Journal of Experimental Child Psychology 230 (2023) 105630

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Children endorse deterrence motivations for third-party punishment but derive higher enjoyment from compensating victims

Rhea L. Arini [a,b,*], Marukh Mahmood [a], Juliana Bocarejo Aljure [c], Gordon P.D. Ingram [c], Luci Wiggs [a], Ben Kenward [a]

[a] Centre for Psychological Research, Oxford Brookes University, Oxford OX3 0BP, UK
[b] School of Anthropology and Museum Ethnography, University of Oxford, Oxford OX2 6PE, UK
[c] Department of Psychology, Universidad de Los Andes, Bogotá, Cundinamarca, Colombia

ARTICLE INFO

Article history:
Received 12 June 2022
Revised 23 November 2022

Keywords:
Third-party punishment
Compensation of victims
Enjoyment and endorsement of third-party interventions
Punishment motives

ABSTRACT

Children's punishment behavior may be driven by both retribution and deterrence, but the potential primacy of either motive is unknown. Moreover, children's punishment enjoyment and compensation enjoyment have never been directly contrasted. Here, British, Colombian, and Italian 7- to 11-year-old children (N = 123) operated a Justice System in which they viewed different moral transgressions in *Minecraft*, a globally popular video game, either face-to-face with an experimenter or over the internet. Children could respond to transgressions by punishing transgressors and compensating victims. The purpose of the system was framed in terms of retribution, deterrence, or compensation between participants. Children's performance, endorsement, and enjoyment of punishment and compensation were measured, along with their endorsement of retribution versus deterrence as punishment justifications, during and/or after justice administration. Children overwhelmingly endorsed deterrence over retribution as their punishment justification irrespective of age. When asked to reproduce the presented frame in their own words, children more reliably reproduced the deterrence frame rather than the retribution frame. Punishment enjoyment decreased while compensation enjoyment increased over time. Despite enjoying compensation more, children preferentially endorsed punishment over compensation, especially with increasing age and

* Corresponding author at: School of Anthropology and Museum Ethnography, University of Oxford, Oxford, OX2 6PE, UK.
E-mail address: rhea88bg@gmail.com (R.L. Arini).

https://doi.org/10.1016/j.jecp.2023.105630
0022-0965/

transgression severity. Reported deterrent justifications, superior reproduction of deterrence framing, lower enjoyment of punishment than of compensation, and higher endorsement of punishment over compensation together suggest that children felt that they *ought* to mete out punishment as a means to deter future transgressions. Face-to-face and internet-mediated responses were not distinguishable, supporting a route to social psychology research with primary school-aged children unable to physically visit labs.

Introduction

When violations of moral norms occur, there is a tension between different possible courses of action that can satisfy the psychological need to see justice restored. Witnessing norm transgressions can trigger at least two different—although non-mutually exclusive—types of third-party interventions: punishment of transgressors (i.e., third-party punishment) and/or compensation of victims. Whereas punitive justice has received a great deal of academic attention, compensatory justice still lags behind (but see Gummerum et al., 2016; Petersen et al., 2012). Although some large-scale societies have moved toward increased use of reparative justice within penal systems during recent decades (including compensation; Johnstone & Van Ness, 2013), the focus remains on punitive sanctions. This is striking in light of anthropological evidence documenting that small-scale societies do not commonly adopt punishment to solve disputes (Marlowe et al., 2008), but they more frequently involve third parties in processes of mediation and arbitration. These forms of triadic settlements are aimed at promoting reconciliation between the antagonists or at rendering decisions about compensation of the wronged party (Fry, 2000; Singh & Garfield, 2022). Moreover, whereas third-party punishment seems to be a uniquely human behavior (Riedl et al., 2012), ethologists have observed triadic postconflict affiliations directed toward victims of aggressions in nonhuman primates (Fraser et al., 2009), some of which might be considered the evolutionary bases of human compensation.

The number of empirical articles on children's third-party punishment has reached double digits during the past years (see Marshall & McAuliffe, 2022, for a review). However, few developmental studies have simultaneously tested different types of third-party interventions in response to moral transgressions (reviewed below). Thus, here we investigated a range of factors that could potentially modulate children's choices between punishment and compensation, the affective states following the enactment of these two types of third-party interventions, and the motivational basis of punishment. Given that most of the literature on children's third-party interventions has been conducted in Northern European and Northern American countries (but see Yang et al., 2021, and Liu et al., 2021), we included children from three different countries (United Kingdom, Italy, and Colombia) to increase diversity in the sample. Our analysis focused on commonalities rather than differences across countries.

Third-party interventions: Compensation and punishment

The two types of third-party interventions have been put in direct comparison to establish whether people tend to be compensation- or punishment-oriented. In the adult literature, some studies have provided evidence for preference for punishment (Adams & Mullen, 2015; van Prooijen, 2010) and others for compensation (Chavez & Bicchieri, 2013; Lotz et al., 2011; Van Doorn, Zeelenberg, & Breugelmans, 2018; Van Doorn et al., 2018b). These mixed results might be due to whether the third-party intervention options at participants' disposal were personally costly or not. In studies showing higher willingness to compensate instead of punishing, both punishment and compensation were economically costly to the participants. In contrast, in studies showing preference for

punishment over compensation, there were no costs associated with either type of third-party intervention. This could be due to participants acting on intuition rather than deliberation when they can carry out third-party interventions at no cost to themselves (Van Doorn & Brouwers, 2017).

There are now a handful of studies that have investigated children's compensatory and punitive tendencies simultaneously. In one such study, 9- to 22-year-old Dutch participants witnessed instances of social exclusion and then played economic games with the victims and transgressors. Older participants paid greater costs to compensate victims and punish transgressors, but it was not possible to establish whether participants preferred compensation or punishment because the two were not contrasted directly (Will et al., 2013). Other developmental studies were explicitly designed to assess whether children tend to be compensation- or punishment-oriented. However, in contrast to the adult literature, it is challenging to identify systematic patterns, in part because compensation was differently operationalized across studies.

In one line of research, compensation entails any prosocial actions toward the victims to make up for the transgressions they suffered, whereas the transgressors' payoff is left untouched. For example, in hypothetical scenarios involving physical harm, 7- to 12-year-old Canadian children were more willing to see the transgressor being punished than the victim being compensated irrespective of age, with this tendency becoming more pronounced for more severe transgressions (Miller & McCann, 1979). Conversely, after witnessing third-party interventions in response to inequity (i.e., selfish resource distributions), 5- to 9-year-old U.S. children evaluated compensation more positively than punishment regardless of their age (Lee & Warneken, 2020). However, when given the choice to intervene themselves in response to inequity, 6- to 9-year-old U.S. children were more likely to pay a cost to enact punishment rather than compensation (McAuliffe & Dunham, 2021).

In another line of research employing ownership transgressions such as theft, punishment was pitted against restitution. Restitution confounds compensation and punishment because the action of returning the resources to the victim also affects the transgressor's payoff. All these studies used non-incentivized paradigms, where children did not need to pay a cost to enact their preferred intervention. For example, in response to theft, 3-year-old German children preferred to return stolen resources to the victim (i.e., restitution) rather than just making them inaccessible to both the thief and the victim (i.e., punishment) or doing nothing (Riedl et al., 2015). In another experiment, 3- to 6-year-old Chinese children also preferred to enact restitution over punishment, but in this case punishment entailed taking away even the resources that the transgressors had prior to the theft (Yang et al., 2021). Finally, when witnessing victims' reactions to ownership transgressions, 4- to 6-year-old Chinese children evaluated restitution more positively than punishment, especially in case of harsh punishment (Liu et al., 2021).

We did not operationalize compensation as restitution. Rather, compensation entailed providing the victim with resources to make up for the damage endured, whereas the transgressor's payoff remained unchanged. Our research intended to expand knowledge about whether children's orientation for punishment or compensation is modulated by factors such as judgment of transgression severity, transgression type, and children's age. To assess orientation, we measured punishment versus compensation endorsement via forced-choice self-report tasks after children had the opportunity to enact both compensation and punishment behaviors. However, we did not assess children's orientation by comparing levels of compensation and punishment because our paradigm used different currencies for each (material resources for compensation and time-outs for punishment).

Specifically, we investigated whether endorsement of compensation versus punishment would change as a function of how seriously children judged transgressions (Question 1 [Q1] in Table 1). Because Miller and McCann (1979) showed that children's preference for punishment over compensation was higher in response to severe transgressions compared with mild transgressions, we expected to observe children increasingly endorsing punishment over compensation the more severely they judged the transgressions.

We were also interested in whether punishment versus compensation endorsement would be affected by the type of moral transgression (Q2 in Table 1). We predicted that children would attribute more importance to punishment in case of physical harm (Miller & McCann, 1979) and to compensation in case of ownership transgressions such as theft (Liu et al., 2021; Riedl et al., 2015; Yang et al., 2021). We were less confident in making predictions about the effect of inequity transgressions given

Table 1
Research questions, related predictions, and whether data supported predictions

Topic	Research question	Prediction	Supported?
Third-party interventions: Compensation and punishment	Q1 (a priori): Does participants' judgment of transgression severity affect compensation vs. punishment endorsement (during justice administration)?	More severe judgments lead to higher endorsement of punishment over compensation.	Yes
	Q2 (a priori): Does transgression type affect compensation vs. punishment endorsement (during justice administration)?	Physical harm elicits higher endorsement of punishment over compensation.	No
		Theft and inequity elicit higher endorsement of compensation over punishment.	No
	Q3 (a priori): Does children's age affect compensation vs. punishment endorsement (during and after justice administration)?	Endorsement of punishment over compensation increases with increasing age.	Yes (but only after justice administration)
Affective states induced by third-party interventions	Q4 (a priori): Does type of third-party intervention affect enjoyment?	Compensation elicits higher enjoyment than punishment.	Yes
	Q5 (a priori): Does time affect enjoyment of punishment and compensation?	Punishment enjoyment decreases over time (no enjoyment by the end of the experiment).	Partly
		No prediction on the temporal pattern of compensation enjoyment.	NA
Punishment motives and justifications: Deterrence and retribution	Q6 (a priori): Does framing condition affect punishment severity (during justice administration)?	The punishment frame most in line with children's preexisting punishment motivation increases punishment severity.	No
	Q7 (a priori): Does children's age affect retribution vs. deterrence endorsement?	Endorsement of deterrence over retribution increases with increasing age.	No
	Q8 (post hoc): Does framing condition affect frame reproduction?	Exploratory analyses; no predictions.	NA

Note. NA, not applicable.

the lack of relevant literature at the time of developing our experiment (contrasting findings were published later; see Lee & Warneken, 2020, and McAuliffe & Dunham, 2021). However, it seemed plausible that children would preferentially endorse compensation in response to inequity because there is evidence that children are motivated to intervene as third parties by the desire to even out the resource imbalances experienced by victims (Arini et al., 2021). Importantly, in our paradigm resource imbalances could be corrected only via compensation, not punishment. Given our interest in both transgression type and transgression severity judgment, which are likely to be associated, we examined their independent effects using models controlling for both.

