Original Article

QJEP

Quarterly Journal of Experimental Psychology
2025, Vol. 78(10) 2091–2109

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/17470218241310829
qjep.sagepub.com
Sage

Cognitive and affective processes in children's third-party punishment

Rhea L. Arini[1,2], Juliana Bocarejo Aljure[3], Nereida Bueno Guerra[4], Clara Bayón González[4], Estrella Fernández Alba[5], Natalia Suárez Fernández[5], Gordon P. D. Ingram[3,6], Luci Wiggs[2] and Ben Kenward[2,7]

Abstract

This study investigated how children's punishment affective states change over time, as well as when children begin to prioritise intentions over outcomes in their punishment decisions. Whereas most prior research sampled children from Anglo-America or Northwestern Europe, we tested 5- to 11-year-old children from Colombia and Spain ($N = 123$). We focused on punishment behaviour in response to ostensibly real moral transgressions, rather than punishment recommendations for hypothetical moral transgressions. We employed moral scenarios involving disloyalty (group-focused moral domain) and unfairness (individual-focused moral domain). Regarding punishment affective states, on average, children did not derive much enjoyment from administering punishment, nor did they anticipate that punishment would feel good. Thus, children did not make the same emotional forecasting error adults commonly commit. Regarding the cognitive integration of outcomes and intentions, children began to punish failed intentional transgressions more harshly than accidental transgression, in both disloyalty and unfairness scenarios, much earlier than in previous behavioural studies: around 7 years of age rather than in late adolescence. This could be due to the lower processing demands and higher intention salience of our paradigm. Exploratory analyses revealed that children showed higher concern for disloyalty than unfairness. Punishment of disloyalty remained relatively stable in severity with increasing age, while punishment of unfairness decreased in severity. This suggests that the relative importance of moral concerns for the individual vs. the group may shift because of culture-directed learning processes.

Keywords

Third-party punishment; moral domains; punishment affective states; punishment motives; outcome-to-intent shift

Received: 19 March 2024; revised: 22 October 2024; accepted: 29 October 2024

Introduction

Morality consists of a set of norms about how people should or should not behave (Janoff-Bulman et al., 2009). These norms, in turn, are the product of selective forces driving people to find solutions to the problems of cooperation that occur in social life (Curry et al., 2019). For cooperation to be maintained, moral norms need to be enforced. Norm enforcement can take two main forms: *second-party punishment* (2PP), that is, punishment of norm transgressors meted out by the victims; and *third-party punishment* (3PP), that is, punishment of norm transgressors by unaffected bystanders who act on behalf of the victims. Whereas second-party punishers correct the behaviour of transgressors essentially for personal

[1]School of Psychology, Sport Science, and Wellbeing, University of Lincoln, Lincoln, UK
[2]Centre for Psychological Research, Oxford Brookes University, Oxford, UK
[3]Department of Psychology, Universidad de los Andes, Bogotá, Colombia
[4]Department of Psychology, Universidad Pontificia Comillas, Madrid, Spain
[5]Faculty of Psychology, Universidad de Oviedo, Oviedo, Spain
[6]School of Science, Engineering & Tech, RMIT University Vietnam, H▯ Chí Minh, Vietnam
[7]Department of Psychology, Uppsala Universitet, Uppsala, Sweden

Corresponding author:
Rhea L. Arini, School of Psychology, Sport Science, and Wellbeing, University of Lincoln, 8 Brayford Wharf E, Lincoln LN5 7AT, UK.
Email: rarini@lincoln.ac.uk

benefits, third-party punishers pay a cost (particularly in terms of risk of counter-retaliation and breakdown of valuable social relationships) for the benefit of others (Jensen, 2010). 3PP has thus received a great deal of scientific attention given its arguably altruistic nature (Fehr & Gächter, 2002, but see the review by Raihani & Bshary, 2019). 3PP has been indicated as a key factor in sustaining the progressive establishment of large-scale cooperative networks in human societies (Boyd & Richerson, 1992). Indeed, population size and complexity of society have been shown to predict the level of 3PP (Marlowe et al., 2008).

