately ($d = 0.17$ or greater in paired-samples t-test between conditions and correlations of $r = 0.16$ or greater).

9.2. Materials and procedure

9.2.1. Evaluations of interpersonal behavior

Participants read 5 vignettes that all described antisocial behavior directed towards a previously-antisocial victim. These vignettes included new scenarios (listed in full in the Supplementary Materials Tables S19 – S23), but had a similar structure where an initial person commits a transgression, and this transgressor becomes the victim of later antisocial behavior by the target "punisher" who was unhelpful or inconsiderate (e.g., not giving someone a ride, not warning someone about a danger that caused an injury, letting someone get stuck in a storage closet, allowing someone's notes to get ruined). These vignettes manipulated whether the antisocial target was ***knowledgeable*** or lacked knowledge of the victim's past immoral behavior (i.e., whether they were present to witness the victim's behavior or absent) and also manipulated whether the antisocial behavior was ***costly*** or costless to the target. For example, for a punitive scenario that involved the target refusing the help a colleague who was stuck in a storage closet, the costly scenario involves the victim telling other employees about her behavior, which damages her reputation, while the costless condition lacked this negative consequence.

A fifth vignette depicted an analogous scenario where the target merely ***reprimands*** the other for their past antisocial behavior. In the reprimand scenarios, the target characters says that their previous bad deed was not okay and warns them of potential negative consequences for antisocial behavior, rather than the penalties imposed by the antisocial behavior in the four other scenarios (e.g., "The next day, Jennifer confronts Charlie and tells him that not helping people who need help in the elevator is not OK. Jennifer warns him that he'll be fired if she sees him acting that way again."). Each scenario depicted a different scenario with different characters, and the matching of scenario type of experimental condition was counterbalanced across participants. Participants then completed the same open-ended explanation question and closed-ended moral evaluation questions (character evaluation, blame, acceptability, and wrongness, $\alpha = 0.93$) used in Studies 1–2.

This study did not include an evaluation of impersonal events, but the interpersonal vignettes were followed by a longer battery of

supernatural belief and social judgments questions, including explicit belief in karma and belief in a just world.

10. Results

10.1. Evaluations of interpersonal behavior

We first focused on analyzing the four antisocial behavior conditions, while excluding the "reprimand" condition where the target does not actually impose any costs on the transgressor. We used a multilevel regression model to predict moral evaluations from the knowledge and costliness of the antisocial action, the interaction between them, and random intercepts by participant and vignette scenario, both in the overall sample and separately for each religious group (full results for each religious group separately are available in the Supplementary Materials). In the overall sample, there was a main effect of knowledge, in that antisocial actions with knowledge of past behavior were evaluated more positively than without knowledge, $b = 0.29$ [0.22, 0.37], $p < .001$. However, there was no significant effect of cost, $b = -0.01$ [−0.09, 0.06], $p = .70$, nor any cost by knowledge interaction, $b = 0.03$ [−0.07, 0.14], $p = .54$.

We then conducted additional regression models that compared all four antisocial behavior conditions to the reprimand condition (dummy coded). As depicted in Fig. 7, reprimand was evaluated more positively than all forms of third-party antisocial behavior: Low-cost knowledge, $b = -1.90$ [−1.98, −1.82], high-cost knowledge, $b = -1.88$ [−1.96, −1.80], low-cost no knowledge, $b = -2.19$ [−2.27, −2.11], and high-cost no knowledge: $b = -2.20$ [−2.28, −2.12], $ps < 0.001$. Reprimand was generally rated as moderately good, indicating that merely condemning and threatening negative consequences for a transgression has much better reputational benefits than actually imposing costly punishment for the transgression.

A final model also included participants' religious group in the model. In addition to the main effect of knowledge, $b = 0.26$ [0.12, 0.40], $p < .001$, there was also a main effect of Christians evaluating antisocial behavior more negatively than members of karmic religions, $b = -0.19$ [−0.34, −0.04], $p = .012$, although not significantly more negative than atheists, $b = -0.10$ [−0.24, 0.04], $p = .15$. No other main effects of interactions were significant (see Supplementary Table S27), $ps > 0.17$, indicating that the effects of knowledge and cost did not significantly differ across members of different religious groups. Evaluations of reprimand also did not significantly differ across members of different religious groups, $ps > 0.56$.