Finally, we tested whether children's endorsement of punishment versus compensation was dependent on age (Q3 in Table 1). Given that Riedl et al. (2015) demonstrated that third-party interventions during early childhood are compensation-oriented, whereas Miller and McCann (1979) showed that third-party interventions during middle to late childhood are punishment-oriented, we hypothesized that there might be a shift toward endorsement of punishment over compensation with increasing age. We note that this is also consistent with some more recent studies that were not available when we designed our experiment (Liu et al., 2021; McAuliffe & Dunham, 2021; Yang et al., 2021).

Affective states induced by third-party interventions

It is unclear whether individuals' endorsement of punishment versus compensation is aligned to enjoyment of punishment versus compensation. A misalignment (i.e., preferential endorsement of

punishment accompanied by higher enjoyment of compensation) would suggest that individuals see punishment as a moral duty, something that ought to be done (Arini et al., 2021).

Neuroscientific studies have begun to clarify the affective components involved in punishment and compensation in adults. The activation of the striatum, a key area in the brain's reward circuitry, seems to reflect anticipatory satisfaction. This kind of striatal response has been observed in people meting out both punishment (Strobel et al., 2011) and compensation (Hu et al., 2015). Activation of the striatum predicts charitable donations to victims of misfortune (Genevsky et al., 2013; Harbaugh et al., 2007).

Research focused on adults' subjective reports of punishment-related emotions has instead produced somewhat mixed results. People confronted with a cooperative norm violation predicted that taking revenge would make them feel better. However, once they enacted punishment, they ended up reporting lower mood (due to rumination) than those who had not punished (Carlsmith et al., 2008). It has been shown that for people to derive satisfaction from punishment, it is important to know how the transgressor reacts to the punishment. Indeed, punishers reported to be satisfied if they saw the transgressor suffer because of punishment and/or if they received proof that the transgressor learned a lesson through punishment (Aharoni et al., 2022). Avengers experienced higher levels of satisfaction than nonavengers upon receiving a message from the transgressors acknowledging that they had understood why they had been punished (Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011) or indicating a change in moral attitude (Funk et al., 2014). Nevertheless, people showed no increased willingness to punish at the prospect of receiving information on the effects of punishment on the transgressor (Funk & Mischkowski, 2022). This may suggest that people fail to predict what would bring them satisfaction following a moral transgression.

Given the gap in knowledge about children's affective states related to third-party interventions, scientific efforts have now begun focusing on the study of the emotional antecedents and consequences of children's punishment behavior. Regarding the emotional antecedents of punishment, self-reported anger after witnessing inequity has been shown to mediate the link between transgression severity and costly punishment in British adults but not in children and adolescents (Gummerum et al., 2020). Moreover, by experimentally manipulating anger (via an autobiographical recall procedure), it was possible to demonstrate that anger has a causal role in punishment severity of inequity in British adults and adolescents but not in children (Gummerum et al., 2022). Regarding the emotional consequences of punishment, 5- to 7-year-old U.S. children who meted out punishment reported higher levels of sadness and lower levels of happiness and excitement than their peers who did not punish (see supplemental material in Marshall et al., 2021). Moreover, 5- to 11-year-old British children were more likely to report no enjoyment when they enacted real punishment rather than pretend punishment (Arini et al., 2021).

In the current study, we investigated whether enjoyment would vary according to the type of third-party intervention children were enacting (Q4 in Table 1) and to the time passed since the intervention (Q5 in Table 1). We measured enjoyment immediately after children decided how to respond to transgressions and then again once they had the time to reflect on their past choices at the end of the experiment. Notably, children were told that their punishment and compensation decisions would be implemented. However, children were neither shown how the transgressors and victims reacted to their decisions nor made to think that they would receive such information. Because children could not ascertain whether the transgressors suffered or learned a moral lesson through punishment, we expected children not to derive satisfaction from punishment. Therefore, we predicted that compensation would generally elicit more enjoyment than punishment. Furthermore, based on Carlsmith et al.'s (2008) findings about the negative effect of punishment-induced rumination on adults' affective states, we expected that children's punishment enjoyment would decrease across time. Because it has been previously demonstrated that British children's emotional experience of punishing was on average neither positive nor negative when measured at the end of the experiment (Arini et al., 2021), we predicted that we would find comparable results also in the current study at the same time point. However, we did not formulate any specific prediction for the temporal pattern that compensation enjoyment would follow.

Punishment motives and justifications: Deterrence and retribution

Clarifying *what* people feel when they engage in punishment could help to explain *why* people punish (but see discussion in Funk & Mischkowski, 2022). The philosophical literature about the motivational basis of punishment can be organized around two main theories of justice: retribution and deterrence. Retributive punishment is rooted in balancing out past injustices by giving the transgressors their "just deserts" (Kant, 1790/1952). In contrast, deterrence conceptualizes punishment as a means to prevent future misbehaviors by transgressors and/or bystanders (Bentham, 1789/1948). Thus, retribution theory is in accord with the idea that punishment is motivated by the desire to see the transgressors suffer in proportion to the wrongdoing committed (suffering hypothesis). Instead, deterrence theory is in accord with the idea that punishment is aimed at communicating to the transgressors that they should make amends for their misbehaviors (understanding hypothesis) (Berman, 2010).

Psychological research has demonstrated that adults are motivated by both retribution and deterrence, but with retribution probably being more important. For example, manipulating punishment severity (as retribution-relevant information) increased participants' punitive tendencies, yet manipulating punishment observability (as deterrence-relevant information) did not (Carlsmith et al., 2002; Molho et al., 2022). Relatedly, before making punishment recommendations, people were most likely to first seek retribution-relevant information because it increased their confidence in the appropriateness of their decisions more than deterrence-relevant information (Carlsmith, 2006; Keller et al., 2010). Moreover, it has been shown that adults were more willing to invest resources into punishing transgressors when punishment satisfied only retributive motives compared with when it satisfied a combination of deterrent and retributive motives (Nockur et al., 2022; but see Crockett et al., 2014).

In the adult literature, a remarkable discrepancy has been noticed between people's actual punitive choices (which tend to be retribution-oriented) and their explicit justifications (which tend to be deterrence-oriented). People have been shown to support deterrence policies in the abstract but to reject them once they saw them contradicting retributive principles (Carlsmith, 2008). People were also shown to persist in justifying their punishment recommendations in deterrent terms even if it was pointed out to them that none of their justifications was applicable to the specific scenario (Aharoni & Fridlund, 2012). People invested resources into retribution of transgressions more often than they reported endorsing retributive justifications (Crockett et al., 2014). This mismatch between implicit punishment motivations and explicit justifications may be driven by people's lack of insight into, or inability to express, the motivations of their own behavior (Carlsmith, 2008). An implication of this argument is that deterrent justifications are primarily post hoc rationalizations of retributive impulses (Aharoni & Fridlund, 2012; Carlsmith & Darley, 2008; Keller et al., 2010; but see Rehren & Zisman, 2022). This mismatch could also indicate the existence of a social desirability bias given that adults may be aware that being perceived as aggressive can have a detrimental effect on their reputation (Eriksson et al., 2016; Gordon et al., 2014; Raihani & Bshary, 2015a).

Regarding the developmental literature, both interview and experimental studies have shown that children are capable of deterrence reasoning from a young age, at least in U.S. contexts. Stern and Peterson (1999) analyzed how 4- to 11-year-old children justify their punishment choices in response to a variety of transgressions. Children of all ages were equally likely to use eye-for-an-eye justifications. However, starting from 7 or 8 years of age, children began to also show awareness of the preventive function of punishment. Bregant et al. (2016) presented 5- to 8-year-old children with a scenario depicting a character stealing a resource from another character. The theft either remained unpunished or was followed by a punishment (not decided by the children themselves). Children predicted that the punished thief would be less likely to misbehave again than the unpunished thief. Dunlea and Heiphetz (2021) found that 6- to 8-year-old children—but not adults—reported that "mean" people became "nicer" after both severe and mild forms of punishment (incarceration and time-out, respectively). In Yudkin et al.'s (2020) study, 3- to 6-year-old children could decide whether to engage in punishment by preventing a harmful peer from accessing a playing opportunity. Once questioned about the reasons for their punitive decisions, they mentioned the desire to see the transgressor change his or her behavior and learn a lesson. Importantly, these expressions of desire for reform correlated with children's actual punishment rates.