From a developmental perspective, it has been shown that children are willing to enact 3PP from a very early age (as young as 19 months; Hamlin et al., 2011), in response to a range of norm transgressions (see the review by Marshall & McAuliffe, 2022). Children engage in 3PP even when it is costly to do so, whether costs are social (Kenward & Östh, 2015), emotional (Arini et al., 2021; Yudkin et al., 2020), or economic (Gonzalez-Gadea et al., 2022; McAuliffe et al., 2015; Yang et al., 2018). Children's 3PP decisions are driven by a variety of motives, concerns, and biases: deterrence of norm transgressors (Arini et al., 2023; Marshall et al., 2021; Twardawski & Hilbig, 2020); justice restoration (Arini et al., 2023; Riedl et al., 2015); equalisation concerns (Arini et al., 2021; Lee & Warneken, 2022); intergroup bias (Gonzalez-Gadea et al., 2022; Gummerum et al., 2009; Jordan et al., 2014); and conformity to a model (House et al., 2020; Salali et al., 2015). However, limited research has been conducted so far on the emotional experiences of children enacting 3PP, as well as on the cognitive integration between different types of information into children's 3PP decisions (reviewed below). This work is thus aimed at shedding light specifically on these two aspects.

Emotional factors in punishment

Research about the relation between punishment and emotions has been focused more on the emotions elicited by moral transgressions (which arguably motivate punishment), rather than on the emotions elicited by enacting or contemplating punishment. Regarding the former strand of research, it was found that in adults, preference for 2PP was associated with anger towards moral transgressions, whereas preference for 3PP was associated with disgust (Molho et al., 2017; Tybur et al., 2020). In addition, 3PP was predicted by compassion towards the victim (Pfattheicher et al., 2019), and moral outrage towards the transgressor (Ginther et al., 2022; Hartsough et al., 2020; Lotz et al., 2011). Thus, it seems that 2PP is consistently elicited by negative emotions, whereas 3PP can be elicited by both negative and positive emotions. A special case of punishment is represented by punishment of free riders by the cooperators in the group in a public goods game: since free riding targets both the self and other group members, punishment combines both 2PP and 3PP. There is evidence that this type of punishment is motivated by anger (Fehr & Gächter, 2002), similarly to 2PP.

Developmental studies on the emotions elicited by moral transgressions have focused on the role of anger and have demonstrated that the relation between anger and punishment depends on the interaction between punishers' age and the type of punishment they engage in (2PP vs. 3PP). It has been shown that violations of both fairness and trustworthiness elicited 2PP, and this relationship was mediated by anger, from childhood to adulthood (Gummerum et al., 2020; van den Bos et al., 2012). Violations of fairness also elicited 3PP, but this relationship was mediated by anger only in adults, not in children or adolescents (Gummerum et al., 2020). Finally, by experimentally manipulating anger, it was demonstrated that this emotion has a causal role in 2PP of unfairness in all age groups, whereas in 3PP, this occurs only in adults and adolescents, but not in children (Gummerum et al., 2022).

Regarding the emotions elicited by punishment (rather than moral transgressions), studies in the adult literature indicate that punishment is expected to be experienced as rewarding. Indeed, adults forecast that punishing uncooperative team members would make them feel better (Carlsmith et al., 2008). Moreover, people show activation in the striatum (a brain area implicated in reward) when determining the punishment for those who acted unfairly towards either them or others, suggesting that they anticipate satisfaction from punishment (De Quervain et al., 2004; Strobel et al., 2011). By contrast, research about the emotional consequences of punishment has produced quite mixed results: whereas some studies indicate that enacting punishment induces negative emotions, others suggest that it can elicit positive emotions under certain conditions. On the one hand, people who inflicted punishment reported feeling worse than individuals who had not been given the possibility to punish—an effect mediated by rumination about the transgression suffered (Carlsmith et al., 2008). On the other hand, seeing the transgressors suffer as a result of punishment has been shown to have a positive effect on punishers' satisfaction (Eder et al., 2020). However, seeing the transgressors acknowledge the wrongfulness of their actions had an even stronger effect on punishers' satisfaction (Aharoni et al., 2022; Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011), since this could be interpreted as a change in moral attitude (Funk et al., 2014). This evidence thus suggests that adults may perceive punishment as a hedonic experience depending on how transgressors react to being punished.