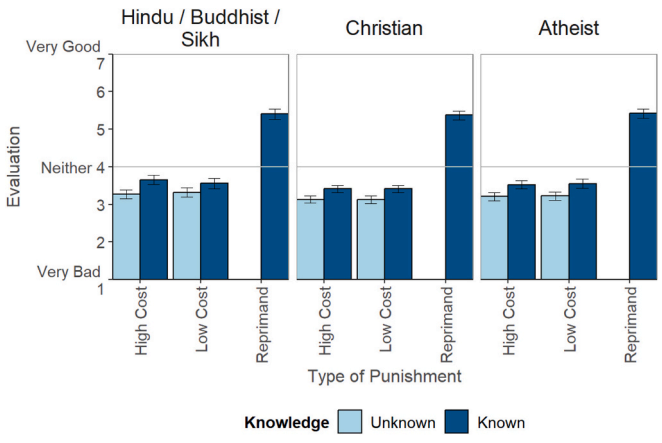

Fig. 7. Mean evaluations (with 95 % confidence intervals) of high cost and low cost third-party antisocial behavior, and reprimand, according to knowledge of the victim's past behavior, Study 3.

10.2. Belief in karma: Evaluations of interpersonal harm

The final set of analyses predicted evaluations of antisocial behavior from (a) knowledge (with vs. without), (b) costliness (no cost vs. costly repercussions of reprimand), (c) participants' self-reported belief in karma, and all interactions between knowledge, costliness, and karma belief. There was a main effect of knowledge, $b = 0.29$ [0.22, 0.37], $p < .001$, and a main effect of belief in karma predicting slightly more positive evaluations of antisocial behavior, $b = 0.13$ [0.07, 0.19], $p < .001$, but belief in karma did not moderate the effects of knowledge or costliness (see Supplementary Table S30). Belief in karma was not significantly associated with evaluations of reprimand, $r = 0.044$ [−0.016, 0.10], $p = .15$.

11. Discussion

Three experiments assessed how several factors impact evaluations of interpersonal and cosmic punishment. We find some evidence of general motives for justice, such that bad outcomes are evaluated more positively when they happen to bad victims than good victims, regardless of whether these outcomes come at the hands of knowing human punishers, unwitting antisocial humans, or unseen forces. However, antisocial humans were consistently evaluated as bad people who performed unacceptable actions, whereas people who merely reprimanded transgressors and impersonal karmic retribution was evaluated much more positively. This was true even among participants who strongly believe in the law of karma, who could interpret antisocial behavior directed to immoral victims as part of the cycles of cosmic justice through which bad people inevitably face bad consequences. Explicit belief in karma predicted increased expectations of impersonal punishments, but karma belief was largely unassociated with evaluations of human punishment and antisocial behavior, implying that somewhat different factors influence perceptions of interpersonal and supernatural punishment.

As evidence of general motives for justice, through which bad people are desired and expected to face bad outcomes (Callan et al., 2014; Carlsmith & Darley, 2008; Hafer & Bègue, 2005; Lerner, 1980), these studies revealed a consistent pattern of more positive evaluations when bad things happen to victims who have behaved badly. Harm that befalls victims who previously committed transgressions was viewed as less bad than harm that befell prosocial victims, both when the harm came at the hands of other people and when harm was not caused by obvious human actions. Participants judged harm to bad victims as less negative even when the target was unaware of the victim's past transgression, and therefore when the "punisher" cannot be interpreted as intentionally enacting third-party punishment of that transgression. However, many participants gave open-ended explanations for harming bad victims that implied that the antisocial target had knowledge of the past transgression, showing that observers were willing to infer connections between past bad deeds and current harms that go beyond the information directly presented in the vignettes. These responses are consistent with general intuitions about immanent justice that are hypothesized to make beliefs about human and supernatural punishment so appealing (Baumard & Boyer, 2013). However, there were other factors relevant to judging human punishers and supernatural justice that go beyond a general desire for justice.

11.1. Judgments of human punishers

When judging bad outcomes caused by other people, participants were also sensitive to whether the target was knowledgeable of the victim's past behavior, being especially lenient in their judgments when harm is knowingly directed towards antisocial victims. Participants were less favourable towards harm unknowingly directed towards antisocial victims, or when harm was directed towards prosocial victims, likely due to different motives and traits signaled by harm in each of

these contexts. Harm must be knowingly directed towards an antisocial victim in order for observers to interpret it as fair punishment directed towards deserving transgressors, and motivated by the punisher's positive moral principles rather than being motivated by selfish or malicious desires. Importantly, knowledge did not impact perceptions of harm or help towards good victims (harm was always bad and help was always good, regardless of whether it was knowingly committed). Knowingly harming an antisocial victim seems to signal a different motive than unknowingly committing the same harm, whereas harming a prosocial victim signals negative traits regardless of knowledge.