Regarding the motivational basis of punishment, there is evidence that, already from early childhood, children prefer to see bad things happen to those who behave badly toward others (Hamlin et al., 2011; Kenward & Östh, 2012, 2015), suggesting incipient retributive desires or expectations. However, a couple of recent studies conducted on older participants (4- to 7-year-old U.S. children and 9- to 12-year-old German children) suggest that children can be motivated by both deterrence and retribution. There is indeed evidence that children punished transgressors at higher rates and invested more resources into punishment when doing so satisfied both retributive and deterrent motives compared with when they satisfied purely retributive motives (Marshall et al., 2021; Twardawski & Hilbig, 2020). In the face of a large body of evidence suggesting that adults are often motivated by retribution despite the deterrent justifications they provide, but more limited evidence in children, it is still unclear whether this mismatch between implicit punishment motivations and explicit justifications is also present when children punish.

Regarding the implicit motivations of punishment, we explored whether children's punishment severity would change depending on the type of punishment frame to which they had been experimentally assigned (Q6 in Table 1), that is, whether children's role as third-party punishers was framed as serving a deterrent or retributive purpose. We argue that the punishment frame most in line with children's preexisting punishment motivation would also be the most effective at increasing their punishment severity (for a similar paradigm, see van Prooijen, 2010). Because there were few grounds for a specific prediction as to whether children are primarily motivated by deterrence or retribution, we did not make one.

Regarding the explicit justifications of punishment, we analyzed how children justified the punishment behaviors they had meted out. More specifically, we assessed endorsement of deterrence or retribution in a forced-choice task, expecting to observe increasing rates of deterrence endorsement with age (Q7 in Table 1), consistent with Stern and Peterson's (1999) findings. Because it has been argued that retribution is driven by intuition and deterrence is driven by deliberation (Aharoni & Fridlund, 2012; Carlsmith & Darley, 2008; Keller et al., 2010; but see Rehren & Zisman, 2022), we predicted that younger children would endorse retribution more often than older children due to their relative lack of inhibitory control and forward-looking reasoning skills.

Method validation: Setting method and believability

This study piloted a new method of conducting experimental research on children—an online virtual environment in the form of a Justice System based on the world of *Minecraft*, a globally popular commercial video game. To test the validity of this method, child participants using this Minecraft Justice System met either face-to-face or over the internet, and we compare results obtained in each way. We also checked that the Minecraft Justice System would appear credible to the children, expecting that the majority of them would believe they had judged misbehaviors that had actually happened on the server, given that player misbehavior and justice administration by other players are now normal in children's online playgrounds (Beale et al., 2016; Kou et al., 2017).

Method

The study was approved by the Oxford Brookes University ethical review committee. All raw data, experimental scripts, and analysis pipelines are available at the Open Science Framework (https://osf.io/4ygw5).

Participants

Exclusions from the dataset amounted to 3 children, with 1 child being excluded because of an experimenter's technical mistake and 2 children being excluded because they had difficulties in comprehending an experimenter's questions due to lack of experience with playing Minecraft (see script in S1 of online supplementary material). After exclusions, participants were 123 children (M_{age} = 9.83 years, *SD* = 1.41, range = 7.05–11.97; distribution: 16 7-year-olds, 16 8-year-olds, 34 9-year-olds,

23 10-year-olds, 34 11-year-olds; 32 girls and 91 boys) residing in the United Kingdom (n = 67), Colombia (n = 23), or Italy (n = 33). Of these 123 children, 43 were assigned to the retribution frame, 40 to the deterrence frame, and 40 to the compensation frame (see Table S3 in supplementary material for sample breakdown according to age, country, and framing condition). Our choice of countries was opportunistic to maximize diversity in the sample. Sample size was determined by logistic constraints; we collected as much data as practically possible within the time periods.

Participants were tested in one of two alternative settings: either face-to-face (at science fairs or a technology-themed summer camp) or remotely over the internet (via Skype or WhatsApp video or voice calls, depending on the reliability of the internet connection that participants had access to from their homes). Data collection lasted from late June 2018 to the beginning of March 2019.

The Italian sample, consisting of 33 children mainly from middle-income backgrounds, was tested entirely over the internet (nationwide recruitment). The Colombian sample, formed by 23 children mainly from middle- to high-income backgrounds and recruited at the same summer camp of a large city, was tested entirely in a face-to-face setting. The British sample, coming from mixed sociodemographic backgrounds, was the only one to be tested in both settings, with 35 children being tested over the internet (nationwide recruitment) and 32 being tested face-to-face (recruitment at two different science fairs in the same medium-sized English city). The categorization of the children in terms of their socioeconomic status was made through informal communications with gatekeepers or knowledge of general characteristics of the catchment areas.

Stimuli

Eight short videos depicting players' behaviors in Minecraft were recorded and embedded into a *Qualtrics* platform questionnaire to create an online Justice System called Squidcraft (link to the Justice System Qualtrics: http://bit.ly/obust33, where video streaming has now been disabled; videos now available at the Open Science Framework: https://osf.io/4ygw5). An offline version that was identical except for minor formatting aspects was also developed for the purpose of testing at science fairs where an internet connection was not reliable. The system was formatted to resemble an administrative control panel interface rather than a questionnaire (Fig. 1).

The videos, varying in length from 25 to 54 s, represented various moral transgressions during Minecraft play (see Table 2 for brief transgression descriptions and Section S1.6 in supplementary material for full descriptions).

Design

We adopted a mixed design in which the within-participant variables were *time* (during and after justice administration) and *transgression type* (see Table 2 and Section 1.6 in supplementary material). The between-participant variables were *country of residence* (United Kingdom, Colombia, or Italy), *setting method* (over the internet or face-to-face), *framing condition* (retribution, deterrence, or compensation), *gender* (male or female), and *age* (7–11 years). These between-participant variables were counterbalanced against two between-participant nuisance variables: *transgression order* (see Table S2 in supplementary material) and *question order* (see Sections S1.8, S1.9, and S1.10 in supplementary material).

The dependent variables were *punishment severity, compensation level, punishment enjoyment, compensation enjoyment, punishment versus compensation endorsement, retribution versus. deterrence endorsement, frame reproduction,* and *believability of the Justice System.* Dependent variables were measured during justice administration (i.e., repeatedly after each moral transgression), after justice administration (i.e., after all moral transgression scenarios were complete), or at both time points (see Table 3 for details). After each moral transgression, we also measured for use as a covariate the participants' *judgment of transgression severity* (on a 6-point ordinal scale ranging from −5 [*very bad*] to 0 [*not bad, not good*]).

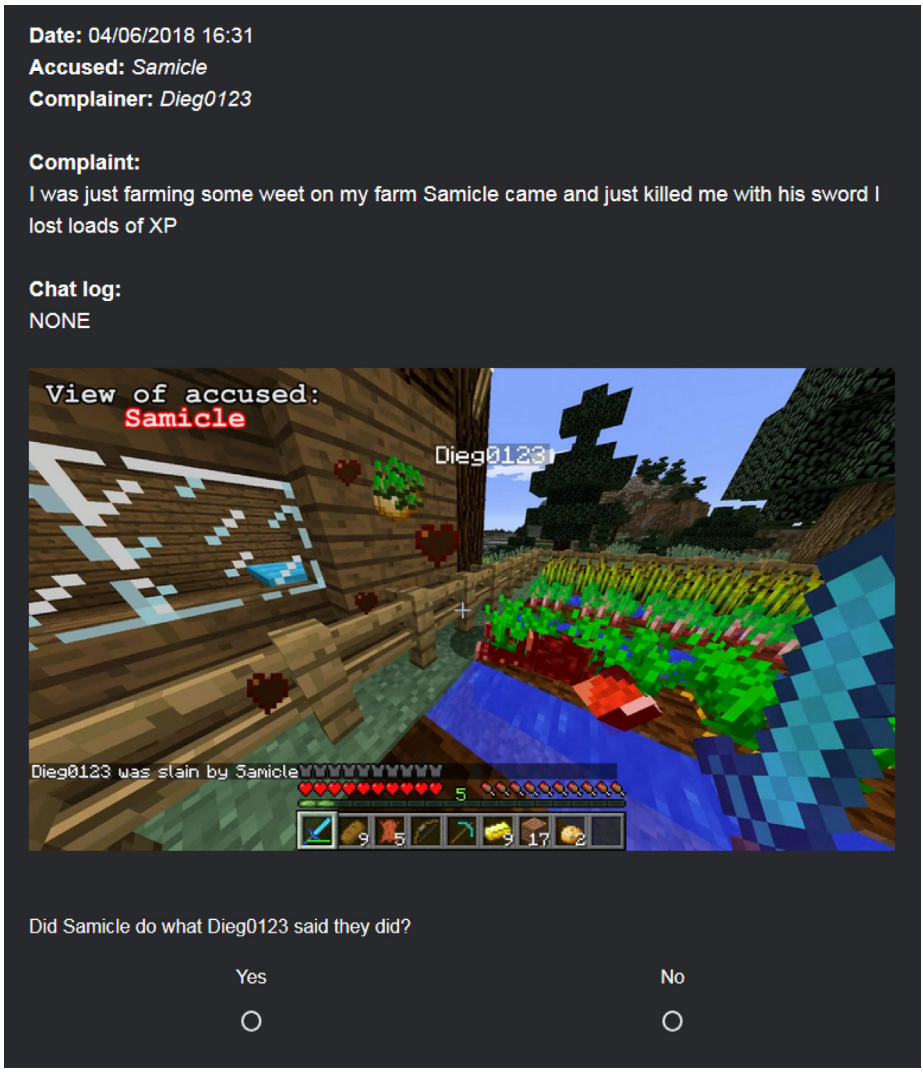

Fig. 1. Screenshot from the Squidcraft Justice System interface operated by participants. A still image from the Physical Harm condition video in which Samicle has just attacked and killed Dieg0123 is shown along with the first question to be answered.

Procedure

Parents gave consent for their children to participate after having received information about the experiment; an opt-in consent system was applied in Italy and the United Kingdom, and an opt-out consent system was applied in Colombia (because usage of Minecraft was already part of the standard activities of the summer camp from which we recruited). In addition to a specific age range (7–11 years), the other requirement for participation was to have already played Minecraft prior to the experiment in order to understand the dynamics between the players (experience with Minecraft was certified by the children's parents).