With respect to the developmental literature, some work suggests that in case of vicarious 2PP and when 3PP is paired with compensation of victims, children may derive enjoyment from punishment to a certain degree. For example, preschool children have been shown to be willing to incur costs to watch an agent that had previously mistreated them being punished by someone else (i.e., vicarious 2PP; Mendes et al., 2018). Although this could be interpreted as evidence that witnessing punishment is experienced as rewarding, the analysis of children's affective indicators depicts a more complex picture. Children showed a combination of both positive (i.e., smiles) and negative emotional expressions (i.e., frowns) while watching the punishment of the antisocial agent (Mendes et al., 2018), suggesting that they felt both pleasure and distress. Furthermore, when given the opportunity to themselves respond to transgressions affecting other people, primary school-aged children reported enjoying enacting 3PP of transgressors, although not as much as compensating victims (Arini et al., 2023). This may indicate that carrying out both types of behaviours contributes to children experiencing an overall sense of justice being restored and consequently enjoyment.

By contrast, in paradigms in which children could only decide whether to assign 3PP but not compensation, the emotional consequences of 3PP seem to be consistently negative. More specifically, children reported experiencing more sadness, less happiness, and less excitement when they engaged in 3PP than when they did not (see Supplementary Information in Marshall et al., 2021). In addition, children were more likely to report lack of enjoyment when they enacted real rather than pretend 3PP (Arini et al., 2021, Study 2). This suggests that children's affective states may be sensitive to the impact of 3PP on transgressors: only when they were really punishing, but not when they were just pretending to punish, could children have felt responsible for the suffering of the transgressor. Knowing to be the cause of someone else's suffering may be responsible for children's lack of punishment enjoyment. However, the fact that children enacted 3PP even though they did not find it enjoyable suggests that they may view 3PP as a moral duty to fulfil for the benefit of others (Arini et al., 2021).

Notably, differently from the procedure used with adults by Carlsmith et al. (2008), both Marshall et al. (2021, Supplementary Information) and Arini et al. (2021, Study 2) asked children to rate their emotions only after they had already assigned punishment. Therefore, these experimental paradigms did not rule out the possibility that children decided to carry out 3PP expecting it to be satisfying, yet they experienced low mood when their expectations were not met (similar to what has been found in adults; Carlsmith et al., 2008). To exclude this alternative explanation, in the present experiment, we investigated the temporal changes in 3PP affective states by asking children to report their affective states before, during, and after punishment allocation. We predicted that neither children's affective states during nor after punishment allocation would be positive, in line with the findings of Marshall et al. (2021, Supplementary Information) and Arini et al. (2021, Study 2). As for children's affective states before punishment allocation, we made no strong predictions. We hypothesised that, if children have hedonic expectations about punishment as adults do (Carlsmith et al., 2008; De Quervain et al., 2004; Strobel et al., 2011), affective states before punishment allocation would be more positive than those reported during and after punishment allocation. If instead the thought of carrying out 3PP in isolation consistently evokes negative emotions in children, affective states before punishment allocation would be no different compared to those reported during and after punishment allocation (Table 1, Q1). Moreover, we investigated whether children are sensitive to the impact of 3PP on transgressors (Arini et al., 2021, Study 2). We hypothesised that, if children are induced to think about the costs they impose on the transgressors with their 3PP decisions, they will experience lowering of their affective states due to feeling responsible for the suffering of the transgressors (Table 1, Q2).

Integration between outcomes and intentions in punishment

People's punishment decisions following a moral transgression are affected by the cognitive integration between different types of information: the *outcome* of the transgressor's action and the transgressor's *intention* behind such action. Importantly, adults tend to attribute more weight to intentions over outcomes, across different operationalisations of punishment and study methodologies (e.g., Barrett et al., 2016; Cushman, 2008; Gummerum & Chu, 2014; Hechler & Kessler, 2022).

Research about the development of children's capability to integrate outcome and intention information has focused much more on *punishment recommendations* rather than actual punishment behaviour. This strand of research has made extensive use of vignette tasks, in which children are presented with hypothetical moral violation scenarios through verbal story-telling and then asked whether they consider punishment of norm transgressors an appropriate response. Moral violations scenarios mostly depict property damage and theft (Baird & Astington, 2004), psychological and physical harm (Helwig et al., 2001; Nobes et al., 2016; Zelazo et al., 1996), or a combination of the two (Cushman et al., 2013; Killen et al., 2011; Margoni & Surian, 2017; Martin et al., 2022; Nobes et al., 2009). Importantly, questions about punishment generally take the form of: "*Should [norm transgressor] get in trouble?*" Since children are not asked whether they themselves would punish the norm transgressor presented in the vignette, they do not even have to imagine themselves as hypothetical punishers but just give an opinion about what would be the right course of action.