Overall, any form of harm was viewed quite negatively (i.e., more often on the "bad" than "good" side of the scale), consistent with prior literature that finds negative reputational consequences of interpersonal third-party punishment. Across many prior studies, where targets have different options to express their disapproval and signal their moral traits, enacting third-party punishments that impose a cost on transgressors is generally evaluated more negatively than alternative responses, such as simple reprimands or compensating the victim (e.g., Chavez & Bicchieri, 2013; Dhaliwal et al., 2021; Ferguson et al., 2019; Heffner & FeldmanHall, 2019). For example, in Study 3 merely reprimanding the transgressor (i.e., calling them out for their bad deed and instructing them to do better) led to much more positive evaluations (on the positive side of the scale), whereas actually inflicting a costly penalty on transgressors was viewed a slightly bad. Reprimanding transgressors may therefore provide the most socially-acceptable way for observers to signal their own moral character and to potentially deter future transgressions by threating future punishments that would be costly to the transgressor, without needing to risk the reputational and practical costs of actually punishing the transgressor now. Although verbal reprimands may not be as effective a deterrent as facing harmful punishments, reprimand may provide an attractive way to reconcile the relative costs and benefits of confronting moral transgressions.

Third-party punishment may only have reputational benefits in specific contexts, and future studies could more systematically investigate moderating factors. For example, in economic games where punishment is the only way to signal moral disapproval, willingness to face personal costs to impose a punishment on someone else may signal one's good moral character (Barclay, 2006; Jordan & Rand, 2019). But there may be other more reliable ways to signal moral character in real-world situations with less constraints, such as when it is possible to compensate the victim (Heffner & FeldmanHall, 2019). The types of financial losses imposed as costly punishment in economic games may also be viewed as a more appropriate penalty than the unhelpful and selfish behavior the current studies. Punishment type and severity may need to be appropriately scaled relative to the original transgression, in ways that were not tested systematically in the current studies. People may also be reluctant to view anyone as a legitimate source of punishment, and instead restrict punishment responsibilities to specific authority figures. For example, social norms may give higher-status individuals the authority to regulate others' behavior through punishment (Redhead et al., 2021), in ways that are deemed inappropriate when someone punishes same-status peers or on strangers. Social institutions (e.g., legal systems, police forces) may also grant certain individuals the right to enact punishments (e.g., prison time, fines) that we do not allow lay people to impose on each other. Social institutions that regulate who is allowed to impose costs on transgressors may mitigate the risk that individual punishers are viewed as acting from their own selfish interests or from a sadistic desire to cause harm to others, thereby yielding greater reputational benefits to those who punish third-party transgressions.

The costliness of punishment may sometimes help to signal that the punisher has altruistic motives, and thereby make punishment a more credible signal of moral character (Rai, 2022). In Study 3 we manipulated the costliness of the antisocial behavior, but we did not find any significant differences between evaluations of costly antisocial behavior and non-costly antisocial behavior, nor any moderation of knowledge effects by costliness. However, the circumstances of the costliness in our studies may not have been the most ideal form of costly punishment to yield reputational benefits for punishers, because they did not obviously know that the punishment would turn out to be costly (see limitations and future directions below). Altogether, these present findings, and inconsistencies between previous studies using various methods to assess third-party punishment, suggest that evaluations of punishment are likely moderated by several contextual variables, that are worth carefully considering in future research. In our studies, which focus on average people interacting in everyday life, participants preferred prosocial over antisocial behavior, regardless of who it was directed towards.

These findings demonstrate the mixed reputational consequences of engaging in third-party punishment. Interpersonal punishment offers benefits to individuals and groups by signaling standards for appropriate moral behavior, and provides a threat that can deter transgressions. There may also be a general desire to see bad things happen to bad people regardless of the mechanism for punishment. However, harming antisocial victims may also signal negative traits in the punisher, who is willing to take time, energy, and potentially risk their safety to cause harm to another person. These risks of human third-party punishers make alternative (less antisocial) responses to transgressions more attractive, such as engaging in social exclusion, reprimand, and gossip, or forgoing human intervention entirely and leaving punishment up to supernatural sources of cosmic justice.