After establishing the connection (for internet data collection) or seating the child at the computer (for face-to-face data collection), the procedure began with the researcher explaining that during the

Table 2
Brief description of transgressions

Transgression type	Description
Physical harm	Player C was farming some wheat when he was unexpectedly killed by Player A with a sword.
Property destruction	Player A and Player C worked together to build a house, but then Player A destroyed it by setting it on fire.
Sanctity/authority transgression	Player A killed a holy squid in a temple while Player C was making an offer of gold to it.
Theft	While Player C and another player were trading an enchanted pick for an emerald, Player A appeared and stole both resources.
Inequity/disloyalty	Two players mining together (Player A and Player C) had promised each other to equally divide any emeralds or diamonds they discovered. When they found two emeralds, Player A seized them both, refusing to share with Player C.
Deception/liberty violation	Player A persuaded Player C to follow him into a place, where Player A trapped Player C inside an obsidian pit.
Harm-related false accusation (control)	Player C complained that she was set on fire by Player A, but the video reveals that this was not the case.
Property-related trivial accusation (control)	Player C, who claimed to need a lot of wood, complained that Player A had harvested a tree in the village common forest.

Note. Player A stands for "accused," while Player C stands for "complainer." Children were asked to decide whether to punish the former and compensate the latter in the Squidcraft Justice System.

Table 3
Description of all dependent variables

Dependent variable	Time of measurement	Measurement scale
Frame reproduction (manipulation check)	Before Trial 1; between Trials 4 and 5	Categorical choices: "punishment with undetermined motivation"; "retribution"; "deterrence"; "compensation"
Punishment severity	During justice administration	11-point ordinal scale ranging from 0 = "no ban from the server" to 10 = "4-week ban"
Compensation level	During justice administration	11-point ordinal scale ranging from 0 = "0 Minecraft diamonds" to 10 = "10 Minecraft diamonds"
Punishment enjoyment	During and after justice administration	11-point ordinal scale: −5 was "very bad"; 0 was "not bad, not good"; +5 was "very good"
Compensation enjoyment	During and after justice administration	11-point ordinal scale: −5 was "very bad"; 0 was "not bad, not good"; +5 was "very good"
Punishment vs. compensation endorsement	During and after justice administration	Binary choice between 0 = "compensation" and 1 = "punishment"
Retribution vs. deterrence endorsement	After justice administration	Binary choice between 0 = "deterrence" and 1 = "retribution"
Believability of the Justice System (manipulation check)	After justice administration	Binary choice between 0 = "not believable" and 1 = "believable"

experiment they would not be Minecraft players themselves but rather judges helping to test a newly set up Justice System for a Minecraft server called Squidcraft (a Minecraft server is an online multiplayer arena where players can interact in numerous ways, with some servers allowing both prosocial and antisocial interactions, with the latter known informally as "griefing"; Beale et al., 2016). Participants were told that players on the server experiencing misbehaviors from other players could log their complaints into the Justice System. These complaints, along with the chat logs between the players and video renditions of the behaviors in question (see Section S1.6 in supplementary material for details of the complaints and chat logs), would then be shown to a Justice System judge for action to be taken (very similar player-operated justice systems are featured in real computer games; e.g., Kou

et al., 2017). In reality, the complaints and chat logs had been previously written, and the videos had been prerecorded. This element of deception was revealed to the children once the experiment was completed.

According to the framing condition to which children were assigned, the purpose of the Justice System was described by emphasizing its retributive, deterrent, or compensatory functions. This frame was repeated twice, paraphrased in different ways (see Sections S1.2 and S1.4 in supplementary material). The experimenter checked twice whether children could reproduce (with their own words) the frame: before Trial 1 (immediately after framing) and then again between Trials 4 and 5. In both frame manipulation checks, children were asked whether they remembered the purpose of the Justice System. The experimenter immediately coded whether their explanations contained mentions of compensation, deterrence, or retribution or of a general punitive motivation with no specific links to retribution or deterrence. When children's answers did not match the assigned frame, the experimenter repeated the frame to the children. Recordings were blind double-coded (see Section S4.1 in supplementary material for coding criteria).

After children had responded to the first frame manipulation check, the experimenter assisted them in navigating to the first complaint of the Justice System (e.g., for internet data collection, by pasting the link into the text-chat channel of the connection), thereby starting the justice administration phase of the experiment. Following the reading of the relevant chat log and the viewing of the video, children were asked whether they believed the accused player had done what the complaining player said he or she had done. In case of an affirmative answer, children were required to judge the accused player's transgression by rating its severity on a scale ranging from *very bad* to *not bad, not good* (first 6 points of the Likert scale in Fig. 2). At this point, children could decide both the amount of compensation (number of diamonds) to allocate to the complainer/victim and the amount of punishment (length of ban from the server) for the accused/transgressor. Children did not need to pay any economic cost to enact their third-party decisions; however, the consequences of these decisions for transgressors and victims were presented as real, and the children used the Justice System interface to make the decisions. To avoid ceiling effects with compensation choices, the experimenter initially specified that diamonds were limited and discouraged the children from always giving the maximum number of diamonds. The order of punishment and compensation-related questions was counterbalanced across participants.

Immediately after children decided to enact punishment and/or compensation (i.e., during justice administration), they were asked to indicate how they felt in punishing and/or compensating on a scale ranging from *very bad* to *very good* (all 11 points of the Likert scale in Fig. 2). The enjoyment question during justice administration was asked multiple times (i.e., every time children had punished a moral transgression). If children had decided to assign both punishment and compensation, they answered a forced-choice question about whether they considered the former or the latter (with order of mentioning counterbalanced) more important in this specific case.

All participants were presented with the same eight complaints, with order of appearance counterbalanced across participants. When all eight complaints had been judged (i.e., after justice administration), participants needed to answer the final block of questions. Children needed to rate on the 11-point Likert scale how performing acts of punishment and compensation had made them feel, whether they attributed more importance to punishing transgressors or compensating victims, and whether their main reason for punishing transgressors was for deterrence or retribution. The

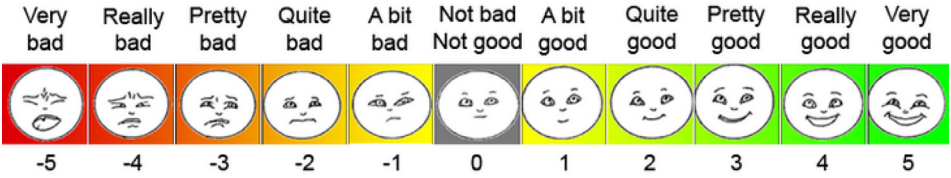

Fig. 2. Likert scale used to measure both judgments of transgression severity and the affective states related to punishment and compensation.

internal order of these questions was counterbalanced across participants. Finally, the experimenter checked whether children truly believed they had judged misbehaviors that had actually happened on the Minecraft server (see Section S1.10 in supplementary material for details of the questions). Of note, self-reported measures (particularly endorsement of punishment vs. compensation and of retribution vs. deterrence) were asked after the behavioral measures (punishment severity and compensation level). We consider this arrangement as best-suited to minimize the risk that individual differences in endorsements influence children's behavioral choices downstream.

Analysis strategy and statistics

To test our research hypotheses, we adopted linear models implemented using the *lme4* package (Version 1.1-26) in the R programming environment (Version 4.0.2; R Core Team, 2020). We used *lmer* to analyze linear mixed-effects models of continuous dependent variables (punishment severity, punishment enjoyment, compensation level, and compensation enjoyment), *glmer* to analyze logistic mixed-effects models of binary dependent variables (punishment vs. compensation endorsement during justice administration); and *glm* to analyze linear fixed-effects models of binary dependent variables (retribution vs. deterrence endorsement, punishment vs. compensation endorsement after justice administration).

To test predictions and to explain variance in dependent variables (and thus increase statistical power to test predictions), models included a range of factors (fixed unless stated otherwise). All models included age, framing condition, setting method, gender, country of residence, question order, and believability (to test predictions and explain variance). All within-participant models include participant ID as a random factor. When dependent variables were modeled trial by trial during justice administration, models included transgression severity judgment and transgression type (to test predictions and explain variance), with transgression type as a random factor to improve generalizability of findings. When dependent variables were modeled to compare their values during and after justice administration, the model compared the mean dependent variable value during justice administration trials with the value from afterward, and models included time-point (during vs. after, to test predictions) and mean transgression severity judgment during trials (to explain variance). Regarding the nuisance variables, only question order—and not transgression order—was included because of stronger theoretical reasons to expect an effect (Condon & DeSteno, 2011). Because of the large number of relations between variables included in these models, we discuss in the main text only relations for which we made predictions, to minimize the false discovery rate. In all our models, we included only main effects, not interaction effects. For intuitive interpretation, some effect sizes for pairwise category comparisons are stated as Cohen's *d* with associated confidence interval (CI); these are calculated by dividing the relevant dummy regression coefficient and associated CI by the standard deviation of the dependent variable. See Tables S4–S11 in supplementary material for full model specifications.

Preliminary analysis of the control scenarios revealed that for the false accusation, none of the participants identified the accused as having done something wrong, and in the trivial accusation children barely expressed any negative judgment (see Section S3.1 of supplementary material). Analyses presented below therefore exclude the two control scenarios, which served their purpose by demonstrating that participants could distinguish substantive accusations from false or trivial accusations.

Results

Method validation: Setting method and believability

The setting method did not have any effect on any of the key dependent variables (all *ps* > .100) (see Tables S4–S11), suggesting no important differences between conducting the experiment remotely over the internet and conducting it face-to-face. In total, 88% (95% CI [82, 94]) of children believed that the events shown had actually happened on the Squidcraft server.