It has been shown that, when young children are asked to evaluate accidental and failed intentional transgressions

Table 1. Summary of research questions, associated predictions, and whether they were supported in the present study.[a].

Topic	Research question	Prediction	Supported?
Punishment Affective States	Q1: Do children enjoy third-party punishment?	Children do not report positive punishment affective states during and after punishment allocation.	Yes
		No prediction about children's affective states before punishment allocation.	NA
	Q2: Are children's punishment affective states influenced by the impact of punishment on transgressors?	Emphasising the impact of punishment on transgressors decreases children's punishment affective states.	No
Integration Between Outcomes and Intentions in Punishment	Q3: When does the outcome-to-intent shift occur in children's third-party punishment behaviour?	Children manifest the outcome-to-intent shift in third-party punishment between 5 and 8 years of age.	Yes
	Q4: Do children attribute different weights to intentions vs. outcomes depending on moral domains when they carry out third-party punishment?	Children punish failed intentional transgressions more severely than accidental transgressions in case of unfairness, but not in the case of disloyalty.	No

[a]The study was originally additionally intended to examine the effect of the presence or absence of an audience on children's 3PP severity. Because the manipulation check indicated that the audience manipulation was unsuccessful, we decided to omit discussion of this research question in the main manuscript (see Supplementary Information – sections S1 and S4 for a full description of this variable as used in this study).

in hypothetical scenarios, the presence of just one negative cue—either relating to outcomes or intentions—is sufficient for them to recommend punishment. They do not usually appear to attribute more weight to intentions over outcomes, differently from adults. In fact, they attribute equal weight to outcomes and intentions (Baird & Astington, 2004; Cushman et al., 2013; Killen et al., 2011; Margoni & Surian, 2017; Nobes et al., 2016), or more to outcomes over intentions (Helwig et al., 2001; Martin et al., 2022; Zelazo et al., 1996). It is later on during development—with the so-called "outcome-to-intent shift"—that children's punishment recommendations tend to become more intention-based. More specifically, condemnation of accidental transgressions begins to decrease (Cushman et al., 2013), while condemnation of failed intentional transgressions either remains steady (Cushman et al., 2013) or increases with age (Martin et al., 2022). Overall, the age of the outcome-to-intent shift for punishment recommendations varies considerably across studies. A couple of studies found that children as young as 3 years are already able to produce punishment recommendations based on intentions (Nobes et al., 2009; Van de Vondervoort & Hamlin, 2018). However, the majority of the studies showed that the outcome-to-intent shift tends to occur in middle childhood, between 5 and 8 years of age (Baird & Astington, 2004; Cushman et al., 2013; Killen et al., 2011; Martin et al., 2022; Nobes et al., 2016). It has been proposed that the outcome-to-intent shift may be promoted by both internal factors, such as the development of theory of mind skills (Killen et al., 2011) and executive functions (Zelazo et al., 1996), and external factors such as social interactions with adults and peers (Tomasello et al., 2005). Indeed, understanding others' mental states such as intentions enables individuals to make predictions about their future behaviours (Young & Tsoi, 2013). This ability, in turn, may prove crucial to avoid engaging in coordination and negotiation efforts with unreliable social partners (Grueneisen & Tomasello, 2020, 2022) (see the review by Margoni & Surian, 2016).

More recently, research efforts in developmental psychology have been also directed towards the investigation of the outcome-to-intent shift in *actual punishment behaviour*. This line of research has made use of behavioural paradigms (especially economic games such as the ultimatum game), in which children are required to react to apparently real (rather than hypothetical) moral violation scenarios, the vast majority of which involve unfair distribution of resources. Most of these studies focus on 2PP rather than 3PP behaviour. Research on 2PP behaviour has produced rather mixed results, ranging from no evidence of sensitivity to intentions in early to middle childhood (Bernhard et al., 2020; Bueno-Guerra et al., 2016; Wittig et al., 2013) to evidence of sensitivity to intentions already fully developed in primary school-aged children (Jaroslawska et al., 2020; Pelligra et al., 2015; Sutter, 2007) or only emerging during adolescence (Gummerum & Chu, 2014; Güroğlu et al., 2009, 2011).