Although our focus in these studies was on evaluations of antisocial behavior, we also explored, in Study 1, whether the reputational effects of *prosocial* behavior depended on knowledge of the recipient's bad behavior. We did not find substantial differences in how prosocial behavior was evaluated when it was directed towards previously-prosocial versus previously-antisocial recipients. Although the least positive ratings were of prosocial behavior knowingly directed to an antisocial recipient, all evaluations of prosocial behavior were very positive. Participants' open-ended explanations for the behavior further show that the target's good character was equally relevant regardless of whether the recipient was previously prosocial or previously antisocial. Because we investigated prosocial behavior that is generally nice and normative regardless of who it is directed towards (e.g., returning a misplaced phone, holding the door open), contextual factors like the recipient's past behavior may be less informative of the prosocial target's moral character. The full context of antisocial behavior may be more relevant when interpreting and evaluating harmful behavior. However, future studies of alternative scenarios may find other situations in which the virtue of prosocial actions differs when directed towards good versus bad beneficiaries. For example, prosocial actions may be viewed as far less positively when they enable or protect other people's immoral actions (e.g., helping someone to cause harm, not reporting someone who committed a crime), but prosocial actions might be viewed far more positively when they benefit transgressors in unrelated contexts (as in our study) or assist with transgressors' rehabilitation.

11.2. Judgments of supernatural punishment

Different psychological factors were at play when harm occurred without any obvious human cause, such as when someone accidentally loses their wallet rather than having it stolen. These impersonal negative events were evaluated much more positively than negative interpersonal behavior, especially when they happened to a bad victim, being closer to a neutral scale midpoint than overtly negative, despite identical negative outcomes to the interpersonal scenarios. This indicates that while many people are pleased when bad things befall bad people, they do not necessarily want human beings to be the source of these harms. Impersonal social institutions (e.g., police and legal frameworks) and supernatural punishment beliefs (e.g., morally concerned gods and karmic payback) may provide attractive ways to regulate punishment, such that transgressors are punished and deterred from future

transgressions, but individual people are not required to harm each other to enact punishment. Theoretical accounts of religion's role in cooperation have outlined how supernatural punishment beliefs can change the incentive structure to mitigate potential risks of human-enforced punishment (e.g., Norenzayan et al., 2016; Schloss & Murray, 2011). Experimental work confirms that people are actually less willing to enact third-party punishment themselves when reminded that God (Laurin et al., 2012) or karma (Zhou et al., 2024) is there to do the punishing instead.

However, the desire to see bad people punished by other people or supernatural forces is not sufficient to explain the prevalence and distribution of explicit belief in karmic causality. As documented in Studies 2 and 3, levels of explicit belief in karmic causality – that the universe operates such that good people are rewarded and bad people punished, both within this lifetime and across multiple cycles of reincarnation – substantially differed between individuals and between members of different religious groups. In contrast, evaluations of interpersonal punishment, reprimand, and impersonal harms did not significantly differ across members of different religious groups, or based on participants' strength of explicit karma belief.

These findings are consistent with prior evidence that explicit belief in supernatural forces, like karma, is shaped by a variety of cultural and cognitive influences that go beyond mere intuitions about justice or expectations of interpersonal fairness (White, Baimel, & Norenzayan, 2021; White & Norenzayan, 2019; White, Willard, Baimel, & Norenzayan, 2021). It also indicates that people's explicit supernatural beliefs are not necessarily applied to how they understand everyday interpersonal behavior. Although there are certainly many occasions in which people use their religious worldviews to understand other people's behavior and choose appropriate actions for themselves (McKay & Whitehouse, 2015; Norenzayan et al., 2016; White & Norenzayan, 2019), in our studies there were no clear associations between explicit karma belief and evaluations of interpersonal punishment. Even participants who strongly believe that karmic principles govern the universe, such that bad people are eventually punished, did not view it as appropriate for human beings to be the source of harm to bad victims. This finding is consistent with other studies showing a negative association between belief in karma and willingness to enact personal revenge (Goyal & Miller, 2023). These findings therefore add to the growing body of literature on how belief in karma can support interpersonal cooperation, by both providing the threat of supernatural punishment that decreases selfishness and deters norm violations (White, Kelly, Shariff, & Norenzayan, 2019; Willard, Baimel, Turpin, Jong, & Whitehouse, 2020), justifying existing social institutions (Basava, Zhang, & Mace, 2021; Cotterill, Sidanius, Bhardwaj, & Kumar, 2014), and also deterring interpersonal punishment that can escalate social tensions (Goyal & Miller, 2023; Zhou et al., 2024).