Third-party interventions: Compensation and punishment

Overall, during justice administration, children expressed preferences for punishment; in 57% (95% CI [53, 61]) of the test trials, children endorsed punishment over compensation when asked to choose. Moreover, punishment versus compensation endorsement during justice administration was not affected by children's age, $\chi^2(1) = 2.68$, $p = .102$, $\Delta R^2 = .006$, odds ratio (OR) = 1.16, 95% CI [0.97, 1.35] (Q3 in Table 1; see Table S6 for the full model for this variable). In contrast, judgment of transgression severity was a significant predictor, $\chi^2(1) = 4.50$, $p = .034$, $\Delta R^2 = .008$ (Q1 in Table 1). Specifically, the more severely children judged the transgressions, the more likely they were to endorse punishment over compensation, OR = 0.85, 95% CI [0.73, 0.98]. Transgression type was also a significant predictor, $\chi^2(1) = 10.73$, $p = .001$, $\Delta R^2 = .032$ (Q2 in Table 1).

Most transgression types did not elicit preferential endorsement of either punishment or compensation; the two exceptions for which punishment was clearly the favorite option were for transgressions related to sanctity/authority and liberty/deception (Fig. 3). There were no transgression types eliciting a preference for compensation. The significant effect of transgression type is explicable with reference to the observation that theft elicited lower endorsement of punishment over compensation than what would be expected from the judgment of the severity of this transgression (Fig. 3).

After justice administration children's preferences for punishment were confirmed, with 60% (95% CI [51, 68]) of children endorsing punishment over compensation in the forced-choice task. This was affected by age, $\chi^2(1) = 7.59$, $p = .006$, $\Delta R^2 = .072$ (Q3 in Table 1; see Table S7 for the full model for this variable); the older the children, the more likely they were to endorse punishment over compensation, OR = 1.58, 95% CI [1.14, 2.25].

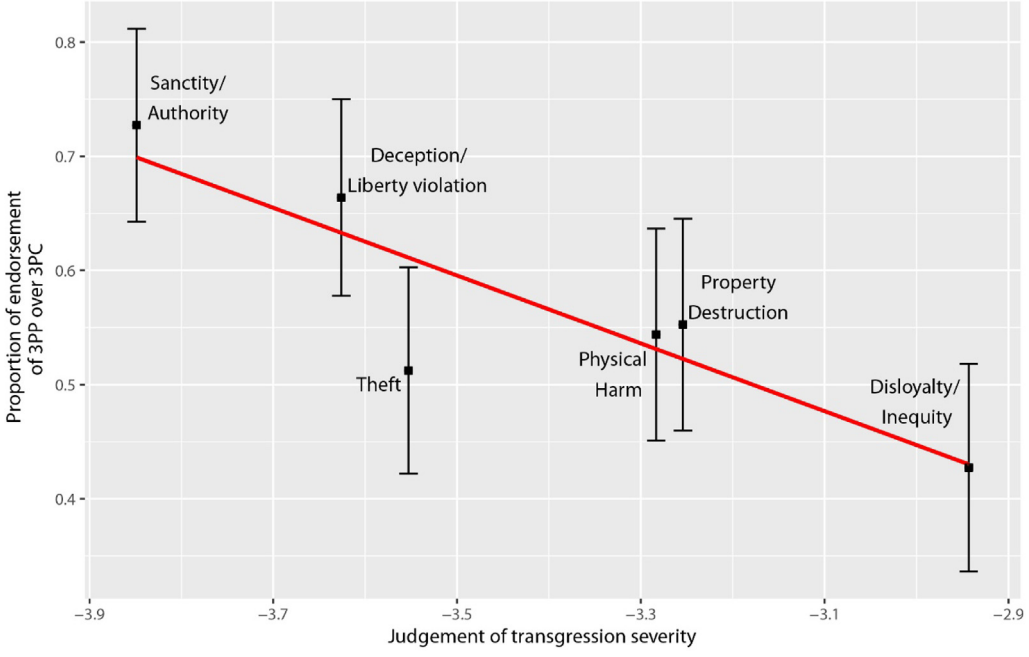

Fig. 3. Proportions of endorsement of punishment over compensation in relation to judgment of transgression severity across transgression types. 3PP, punishment; 3PC, compensation. In judgment of transgression severity, more negative numbers indicate more severe judgments. 95% confidence intervals are shown for each transgression type; the regression line is based on the proportions for each transgression type.

Affective states induced by third-party interventions

Children's enjoyment was predicted by type of third-party intervention, $\chi^2(1) = 76.38$, $p < .001$, $\Delta R^2 = .093$ (Q4 in Table 1; see Table S8 for the full model), with compensation eliciting more enjoyment ($M = 3.31$, $SD = 1.34$) than punishment ($M = 2.27$, $SD = 1.96$), $d = 0.61$, 95% CI [0.48, 0.74] (Fig. 4). Punishment enjoyment was predicted by time, $\chi^2(1) = 7.19$, $p = .007$, $\Delta R^2 = .015$ (Q5 in Table 1; see Table S9 for the full model); punishment enjoyment was lower when measured after justice administration ($M = 2.04$, $SD = 2.31$) than when measured during justice administration ($M = 2.50$, $SD = 1.53$), $d = -0.24$, 95% CI [−0.42, −0.07].

Compensation enjoyment was also predicted by time, $\chi^2(1) = 19.81$, $p < .001$, $\Delta R^2 = .035$ (Q5 in Table 1; see Table S10 for the full model), but the temporal pattern was different from punishment enjoyment; compensation enjoyment was higher after justice administration ($M = 3.59$, $SD = 1.44$) than during justice administration ($M = 3.05$, $SD = 1.17$), $d = 0.37$, 95% CI [0.21, 0.53].

Punishment motives and justifications: Deterrence and retribution

During justice administration, punishment severity did not change across framing conditions, $\chi^2(1) = 2.26$, $p = .324$, $\Delta R^2 = .008$ (Q6 in Table 1; see Table S4 for the full model for this variable). For the important retribution versus deterrence comparison, $d = 0.16$, 95% CI [−0.12, 0.43] (a positive value means higher punishment severity in the deterrence frame), indicating the possibility of a small undetected effect. After justice administration, 88% (95% CI [80, 93]) of children endorsed deterrence over retribution in the forced-choice task. This overwhelming endorsement of deterrence did not vary as a function of age, $\chi^2(1) = 0.18$, $p = .668$, $\Delta R^2 = .002$, OR = 0.91, 95% CI [0.57, 1.43] (Q7 in Table 1; see Table S11 for the full model for this variable).

Given the lack of framing effects on punishment severity (as well as on all other key variables), we examined whether there were framing effects on children's answers to the framing manipulation checks (Q8 in Table 1), which assessed children's understanding of the Justice System's presented purpose (a post hoc investigation). Children's mentions of retribution, deterrence, and compensation significantly varied across framing conditions (all $ps \leq .001$ at Checks 1 and 2, Fisher's exact tests) (Fig. 5), indicating that the framing manipulation did affect children's understanding of the Justice System's purpose. Specifically, mentions of retribution were more common in the retribution condition than in the compensation condition (Checks 1 and 2, $ps < .010$) and deterrence condition (Checks 1 and 2, $ps < .050$). Mentions of deterrence were more common in the deterrence condition than in the

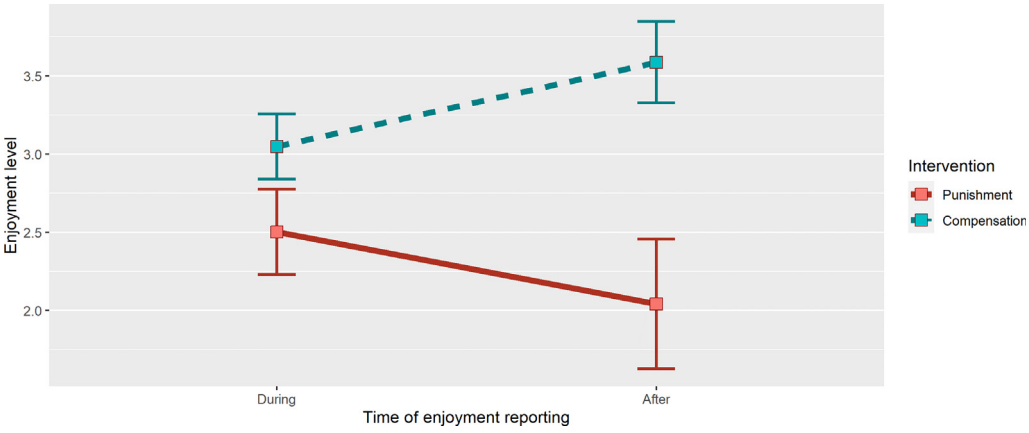

Fig. 4. Enjoyment level reported at different time points for the two types of third-party intervention. Error bars represent 95% confidence intervals of means. Note that enjoyment after justice administration (one data point per individual) was compared with enjoyment during justice administration (data points averaged over each individual's 6 trials).

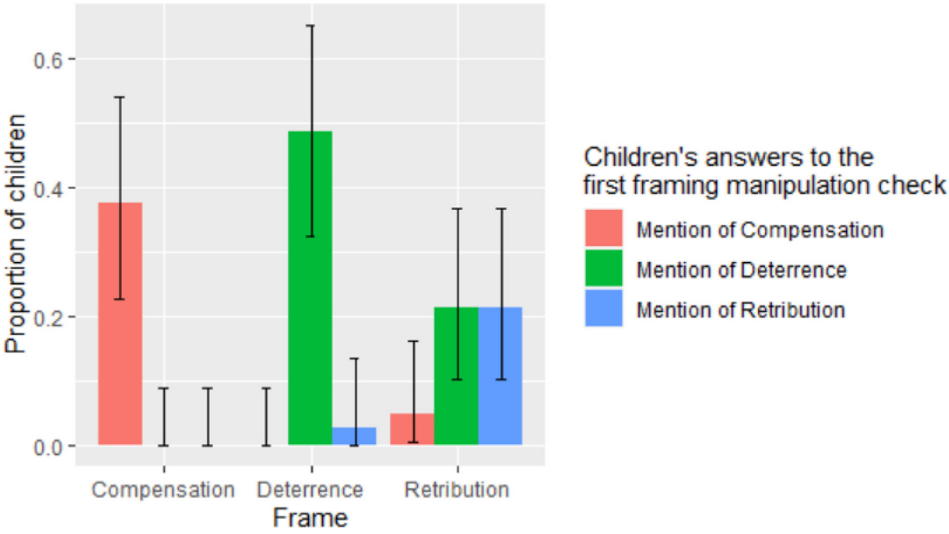

Fig. 5. Proportions of children who mentioned compensation, deterrence, and retribution across framing conditions at the first framing manipulation check (responses at the second manipulation check were very similar; see Fig. S2 in supplementary material). Error bars represent 95% confidence intervals of proportions.

compensation condition (Checks 1 and 2, *p*s < .001) and retribution condition (Checks 1 and 2, *p*s < .050). Mentions of compensation were more common in the compensation condition than in the deterrence condition (Checks 1 and 2, *p*s < .001) and retribution condition (Checks 1 and 2, *p*s < .010).