Regarding the research on the outcome-to-intent shift in 3PP behaviour, it has been shown that 4- to 7-year-old children did not differentiate between unequal distributions stemming from chance or negative intentions (Bernhard et al., 2020), and that both children and adolescents until 15 years of age consistently based their 3PP responses on outcome information (Gummerum & Chu, 2014). To date, evidence of the capability to integrate outcomes and intentions in 3PP behaviour has been found only in adults (Gummerum & Chu, 2014; Hechler & Kessler, 2022), suggesting that the outcome-to-intent shift in their 3PP behaviour may take place in late adolescence. However, it has been also shown that, after having witnessed an adult inflicting 3PP on a norm transgressor, 3- and 4-year-old children were more likely to intervene to reduce the amount of punishment when the transgressor's misbehaviour was accidental rather than intentional (Chernyak & Sobel, 2016). This indicates that children may have some degree of sensitivity to intentions in third-party contexts, even when they are not third-party punishers themselves. Finally, when we consider partner choice behaviours (i.e., avoiding a norm transgressor could be seen as a form of indirect punishment), sensitivity to intentions is detectable even in infants: 8-month-olds preferred to reach for a puppet who was involved in an accidental transgression rather than a failed intentional transgression (Hamlin, 2013, Study 2).

To sum up, the outcome-to-intent shift has been shown to occur, on average, in middle childhood for punishment recommendations (Baird & Astington, 2004; Cushman et al., 2013; Killen et al., 2011; Martin et al., 2022; Nobes et al., 2016), and supposedly in late adolescence for 3PP behaviour (Gummerum & Chu, 2014). The developmental lag between expressing intention-based punishment recommendations and enacting intention-based 3PP behaviour could be due to the different cognitive demands of different experimental paradigms (vignette tasks vs. behavioural paradigms; Hilton & Kuhlmeier, 2019). Another, not mutually exclusive explanation is that this developmental lag is an example of the knowledge-behaviour gap (Blake, 2018): children may have beliefs about the right thing to do in response to a moral transgression that they struggle to implement in practice because of the lack of cognitive control skills.

Since reducing cognitive demands of tasks has been shown to lower the age at which the outcome-to-intent occurs in moral judgements (Margoni & Surian, 2020), we took a similar approach to investigate the outcome-to-intent shift in 3PP behaviour. More specifically, we developed a behavioural paradigm with arguably lower cognitive demands than the one used by Gummerum and Chu (2014) to assess whether children can integrate outcome and intention information into their 3PP behaviour. Whereas in Gummerum and Chu's (2014) paradigm children had to predict how they would react to a range of possible moral transgressions before observing them, in our paradigm children were asked to make 3PP decisions after being shown the transgressions. Moreover, in our paradigm, moral scenarios were presented in such a way that children could infer intentions by observing actors' behaviour and listening to their dialogues as opposed to having to represent their mental states. By reducing processing demands and increasing intention salience, we predicted that children would manifest the outcome-to-intent shift in their 3PP behaviour earlier than in late adolescence (Gummerum & Chu, 2014), and potentially within the same age range commonly observed in punishment recommendations, that is, between 5 and 8 years of age (Baird & Astington, 2004; Cushman et al., 2013; Killen et al., 2011; Martin et al., 2022; Nobes et al., 2016) (Table 1, Q3).