By studying karma, we add to growing literature on the diverse forms of supernatural beliefs that people apply to explain life events (Exline & Wilt, 2023; Johnson, 2021), and future research may gain further insights from studying other potential sources of supernatural punishment. Some supernatural beings, like an omniscient, loving God, might be viewed as an appropriate source of punishment, that justly enforces moral standards as part of a larger plan for moral order. But other supernatural agents, like spirits or demons who are believed to have more malicious motives for causing human harm, might be evaluated more similarly to human punishers in our studies, and condemned even when they cause misfortune to befall bad people. Future research across more diverse cultural groups, or experiments highlighting different sources of secular and supernatural punishment, could provide valuable insight into what types of punishment are viewed as most legitimate and appropriate.

11.3. Limitations and future directions

The present studies investigated evaluations of third-party punishment, as well as evaluations of other sorts of prosocial and antisocial behavior that is less readily classified as "punishment," in the limited context of imagined vignettes from everyday life. It would be valuable for future research to investigate the present variables (e.g., victim's past behavior, target's knowledge, costliness of action) across other paradigms that provide opportunities for punishment. For example, future studies could systematically manipulate these variables in the third-party punishment game, which provides certain constraints on behavior that do not exist in the real-world scenarios that we describe, or in contexts of interpersonal revenge. The relative severity of the initial transgression and the "punishment" may also matter, such that participants may view it as inappropriate to direct equally-harmful behavior at bad victims (as in our vignettes), but view it as fair and acceptable to cause minor hassles and inconveniences that are minor compared to the initial transgression. In all of these cases, some form of punishment may be viewed as more appropriate than in the present vignettes, and less likely to signal the negative traits common to the present studies. Overall higher baseline approval of punishment may also provide more opportunity to detect cultural differences or associations with belief in karma, which were not observed in the present research.

Future studies could also try other means of manipulating costliness of antisocial behavior, to better test whether costly punishment is evaluated differently from non-costly punishment. For example, in our current studies, it was not obvious that the punisher knew ahead of time that their punishment would be costly (i.e., have negative consequences for them). It may be a clearer signal of a positive reputation for targets to choose punishment even when they know their actions will be costly, rather than when they think their punishment will have no impact or even benefit themselves. Future studies could also manipulate other factors that may make third-party punishment more or less appealing, such as whether the target had other options besides punishment to signal their disapproval, relative severity of the punishment's harm relative to the initial transgression, or whether their punishment occurred in addition to other more prosocial responses, such as only punishing the transgressor after helping the victim to cope with the transgression against them.

Additional research may also benefit from including additional dependent measures. Our studies focused on a composite evaluation measure made up of judgments of the target's moral character, their blameworthiness, and the behavior's acceptability and badness. These types of judgments encompass several reactions to the person and their behavior that are relevant to social impressions, and were consistently judged quite negatively, even when harm was knowingly directed to antisocial victims. However, future studies could also ask about whether the harmful action was justified, legitimate, or deserved. Participants may be more willing to admit that harm to transgressors is deserved, even if they condemn the person who committed the harmful action, and direct ratings of justice/injustice in the situation may more closely track individual differences in karmic beliefs than do moral evaluations of antisocial people. These measures may also be less susceptible to social desirability concerns, such that participants who are pressured by social norms to condemn harm (even when enacted as "punishment") might be more willing to admit that a transgressor's bad outcomes are justified. Such social desirability concerns are unlikely to be a major influence in our study, given that all participants were expressing their opinions anonymously, and that negative evaluations were also present among believers who could use supernatural concepts like karma to legitimize the "punishment" behavior if they felt it was deserved. Future studies could use additional measures to confirm the robustness of these patterns.

Testing these ideas across a broader array of cultural contexts may also reveal important cultural influences on punishment that were not detected in the present samples. The current Studies 2 and 3 did rely on ethnically and religiously diverse samples, who varied systematically in their explicit belief in karmic causality. But all participants were

recruited from the same Canadian university and therefore may share similar ideas for how human beings should treat one another in daily life. More extreme cultural comparisons, such as between samples for whom karmic ideas are strongly endorsed and embedded in daily life and samples that explicitly deny the existence of supernatural justice, may reveal stronger associations between beliefs and culture and their evaluations of interpersonal punishment behavior.