A salient observation is that deterrence was more commonly mentioned than retribution (in line with the results for deterrence vs. retribution endorsement after justice administration). For example, retribution was almost never mentioned by participants in the deterrence frame condition, whereas deterrence was mentioned as often as retribution by participants in the retribution frame condition (Fig, 5). Intercoder reliability was good with regard to the proportion of children mentioning each motivation (see Section S4.4 in supplementary material).

Discussion

Our research advanced knowledge about important but relatively uninvestigated topics in developmental psychology: the modulating factors of children's orientation for punishment or compensation, the emotional consequences of enacting punishment and compensation, and the justifications and motives (deterrence vs. retribution) behind children's decisions to inflict punishment.

Interestingly, high levels of endorsement of deterrence over retribution were found irrespective of children's age, confirming that it is normal for children (as well as adults) to conceive of punishment as being for deterrence (Bregant et al., 2016; Dunlea & Heiphetz, 2021; Marshall et al., 2022; Stern & Peterson, 1999; Yudkin et al., 2020). However, because children, relative to adults, have less developed inhibitory control and forward-looking reasoning skills, it seems less likely that their high levels of deterrence endorsement result from sophisticated deliberative processes (as discussed by Bregant et al., 2016). The observed tendency of children to default to deterrence-based explanations of punishment raises the possibility that deterrence reasoning is more intuitive, which would contravene some assumptions in the adult literature (Aharoni & Fridlund, 2012; Carlsmith & Darley, 2008; Keller et al., 2010; but see Rehren & Zisman, 2022). However, this should be tested with children younger than those in our sample.

Moreover, contrary to our predictions, whether the punishment frame was deterrence or retribution had no effect on children's punishment severity. This suggests that, in this context, children tend

to act according to their preconceived notions of what is right, not merely what they are told should be done. Importantly, we can rule out that these null effects are due to the total ineffectiveness of our framing manipulation; we did observe framing effects in the manipulation check. The frequency of children's mentions of retribution, deterrence, and compensation depended on the framing condition, indicating that children tended to remember the experimenter's framing explanations; they did not purely report their preexisting justice beliefs. However, a further salient result from the manipulation checks was that children assigned to punishment frames were biased toward explaining the Justice System in terms of deterrence, showing that the preference for deterrence over retribution evident in the forced-choice endorsement task generalized to this open-ended measure. These results are also in line with previous work showing that children are sensitive to deterrence cues (Marshall et al., 2021; Twardawski & Hilbig, 2020). We note further that our between-participant statistical power was lower than ideal, and the confidence interval for the size of the effect on punishment severity of deterrence versus retribution frames allows for a small undetected effect; children might punish slightly more given a deterrence frame in this context.

It is possible that explaining the experiment to children as a Justice System where they would take the role of judges affected our results (Gonzalez-Gadea et al., 2022); this may have geared children toward remembering and endorsing more deterrence-based explanations for punishment. However, it has also been demonstrated that children explain punishment in terms of deterrence irrespective of social roles given that they overwhelmingly attribute deterrent motives to institutional and peer punishers alike (Marshall et al., 2022). It remains to be clarified whether children's endorsement of deterrence in our experiment was due to the high frequency at which children in everyday life are familiarized with pedagogical uses of punishment (Marshall et al., 2022), an experience that would promote internalization of deterrent messages.

With respect to children's affective states related to third-party interventions, it was found that compensation elicited more enjoyment than punishment, as predicted. This could be due to a "warm glow" effect deriving from the experience of giving to people in need (Andreoni, 1990). From an early age, children show sympathetic behavior toward victims of transgressions (Vaish et al., 2009) and are motivated to see others get the help they need (Hepach et al., 2016). Moreover, both compensation and punishment enjoyment were time dependent but followed different temporal patterns; compensation enjoyment increased, whereas punishment enjoyment declined over time. The decrease in punishment enjoyment is unlikely to be due to emotional memory extinction (LaBar & Cabeza, 2006) because the same process would probably have governed compensation enjoyment too, which instead showed an increase over time. Whereas the temporal pattern of compensation enjoyment might be indicative of children's positive reappraisal of the impact of their action on the victims, the temporal pattern of punishment enjoyment is in accordance with Carlsmith et al.'s (2008) finding that enacting punishment causes rumination and thus lowering of mood. It is also possible that the decrease of punishment enjoyment over time might also indicate that children experienced a social desirability bias to show regret.

Children enjoyed compensating victims more than punishing transgressors, yet they preferentially endorsed punishment over compensation (after justice administration and with increasing judgment severity during justice administration). This misalignment between affective states and endorsements suggests that children see punishment as a duty to fulfil even if unpleasant. This sense of duty might arise in part because of demand characteristics of the situation (being repeatedly asked to contemplate allocated punishment and compensation), but this does not explain the mismatch between affective states and endorsement. Whether this duty is about meeting the retributive goal (i.e., transgressors' suffering) or the deterrent goal (i.e., transgressors learning their lesson) is difficult to establish from our affective results given that previous literature demonstrated that both are linked to higher punishers' satisfaction (Aharoni et al., 2022). However, a retributive explanation for children's punishment seems unlikely given our findings that children overwhelmingly endorsed deterrence irrespective of age and had a recollection of the purpose of the Justice System biased toward deterrence.

Children's punishment-related affective states were positive, albeit not as positive as compensation-related affective states. Thus, this finding is in contrast to a previous study's finding that children usually did not enjoy enacting their punishment decisions (Arini et al., 2021). The

specificities of the different experiments might account for these contrasting results. In the paradigm used by Arini et al. (2021), the allocation of punishment to the transgressor was more visually and auditorily salient for the participant than in the Minecraft paradigm, which might have elicited a sense of compassion for the punished transgressor. Moreover, in Arini et al.'s experiments, children did not have previous experience with the game, whereas participants in the Minecraft experiment were familiar with Minecraft and generally enjoyed playing and thinking about it. Most important, in Arini et al.'s paradigm children could only assign punishment, whereas in the Minecraft paradigm they could both punish transgressors and compensate victims, thereby contributing to a greater overall sense of justice being restored.

As expected, the seriousness of a transgression influenced children's third-party interventions. Analyses showed that children preferred to endorse punishment over compensation during justice administration when they judged transgressions more severely. Furthermore, children's endorsement of punishment versus compensation during justice administration was affected by transgression type. Contrary to our predictions, children did not preferentially endorse punishment in cases of harm violation (Miller & McCann, 1979) and compensation in cases of theft (Liu et al., 2021; Riedl et al., 2015; Yang et al., 2021) and inequity (Lee & Warneken, 2020). Transgressions in sanctity/authority and liberty/deception were in fact the only contexts eliciting clear preferential endorsement of punishment; for the majority of transgression types, children did not express preferential endorsement of either punishment or compensation during justice administration.

To note, the sanctity/authority transgression (i.e., killing a squid, which is usually allowed but not on this particular Minecraft server because of its status as a holy animal) was the scenario prompting the highest rates of punishment endorsement. Whereas all the other moral norms such as killing players and stealing were independent of this Minecraft server, this norm was novel and unique. This may relate to the particular importance of sacred ingroup values (Tetlock, 2003) and is telling of the malleable nature of children's norm learning (Rakoczy et al., 2008) and of the volatility of moral norms on the internet. This may also speak to the strong potential for information associated with religious contexts to be quickly accepted by children (Vaden & Woolley, 2011).

Interestingly, theft elicited higher endorsement of compensation over punishment than expected on the basis of judgment of transgression severity. To explain this result, we refer to the evidence that children use punishment as an opportunity to directly right the wrong when it concerns resources (Arini et al., 2021). In the current experimental setting, punishment (i.e., banning the transgressor from the game) was not suitable for equalizing the resource imbalance between victim and transgressor after a theft, so children reacted to this scenario by instead using compensation (i.e., giving diamonds to the victim) to fulfil their equalization purposes.

Whereas several transgression types did not cause preferential endorsement of punishment over compensation during justice administration, a clear preference for endorsement of punishment was observed after justice administration. This result could be because a generic and abstract sense that moral transgressions have occurred in the past might elicit children's intuitive reaction to endorse punishment. On the other hand, requiring children to attend in real time to the details of the different transgression types might induce them to engage in careful deliberation about whether it is more appropriate to endorse punishment or compensation (Van Doorn & Brouwers, 2017). The preferential endorsement of punishment after justice administration is also indicative of children's transgressor-centered approach to justice restoration. This is in accordance with the other behavioral studies in the developmental literature where punishment was preferred over compensation when the latter was operationalized as any prosocial actions toward the victim (McAuliffe & Dunham, 2021) other than restitution of previously subtracted resources (Riedl et al., 2015; Yang et al., 2021). This is also in accordance with the studies in the adult literature where third-party interventions were not costly to the participants (Adams & Mullen, 2015; van Prooijen, 2010), as in our experiment. In contrast, studies with adult participants providing evidence for preference for compensation commonly employed paradigms where third parties' economic resources were at stake (Chavez & Bicchieri, 2013; Lotz et al., 2011; Van Doorn et al., 2018a, 2018b).