Regarding the moral scenarios in our paradigm, we chose unfairness for comparability with previous literature on the outcome-to-intent shift in 3PP behaviour (Bernhard et al., 2020; Gummerum & Chu, 2014; Hechler & Kessler, 2022) and disloyalty to assess generalisability of the findings. This comparison is relevant in light of moral foundations theory, according to which people's moral concerns pertain to two main domains: an individual-focused domain (including fairness and harm) aimed at the protection of individuals' rights, and a group-focused domain (including loyalty, authority, and purity) aimed at the formation and maintenance of cohesive social groups (Graham et al., 2011). Interestingly, it has been found that among adults, the role of intentions varies across different types of moral domains: intentions matter more when evaluating harm (individual-focused domain) and less when evaluating purity violations (group-focused domain) in both U.S. American and British adults (Chakroff et al., 2016; Sweetman & Newman, 2020a, 2020b; Young & Saxe, 2011; Young & Tsoi, 2013). This finding has been also replicated in a large, multi-site study that included even a broad range of small-scale societies, practising foraging, pastoralism, or horticulture (Barrett et al., 2016). By contrast, nothing is currently known about whether the type of moral domain can influence the weight of intentions in children's 3PP behaviour (although speculations have been made by Bernhard et al., 2020). We reasoned that, if the pattern of attributing more importance to the role of intentions in individual- over group-focused domains is generalisable, children would assign more weight to intentions vs. outcomes for 3PP of unfairness than disloyalty. In other words, children would punish failed intentional transgressions more severely than accidental transgressions in case of unfairness, but not in case of disloyalty (Table 1, Q4).

Method

Sample

The choice of countries for this experiment—Colombia and Spain—was opportunistic but motivated by the desire to counteract sampling bias in developmental psychology (Amir & McAuliffe, 2020; Nielsen et al., 2017) given that the vast majority of studies about punishment mentioned in the Introduction was conducted in Anglo-America or

Northwestern Europe. Latin American and Mediterranean societies endorse more collectivist (vs. individualistic) values than Anglo-American and Northwestern European societies (Hofstede, 2001), meaning that they place a relatively stronger emphasis on the group (vs. the individual). However, different from commonly held assumptions, people from Latin American and Mediterranean societies present a distinctive mixture of independent and interdependent traits in how they relate to others or define themselves. This differentiates them from other collectivist cultures, such as Confucian Asia, where people tend to have more markedly interdependent traits (Krys et al., 2022; Uskul et al., 2023).

We allowed logistical constraints to determine effect sizes; the stopping rule was to collect as much data as possible in the period of time at our disposal. As a result, participants were 123 primary school-aged children, who were tested face-to-face at their schools by the researchers. Of these 123 children, 44 lived in Colombia (*mean age*: 7.7 years; *SD age*: 1.6 years; age range: 5.0–10.8 years; gender distribution: 12 girls and 32 boys), and the remaining 79 in Spain (*mean age*: 8.7 years; *SD age*: 1.7 years; age range: 5.3–11.8 years; gender distribution: 42 girls and 37 boys). Colombian children were all recruited from the same public school in inner Bogotá and were tested from July 2018 to March 2019. Spanish children were instead recruited from multiple schools—one mixed public-private school in Oviedo (Asturias), as well as one public school and two mixed public-private schools in the Madrid region—and tested from November 2019 to January 2020. Regarding the Colombian sample, all caregivers partially or fully completed a socio-demographic questionnaire, indicating that they were all of Colombian nationality, with low-to-middle income and education level (the majority of respondents had a secondary school qualification). As for the Spanish sample, socio-demographic data were not systematically collected but inferred through experimenters' knowledge of the catchment areas: caregivers were predominantly of Spanish nationality, with middle-to-high income and education level. The study was approved by Oxford Brookes University Ethical Review Committee (study number 171101, Children's Social Judgement in a Computer Game) and received Chair's approval by the Universidad de los Andes and Universidad Pontificia Comillas, as well as by the Research Ethics Committee of the Principality of Asturias.

Materials

We developed a spaceship computer game as a variation of the *MegaAttack* game that had previously been employed to test British children (Arini et al., 2021, Study 2). The game was programmed in LÖVE, an open-source game-development environment using the Lua programming language. We then installed the game on various laptop computers that we took to the test locations to conduct testing sessions in-person. Participants saw on the laptop game bouts that they were told were being played and commented on live by internet players (but were in fact pre-recorded). The children's role was to referee internet players in the *MegaAttack* game, judging whether they behaved badly or not. In the former case, children could decide whether to assign punishment to misbehaving internet players and, if so, how much.

The game involved a team of two player-controlled spaceships, shooting enemies and collecting gems. Two in-game tasks subject to potential norm violations were distributing bombs (one player made the allocation between themself and their team-member, potentially unfairly) and participating in a cooperative task for the collection of a mega-gem. For the cooperation task to be successful, both players needed to attach to the mega-gem. If only one of the players attached, with the other player disloyally ignoring them, the attached player would remain trapped.