Overall, the present results provide initial evidence that victim's past behavior and transgressor's knowledge of it influence evaluations of antisocial behavior, and that impersonal punishment and reprimand is often preferred to direct interpersonal punishment. Further, explicit beliefs that supernatural forces ensure that bad things happen to bad people does not lead people to license antisocial behavior towards bad victims, and desiring bad things to befall bad victims cannot explain individual and cultural differences in explicit supernatural justice beliefs. Rather, a variety of factors govern perceptions of interpersonal antisocial behavior, and additional cognitive and cultural factors encourage supernatural justice beliefs, with somewhat different considerations being relevant when interpreting a situation as caused by interpersonal or impersonal forces.

CRediT authorship contribution statement

Cindel J.M. White: Writing – review & editing, Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Julia W. Van de Vondervoort:** Writing – review & editing, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare no conflict of interest.

This research was funded by Social Sciences and Humanities Research Council of Canada grants to Ara Norenzayan (#410-2010-0297) and Cindel J.M. White (#430-2022-00762).

Acknowledgements

We thank Ara Norenzayan for his support and advice in planning this project.

Open practices

Preregistered methods and analysis plans for all studies are available on the Open Science Framework:

Sample 1A: https://osf.io/wp6gj/
Sample 1B: https://osf.io/6r5pe/
Study 2: https://osf.io/wtc58/
Study3: https://osf.io/u37mf/
All data and analysis code is available at https://osf.io/85639/

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2025.104761.

References

Atwater, L. E., Waldman, D. A., Carey, J. A., & Cartier, P. (2001). Recipient and observer reactions to discipline: Are managers experiencing wishful thinking? *Journal of Organizational Behavior, 22*(3), 249–270. https://doi.org/10.1002/job.67

Balafoutas, L., & Nikiforakis, N. (2012). Norm enforcement in the city: A natural field experiment. *European Economic Review, 56*(8), 1773–1785. https://doi.org/10.1016/j.euroecorev.2012.09.008

Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences of the United States of America, 111*(45), 15924–15927. JSTOR.

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior, 27*(5), 325–344. https://doi.org/10.1016/j.evolhumbehav.2006.01.003

Basava, K., Zhang, H., & Mace, R. (2021). A phylogenetic analysis of revolution and afterlife beliefs. *Nature Human Behaviour, 1–8*. https://doi.org/10.1038/s41562-020-01013-4

Batistoni, T., Barclay, P., & Raihani, N. J. (2022). Third-party punishers do not compete to be chosen as partners in an experimental game. *Proceedings of the Royal Society B: Biological Sciences, 289*(1966), 20211773. https://doi.org/10.1098/rspb.2021.1773

Baumard, N., & Boyer, P. (2013). Explaining moral religions. *Trends in Cognitive Sciences, 17*(6), 272–280. https://doi.org/10.1016/j.tics.2013.04.003

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America, 100*(6), 3531–3535. https://doi.org/10.1073/pnas.0630443100

Bronkhorst, J. (2011). *Karma*. University of Hawai'i Press. http://muse.jhu.edu/book/1739.

Callan, M. J., Ellard, J. H., & Nicol, J. E. (2006). The belief in a just world and immanent justice reasoning in adults. *Personality and Social Psychology Bulletin, 32*(12), 1646–1658. https://doi.org/10.1177/0146167206292236

Callan, M. J., Sutton, R. M., Harvey, A. J., & Dawtry, R. J. (2014). Immanent justice reasoning. In M. P. Zanna, & J. M. Olson (Eds.), *Vol. 49. Advances in experimental social psychology* (pp. 105–161). Elsevier. https://doi.org/10.1016/B978-0-12-800052-6.00002-0.

Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. In , *Vol. 40. Advances in experimental social psychology* (pp. 193–236). Academic Press. https://doi.org/10.1016/S0065-2601(07)00004-4.

Chavez, A. K., & Bicchieri, C. (2013). Third-party sanctioning and compensation behavior: Findings from the ultimatum game. *Journal of Economic Psychology, 39*, 268–277. https://doi.org/10.1016/j.joep.2013.09.004

Cotterill, S., Sidanius, J., Bhardwaj, A., & Kumar, V. (2014). Ideological support for the Indian caste system: Social dominance orientation, right-wing authoritarianism and karma. *Journal of Social and Political Psychology, 2*(1), 98–116. https://doi.org/10.5964/jspp.v2i1.171

Dhaliwal, N. A., Patil, I., & Cushman, F. (2021). Reputational and cooperative benefits of third-party compensation. *Organizational Behavior and Human Decision Processes, 164*, 27–51. https://doi.org/10.1016/j.obhdp.2021.01.003