Finally, regarding the investigation of developmental patterns, as hypothesized on the basis of the studies conducted by Riedl et al. (2015) and Miller and McCann (1979), we observed a developmental increase in the proportion of children endorsing punishment over compensation after justice

administration (a finding consistent with later studies by Yang et al., 2021, and McAuliffe & Dunham, 2021). This points to the possibility that, although children are willing to punish transgressors from an early age (e.g., Kenward & Östh, 2012, 2015), attitudes toward this type of third-party intervention are further subject to learning processes. Because third-party punishment is administered in schools and households, children may learn that punishment is a socially approved choice (Marshall et al., 2022), and indeed there is evidence that young children attribute reputational benefits to punishing (Vaish et al., 2016; but see Dhaliwal et al., 2021, Eriksson et al., 2016, and Raihani & Bshary, 2015b, for opposite examples in adults).

Regarding the limitations of our study, as noted above, some effect size confidence intervals reflect the possibility of undetected effects due to lower than ideal power, and further the samples of recruited children were not necessarily representative of the respective national populations. Children tested face-to-face while attending science-themed fairs and summer camps came from households characterized by higher education and socioeconomic conditions than the national average. In comparison, online testing was more able to reach children of diverse backgrounds. The internal cultural diversity of our sample is in some ways a strength, but the relatively small sample from each country (in part due to greater difficulties with internet recruiting than expected) also represents a weakness. We were unable to fully counterbalance the setting method across different countries. Only British children were tested in both settings, whereas Colombian children were tested exclusively face-to-face and Italian children were tested exclusively over the internet. Having said that, the statistical analyses we conducted always controlled for children's country of residence and setting method, and our aim to test a broad range of children in terms of nationalities was mainly motivated by the desire to maximize the chances of detecting common patterns of moral behavior rather than cross-cultural differences.

Future avenues for investigating children's third-party interventions should take advantage of multiple methodologies. Qualitative and possibly longitudinal studies could provide a more detailed insight into the development of children's concepts about punishment justifications. From interviews with children and their parents and teachers, it would also be possible to discern to what extent children's beliefs about punishment are affected by their familiarity with deterrent justifications in the family and school settings (Sorbring et al., 2006). Questionnaire studies could shed light on personality differences in children endorsing punishment versus compensation and retribution versus deterrence. Finally, experimental studies could complement the picture by measuring children's affective states in three different conditions: when they are given only the opportunity to enact punishment of transgressors, when they are given only the opportunity to compensate victims, and when they can choose to either punish or compensate.

Before concluding, we briefly address methodological implications. To our knowledge, this is the first study in which developmental psychologists have tested children in a virtual environment (i.e., a Justice System) relating to an online game of their choice by making use of video chat and voice call applications. The lack of any differences in the key variables depending on whether children were tested over the internet or face-to-face and the high rates at which children believed they were enacting interventions with real consequences for victims and transgressors together provide evidence that our innovative computer-mediated paradigm has the potential to fundamentally change the practicalities of collecting some types of behavioral data. However, care must be taken not to overgeneralize this result to different studies. Furthermore, although online testing is arguably more scalable than in-person testing, online recruitment of children is not without its own specific set of challenges (e.g., parents might have lower trust in researchers they have not met). Future studies therefore should systematically investigate different online recruitment methods in order to identify the most effective ones.

Conclusion

We demonstrated that children overwhelmingly reported deterrence as their punishment justification across a wide age range. In addition, the more severely children judged the transgressions, the more they endorsed punishment over compensation during justice administration, revealing a transgressor-centered approach to justice restoration. Moreover, even though children explicitly

endorsed punishment over compensation after justice administration, they derived higher enjoyment from compensating victims than from punishing transgressors. Finally, whereas compensation enjoyment increased, punishment enjoyment decreased over time. Results from enjoyment, endorsement, and justification measures together suggest that children enacting punishment behavior are motivated by a sense that they *ought* to punish to fulfill social obligations and achieve deterrence. How children's punitive sentiment becomes more motivated by retribution, as it is in adults, is a question that awaits future investigation.

Data availability

Data has been submitted with the manuscript

Acknowledgments

We thank Morag MacLean and Michaela Gummerum for their precious feedback on an early draft of the manuscript, all the families who took part in this project, and the Minecraft gaming clubs that supported it. In addition, we thank Juan Sebastian Nassar Pereira for assistance in data collection in Colombia, and Salvatore Arini for assistance in the recruitment of participants in Italy. This work was supported by the Nigel Groome Studentship and Santander Research Scholarship (both awarded to Rhea L. Arini) and internal research funding from the Faculty of Social Sciences, Universidad de los Andes, Colombia (awarded to Gordon P. D. Ingram).

Data availability

Data, script, stimuli, and analysis pipelines are available at the Open Science Framework (https://osf.io/4ygw5).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2023.105630.

References

Adams, G. S., & Mullen, E. (2015). Punishing the perpetrator decreases compensation for victims. *Social Psychological and Personality Science, 6*(1), 31–38. https://doi.org/10.1177/1948550614542346.

Aharoni, E., & Fridlund, A. J. (2012). Punishment without reason: Isolating retribution in lay punishment of criminal offenders. *Psychology, Public Policy, and Law, 18*(4), 599–625. https://doi.org/10.1037/a0025821.

Aharoni, E., Simpson, D., Nahmias, E., & Gollwitzer, M. (2022). A painful message: Testing the effects of suffering and understanding on punishment judgments. *Zeitschrift für Psychologie, 230*(2), 138–151. https://doi.org/10.1027/2151-2604/a000460.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal, 100*(401), 464–477. https://doi.org/10.2307/2234133.

Arini, R. L., Wiggs, L., & Kenward, B. (2021). Moral duty and equalization concerns motivate children's third-party punishment. *Developmental Psychology, 57*(8), 1325–1341. https://doi.org/10.1037/dev0001191.

Beale, M., McKittrick, M., & Richards, D. (2016). "Good" grief: Subversion, praxis, and the unmasked ethics of griefing guides. *Technical Communication Quarterly, 25*(3), 191–201. https://doi.org/10.1080/10572252.2016.1185160.

Bentham, J. (1948). *A fragment on government and an introduction to the principles of morals and legislation* (W. Harrison, Ed.). Macmillan. (Original work published 1789)

Berman, M. N. (2010). *Two kinds of retributivism* (Public Law Research Paper No. 171). University of Texas School of Law.

Bregant, J., Shaw, A., & Kinzler, K. D. (2016). Intuitive jurisprudence: Early reasoning about the functions of punishment. *Journal of Empirical Legal Studies, 13*(4), 693–717. https://doi.org/10.1111/jels.12130.

Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology, 42*(4), 437–451. https://doi.org/10.1016/j.jesp.2005.06.007.

Carlsmith, K. M. (2008). On justifying punishment: The discrepancy between words and actions. *Social Justice Research, 21*(2), 119–137. https://doi.org/10.1007/s11211-008-0068-x.

Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. *Advances in Experimental Social Psychology, 40*, 193–236. https://doi.org/10.1016/S0065-2601(07)00004-4.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*(2), 284–299. https://doi.org/10.1037/0022-3514.83.2.284.
Carlsmith, K. M., Wilson, T. D., & Gilbert, D. T. (2008). The paradoxical consequences of revenge. *Journal of Personality and Social Psychology, 95*(6), 1316–1324. https://doi.org/10.1037/a0012165.
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.
Condon, P., & DeSteno, D. (2011). Compassion for one reduces punishment for another. *Journal of Experimental Social Psychology, 47*(3), 698–701. https://doi.org/10.1016/j.jesp.2010.11.016.
Crockett, M. J., Özdemir, Y., & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *Journal of Experimental Psychology: General, 143*(6), 2279–2286. https://doi.org/10.1037/xge0000018.
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.
Dunlea, J. P., & Heiphetz, L. (2021). Children's and adults' views of punishment as a path to redemption. *Child Development, 92*(4), e398–e415. https://doi.org/10.1111/cdev.13475.
Eriksson, K., Andersson, P. A., & Strimling, P. (2016). Moderators of the disapproval of peer punishment. *Group Processes & Intergroup Relations, 19*(2), 152–168. https://doi.org/10.1177/1368430215583519.
Fraser, O. N., Koski, S. E., Wittig, R. M., & Aureli, F. (2009). Why are bystanders friendly to recipients of aggression? *Communicative & Integrative Biology, 2*(3), 285–291. https://doi.org/10.4161/cib.2.3.8718.
Fry, D. P. (2000). Conflict management in cross-cultural perspective. In F. Aureli & F. B. M. de Waal (Eds.), *Natural conflict resolution* (pp. 334–351). University of California Press.
Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the message: Punishment is satisfying if the transgressor responds to its communicative intent. *Personality and Social Psychology Bulletin, 40*(8), 986–997. https://doi.org/10.1177/0146167214533130.
Funk, F., & Mischkowski, D. (2022). Examining consequentialist punishment motives in one-shot social dilemmas. *Zeitschrift für Psychologie, 230*(2), 127–137. https://doi.org/10.1027/2151-2604/a000459.
Genevsky, A., Västfjäll, D., Slovic, P., & Knutson, B. (2013). Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. *Journal of Neuroscience, 33*(43), 17188–17196. https://doi.org/10.1523/JNEUROSCI.2348-13.2013.
Gollwitzer, M., & Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology, 45*(4), 840–844. https://doi.org/10.1016/j.jesp.2009.03.001.
Gollwitzer, M., Meder, M., & Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology, 41*(3), 364–374. https://doi.org/10.1002/ejsp.782.
Gonzalez-Gadea, M. L., Dominguez, A., & Petroni, A. (2022). Decisions and mechanisms of intergroup bias in children's third-party punishment. *Social Development, 31*(4), 1194–1210. https://doi.org/10.1111/sode.12608.
Gordon, D. S., Madden, J. R., & Lea, S. E. (2014). Both loved and feared: Third party punishers are viewed as formidable and likeable, but these reputational benefits may only be open to dominant individuals. *PLoS One, 9*(10). https://doi.org/10.1371/journal.pone.0110045. Article e110045.
Gummerum, M., López-Pérez, B., Van Dijk, E., & Van Dillen, L. F. (2020). When punishment is emotion-driven: Children's, adolescents', and adults' costly punishment of unfair allocations. *Social Development, 29*(1), 126–142. https://doi.org/10.1111/sode.12387.
Gummerum, M., López-Pérez, B., Van Dijk, E., & Van Dillen, L. F. (2022). Ire and punishment: Incidental anger and costly punishment in children, adolescents, and adults. *Journal of Experimental Child Psychology, 218*. https://doi.org/10.1016/j.jecp.2022.105376 105376.
Gummerum, M., Van Dillen, L. F., Van Dijk, E., & López-Pérez, B. (2016). Costly third-party interventions: The role of incidental anger and attention focus in punishment of the perpetrator and compensation of the victim. *Journal of Experimental Social Psychology, 65*, 94–104. https://doi.org/10.1016/j.jesp.2016.04.004.
Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the United States of America, 108*(50), 19931–19936. https://doi.org/10.1073/pnas.1110306108.
Harbaugh, W. T., Mayr, U., & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science, 316*(5831), 1622–1625. https://doi.org/10.1126/science.1140738.
Hepach, R., Vaish, A., Grossmann, T., & Tomasello, M. (2016). Young children want to see others get the help they need. *Child Development, 87*(6), 1703–1714. https://doi.org/10.1111/cdev.12633.
Hu, Y., Strang, S., & Weber, B. (2015). Helping or punishing strangers: Neural correlates of altruistic decisions as third-party and of its relation to empathic concern. *Frontiers in Behavioral Neuroscience, 9*, Article 24. https://doi.org/10.3389/fnbeh.2015.00024.
Johnstone, G., & Van Ness, D. (Eds.). (2013). *Handbook of restorative justice*. Routledge.
Kant, I. (1952). *The critique of judgement* (J. C. Meredith, Trans.). Clarendon. (Original work published 1790).
Keller, L. B., Oswald, M. E., Stucki, I., & Gollwitzer, M. (2010). A closer look at an eye for an eye: Laypersons' punishment decisions are primarily driven by retributive motives. *Social Justice Research, 23*(2–3), 99–116. https://doi.org/10.1007/s11211-010-0113-4.
Kenward, B., & Östh, T. (2012). Enactment of third-party punishment by 4-year-olds. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00373. Article 373.
Kenward, B., & Östh, T. (2015). Five-year-olds punish antisocial adults. *Aggressive Behavior, 41*(5), 413–420. https://doi.org/10.1002/ab.21568.
Kou, Y., Johansson, M., & Verhagen, H. (2017). In *Prosocial behavior in an online game community: An ethnographic study* (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3102071.3102078.
LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience, 7*(1), 54–64. https://doi.org/10.1038/nrn1825.
Lee, Y. E., & Warneken, F. (2020). Children's evaluations of third-party responses to unfairness: Children prefer helping over punishment. *Cognition, 205*. https://doi.org/10.1016/j.cognition.2020.104374 104374.

Liu, X., Yang, X., & Wu, Z. (2021). To punish or to restore: How children evaluate victims' responses to immorality. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.696160 696160.
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.
Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., ... Tracer, D. (2008). More "altruistic" punishment in larger societies. *Proceedings of the Royal Society B: Biological Sciences, 275*(1634), 587–592. https://doi.org/10.1098/rspb.2007.1517.
Marshall, J., Gollwitzer, A., & Bloom, P. (2022). Why do children and adults think other people punish? *Developmental Psychology, 58*(9), 1783–1792. https://doi.org/10.1037/dev0001378.
Marshall, J., & McAuliffe, K. (2022). Children as assessors and agents of third-party punishment. *Nature Reviews Psychology, 1*, 334–344. https://doi.org/10.1038/s44159-022-00046-y.
Marshall, J., Yudkin, D. A., & Crockett, M. J. (2021). Children punish third parties to satisfy both consequentialist and retributive motives. *Nature Human Behaviour, 5*(3), 361–368. https://doi.org/10.1038/s41562-020-00975-9.
McAuliffe, K., & Dunham, Y. (2021). Children favor punishment over restoration. *Developmental Science, 24*(5). https://doi.org/10.1111/desc.13093. Article e13093.
Miller, D. T., & McCann, C. D. (1979). Children's reactions to the perpetrators and victims of injustices. *Child Development, 50*, 861–868. https://doi.org/10.2307/1128955.
Molho, C., Twardawski, M., & Fan, L. (2022). What motivates direct and indirect punishment? Extending the "intuitive retributivism" hypothesis. *Zeitschrift für Psychologie, 230*(2), 84–93. https://doi.org/10.1027/2151-2604/a000455.
Nockur, L., Kesberg, R., Pfattheicher, S., & Keller, J. (2022). Why do we punish? On retribution, deterrence, and the moderating role of punishment system. *Zeitschrift für Psychologie, 230*(2), 104–113. https://doi.org/10.1027/2151-2604/a000457.
Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2012). To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice. *Evolution and Human Behavior, 33*(6), 682–695. https://doi.org/10.1016/j.evolhumbehav.2012.05.003.
R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org.
Raihani, N. J., & Bshary, R. (2015a). The reputation of punishers. *Trends in Ecology & Evolution, 30*(2), 98–103. https://doi.org/10.1016/j.tree.2014.12.003.
Raihani, N. J., & Bshary, R. (2015b). Third-party punishers are rewarded, but third-party helpers even more so. *Evolution, 69*(4), 993–1003. https://doi.org/10.1111/evo.12637.
Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology, 44*(3), 875–881. https://doi.org/10.1037/0012-1649.44.3.875.
Rehren, P., & Zisman, V. (2022). Testing the intuitive retributivism dual process model. *Zeitschrift für Psychologie, 230*(2), 152–163. https://doi.org/10.1027/2151-2604/a000461.
Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America, 109*(37), 14824–14829. https://doi.org/10.1073/pnas.1203179109.
Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2015). Restorative justice in children. *Current Biology, 25*(13), 1731–1735. https://doi.org/10.1016/j.cub.2015.05.014.
Singh, M., & Garfield, Z. H. (2022). Evidence for third-party mediation but not punishment in Mentawai justice. *Nature Human Behaviour, 6*, 930–940. https://doi.org/10.1038/s41562-022-01341-7.
Sorbring, E., Deater-Deckard, K., & Palmérus, K. (2006). Girls' and boys' perception of mothers' intentions of using physical punishment and reasoning as discipline methods. *European Journal of Developmental Psychology, 3*(2), 142–162. https://doi.org/10.1080/17405620500398748.
Stern, B. L., & Peterson, L. (1999). Linking wrongdoing and consequence: A developmental analysis of children's punishment orientation. *Journal of Genetic Psychology, 160*(2), 205–224. https://doi.org/10.1080/00221329909595393.
Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., & Kirsch, P. (2011). Beyond revenge: Neural and genetic bases of altruistic punishment. *NeuroImage, 54*(1), 671–680. https://doi.org/10.1016/j.neuroimage.2010.07.051.
Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences, 7*(7), 320–324. https://doi.org/10.1016/S1364-6613(03)00135-9.
Twardawski, M., & Hilbig, B. E. (2020). The motivational basis of third-party punishment in children. *PLoS One, 15*(11). https://doi.org/10.1371/journal.pone.0241919. Article e241919.
Vaden, V. C., & Woolley, J. D. (2011). Does God make it real? Children's belief in religious stories from the Judeo-Christian tradition. *Child Development, 82*(4), 1120–1135. https://doi.org/10.1111/j.1467-8624.2011.01589.x.
Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology, 45*(2), 534–543. https://doi.org/10.1037/a0014322.
Vaish, A., Herrmann, E., Markmann, C., & Tomasello, M. (2016). Preschoolers value those who sanction non-cooperators. *Cognition, 153*, 43–51. https://doi.org/10.1016/j.cognition.2016.04.011.
Van Doorn, J., & Brouwers, L. (2017). Third-party responses to injustice: A review on the preference for compensation. *Crime Psychology Review, 3*(1), 59–77. https://doi.org/10.1080/23744006.2018.1470765.
Van Doorn, J., Zeelenberg, M., & Breugelmans, S. M. (2018a). An exploration of third parties' preference for compensation over punishment: Six experimental demonstrations. *Theory and Decision, 85*(3–4), 333–351. https://doi.org/10.1007/s11238-018-9665-9.
Van Doorn, J., Zeelenberg, M., Breugelmans, S. M., Berger, S., & Okimoto, T. G. (2018b). Prosocial consequences of third-party anger. *Theory and Decision, 84*(4), 585–599. https://doi.org/10.1007/s11238-017-9652-6.
van Prooijen, J. W. (2010). Retributive versus compensatory justice: Observers' preference for punishing in response to criminal offenses. *European Journal of Social Psychology, 40*(1), 72–85. https://doi.org/10.1002/ejsp.611.
Will, G.-J., Crone, E. A., van den Bos, W., & Güroğlu, B. (2013). Acting on observed social exclusion: Developmental perspectives on punishment of excluders and compensation of victims. *Developmental Psychology, 49*(12), 2236–2244. https://doi.org/10.1037/a0032299.

Yang, X., Wu, Z., & Dunham, Y. (2021). Children's restorative justice in an intergroup context. *Social Development, 30*(3), 663–683. https://doi.org/10.1111/sode.12508.
Yudkin, D. A., Van Bavel, J. J., & Rhodes, M. (2020). Young children police group members at personal cost. *Journal of Experimental Psychology: General, 149*(1), 182–191. https://doi.org/10.1037/xge0000613.