Each video presenting the moral scenarios via game bouts featured a different pair of player avatars (different animals inside spaceships) to aid memorisation of their different behaviours and was kept short (~1 minute each) with the aim of not excessively taxing children's working memory. Questions being asked of the children did not require articulated verbal responses. All these precautions were made to minimise the cognitive demands of our task (Hilton & Kuhlmeier, 2019; Margoni & Surian, 2020).

Regarding the content of the moral scenarios, the *control trial* portrayed cases of moral norm conformity (i.e., no moral transgressions) in both the fairness and loyalty domains. Therefore, in the control trial, both outcomes and intentions of the players had the same valence (*positive intention, positive outcome*). Instead, the *test trials* portrayed cases of moral transgressions in either the fairness or loyalty domains. Moreover, in the test trials, outcomes and intentions of the players had opposite valences, namely accidental transgressions (*positive intention*, *negative outcome*) and failed intentional transgressions (*negative intention*, *positive outcome*). These two cases are the most informative to study how the relative weight of intentions and outcome changes with age (Ingram & Moreno-Romero, 2021). In addition to that, the fact that each video in the test trials contained only one negative cue, relating to either outcomes or intentions, ruled out the potential inconvenience that children could merely anchor their 3PP decisions to the first negative cue appearing in the scenarios (Nelson, 1980).

Design

We adopted a mixed design in which the factors were *Moral domain* (two within-subject levels: fairness domain and loyalty domain); *Intentionality* (three within-subject

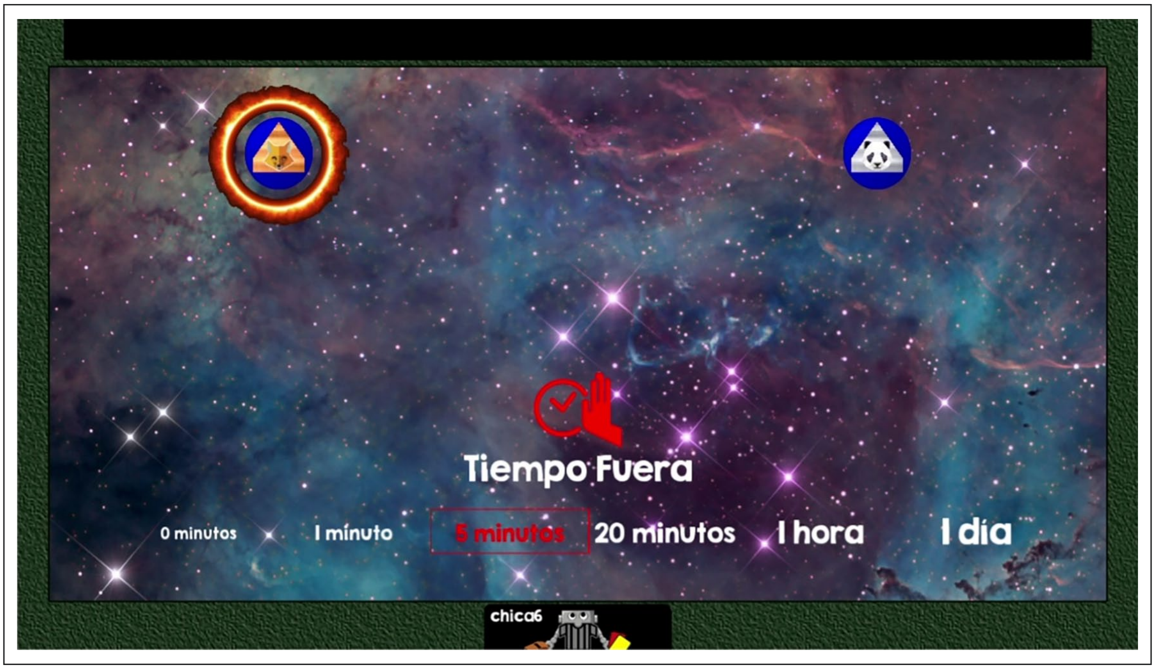

Figure 1. Scale used to measure punishment severity. Punishment was a time-out from the game; children were asked to decide the length of the time-out ("*How long do you want the time out to be?*") among the following options: 0 minutes; 1 minute; 5 minutes; 20 minutes; 1 hour; 1 day.

levels: failed intentional transgression, accidental transgression, and no moral transgression); *Question time* (three within-subject levels: before, during, and after); *Question focus* (two between-subject levels: focus on 3PP impact and no focus on 3PP impact).