Exline, J. J., & Wilt, J. A. (2023). Supernatural attributions: Seeing god, the devil, demons, spirits, fate, and karma as causes of events. *Annual Review of Clinical Psychology, 19*, 461–487. https://doi.org/10.1146/annurev-clinpsy-080921-081114

Feather, N. T., & Sherman, R. (2002). Envy, resentment, schadenfreude, and sympathy: Reactions to deserved and undeserved achievement and subsequent failure. *Personality and Social Psychology Bulletin, 28*(7), 953–961. https://doi.org/10.1177/014616720202800708

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Ferguson, E., Quigley, E., Powell, G., Stewart, L., Harrison, F., & Tallentire, H. (2019). To help or punish in the face of unfairness: Men and women prefer mutually-beneficial strategies over punishment in a sexual selection context. *Royal Society Open Science, 6*(9), Article 181441. https://doi.org/10.1098/rsos.181441

Goyal, N., & Miller, J. G. (2023). Beliefs in inevitable justice curb revenge behaviours: Cultural perspectives on karma. *European Journal of Social Psychology, 53*(4), 732–745. https://doi.org/10.1002/ejsp.2933

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Hafer, C. L., & Bègue, L. (2005). Experimental research on just-world theory: Problems, developments, and future challenges. *Psychological Bulletin, 131*(1), 128–167. https://doi.org/10.1037/0033-2909.131.1.128

Heffner, J., & FeldmanHall, O. (2019). Why we don't always punish: Preferences for non-punitive responses to moral violations. *Scientific Reports, 9*(1). https://doi.org/10.1038/s41598-019-49680-2. Article 1.

Johnson, K. A. (2021). God Karma, Jinn, spirits, and other metaphysical forces. *Current Opinion in Psychology, 40*, 10–14. https://doi.org/10.1016/j.copsyc.2020.08.001

Johnson, K. A., Okun, M. A., Cohen, A. B., Sharp, C. A., & Hook, J. N. (2019). Development and validation of the five-factor LAMBI measure of god representations. *Psychology of Religion and Spirituality, 11*(4), 339–349. https://doi.org/10.1037/rel0000207

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. Nature (London), 530(7591), 473-476D. https://doi.org.ezproxy.library.ubc.ca/10.1038/nature16981.

Jordan, J. J., & Rand, D. G. (2019). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000186

Laurin, K., Shariff, A. F., Henrich, J., & Kay, A. C. (2012). Outsourcing punishment to god: Beliefs in divine control reduce earthly punishment. *Proceedings of the Royal Society of London - Series B: Biological Sciences*. https://doi.org/10.1098/rspb.2012.0615. rspb20120615.

Lee, Y., & Warneken, F. (2020). Children's evaluations of third-party responses to unfairness: Children prefer helping over punishment. *Cognition, 205*, Article 104374. https://doi.org/10.1016/j.cognition.2020.104374

Leliveld, M. C., van Dijk, E., & van Beest, I. (2012). Punishing and compensating others at your own expense: The role of empathic concern on reactions to distributive injustice. *European Journal of Social Psychology, 42*(2), 135–140. https://doi.org/10.1002/ejsp.872

Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion*. Plenum Press.

Li, J., Li, S., Wang, P., Liu, X., Zhu, C., Niu, X., … Yin, X. (2018). Fourth-party evaluation of third-party pro-social help and punishment: An ERP study. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00932

Lotz, S., Okimoto, T. G., Schlösser, T., & Fetchenhauer, D. (2011). Punitive versus compensatory reactions to injustice: Emotional antecedents to third-party interventions. *Journal of Experimental Social Psychology, 47*(2), 477–480. https://doi.org/10.1016/j.jesp.2010.10.004

McKay, R., & Whitehouse, H. (2015). Religion and morality. *Psychological Bulletin, 141*(2), 447–473. https://doi.org/10.1037/a0038455

Nelissen, R. M. A. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior, 29*(4), 242–248. https://doi.org/10.1016/j.evolhumbehav.2008.01.001

Nikiforakis, N., & Mitchell, H. (2014). Mixing the carrots with the sticks: Third party punishment and reward. Experimental Economics; New York, 17(1), 1–23. https://doi.org.ezproxy.library.ubc.ca/10.1007/s10683-013-9354-z.

Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., & Henrich, J. (2016). The cultural evolution of prosocial religions. *Behavioral and Brain Sciences, 39*. https://doi.org/10.1017/S0140525X14001356. e1 (19 pages).

Pedersen, E. J., McAuliffe, W. H. B., & McCullough, M. E. (2018). The unresponsive avenger: More evidence that disinterested third parties do not punish altruistically. *Journal of Experimental Psychology: General, 147*(4), 514–544. https://doi.org/10.1037/xge0000410

Pedersen, E. J., McAuliffe, W. H. B., Shah, Y., Tanaka, H., Ohtsubo, Y., & McCullough, M. E. (2020). When and why do third parties punish outside of the lab? A cross-cultural recall study. *Social Psychological and Personality Science, 11*(6), 846–853. https://doi.org/10.1177/1948550619884565

Przepiorka, W., & Liebe, U. (2016). Generosity is a sign of trustworthiness—The punishment of selfishness is not. *Evolution and Human Behavior, 37*(4), 255–262. https://doi.org/10.1016/j.evolhumbehav.2015.12.003

Rai, T. S. (2022). Material benefits crowd out moralistic punishment. *Psychological Science, 33*(5), 789–797. https://doi.org/10.1177/09567976211054786

Raihani, N. J., & Bshary, R. (2015). Third-party punishers are rewarded, but third-party helpers even more so. *Evolution, 69*(4), 993–1003. JSTOR.

Redhead, D., Dhaliwal, N., & Cheng, J. T. (2021). Taking charge and stepping in: Individuals who punish are rewarded with prestige and dominance. *Social and Personality Psychology Compass, 15*(2), Article e12581. https://doi.org/10.1111/spc3.12581

dos Santos, M., Rankin, D. J., & Wedekind, C. (2013). Human cooperation based on punishment reputation. *Evolution, 67*(8), 2446–2450. JSTOR.

Schloss, J. P., & Murray, M. J. (2011). Evolutionary accounts of belief in supernatural punishment: A critical review. *Religion, Brain & Behavior, 1*(1), 46–99. https://doi.org/10.1080/2153599X.2011.558707

Taylor, E., Clutterbuck, R., Player, L., Shah, P., & Callan, M. (2022). The role of karmic beliefs in immanent justice reasoning. *Psychology of Religion and Spirituality, 14*(2), 278–282. https://doi.org/10.1037/rel0000368

Vaccarino, E., & Kawakami, K. (2021). In the office or at the gym: The impact of confronting sexism in specific contexts on support for confrontation and perceptions of others. *Self and Identity, 20*(7), 893–912. https://doi.org/10.1080/15298868.2020.1832566

White, C. J. M., Baimel, A., & Norenzayan, A. (2021). Cultural learning and cognitive biases shape religious beliefs. *Current Opinion in Psychology, 40*, 34–39. https://doi.org/10.1016/j.copsyc.2020.07.033

White, C. J. M., Kelly, J. M., Shariff, A. F., & Norenzayan, A. (2019). Supernatural norm enforcement: Thinking about karma and god reduces selfishness among believers. *Journal of Experimental Social Psychology, 84*, Article 103797. https://doi.org/10.1016/j.jesp.2019.03.008

White, C. J. M., & Norenzayan, A. (2019). Belief in karma: How cultural evolution, cognition, and motivations shape belief in supernatural justice. In J. M. Olson (Ed.), *Vol. 60. Advances in experimental social psychology* (pp. 1–63). Academic Press. https://doi.org/10.1016/bs.aesp.2019.03.001.

White, C. J. M., & Norenzayan, A. (2022). Karma and god: Convergent and divergent mental representations of supernatural norm enforcement. *Psychology of Religion and Spirituality, 14*(1), 70–85. https://doi.org/10.1037/rel0000436

White, C. J. M., Norenzayan, A., & Schaller, M. (2019). The content and correlates of belief in karma across cultures. *Personality and Social Psychology Bulletin, 45*(8), 1184–1201. https://doi.org/10.1177/0146167218808502

White, C. J. M., Willard, A. K., Baimel, A., & Norenzayan, A. (2021). Cognitive pathways to belief in karma and belief in god. *Cognitive Science, 45*(1), Article e12935. https://doi.org/10.1111/cogs.12935

Willard, A. K., Baimel, A., Turpin, H., Jong, J., & Whitehouse, H. (2020). Rewarding the good and punishing the bad: The role of karma and afterlife beliefs in shaping moral norms. *Evolution and Human Behavior, 41*(5), 385–396. https://doi.org/10.1016/j.evolhumbehav.2020.07.001

Zhou, K., Baimel, A., & White, C. (2024). Outsourcing punishment to karma: Thinking about karma reduces the punishment of transgressors. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/w5x23