All children participating in the experiment were presented first with the control trial (portraying no moral transgressions), followed by the four test trials (portraying moral transgressions) in a counterbalanced order. Order with respect to failed intentional/accidental transgression was ABBA or BAAB, and with respect to disloyalty/unfairness, transgression was ABAB or BABA (see Supplementary Information—Table S1). Notably, by consistently showing the control trial at the beginning of the refereeing sessions, we ensured that participants were always exposed to the same reference point against which to compare subsequent trials (e.g., see the studies by Twardawski & Hilbig, 2020 and Arini et al., 2021 for similar design choices). We reasoned that this approach would aid children's understanding of what was expected of them as referees (i.e., allocating punishment only in case of moral transgressions).

The dependent variables measured were *Punishment severity* (six ordinal levels ranging from 1, "no punishment," to 6, "1-day ban," Figure 1) and *Punishment affective states* (11 ordinal levels from –5, "very bad," to +5, "very good," Figure 2). We also measured *Judgement of transgression severity* for use as a control covariate (six ordinal levels, ranging from −5, "very bad," to 0, "not bad not good," Figure 2), given that substantial variance in punishment severity and affective states has been shown to be explained by this variable (Arini et al., 2021, 2023).

Procedure

The experimental phases were playing familiarisation, refereeing introduction, refereeing sessions, and manipulation check questions. In *playing familiarisation*, the experimenter and the child played together as a team, and the experimenter illustrated to the child the moral norms applied to the *MegaAttack* game. Teammates were expected to equally divide some bombs among themselves (i.e., fairness norm) and to offer each other help in a cooperative task for the collection of a mega-gem (i.e., loyalty norm). In the *refereeing introduction*, the child was told that they would switch from the role of player to that of referee in the *MegaAttack* game. In this new role, the child would have to judge the behaviour of some internet players, with the possibility to give them a time-out from the game when a moral transgression had occurred. We chose this form of punishment for its ecological validity: real computer games have implemented similar systems that give players the possibility to punish misbehaving players by temporarily or permanently banning them from the game (Kou et al., 2017).

In the actual *refereeing sessions* of the experiment (consisting of one control trial and four test trials), the refereeing child watched five purportedly live game bouts that had been actually pre-recorded. During these game bouts, the child could hear dialogues between the internet players

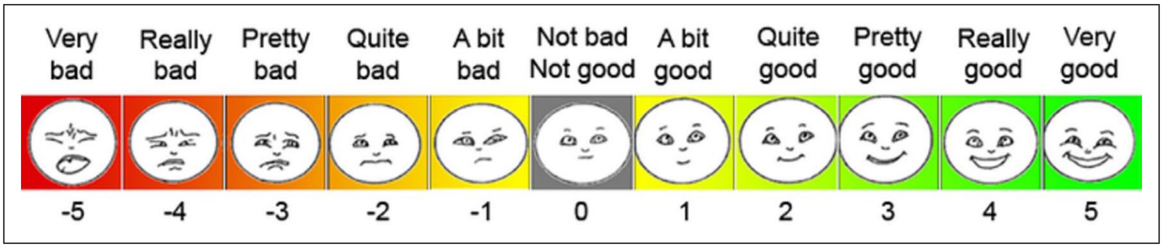

Figure 2. Scale used to measure both judgements of transgression severity and punishment affective states. When children thought that a player had misbehaved in the game, they were asked to judge the severity of the player's transgression ("*How bad do you think that was?*"). Their options ranged from "very bad" to "not bad, not good" (first 6 points of the scale). In addition, children were asked to rate their punishment feelings at three time points: before ("*How do you think it will feel to do that?*"), during ("*How did it feel to do that?*"), and after punishment allocation ("*How did it make you feel?*"). The options they were given to rate their feelings ranged from "very bad" to "very good" (all 11 points of the scale).

describing their own intentions; gender of these voice-overs was matched with that of the child being tested (video rendition of the refereeing sessions with dialogues in Spanish and English translation available at the Open Science Framework: https://osf.io/c9w2a/). In the *control trial*, the child watched one game bout in which no moral norms were violated by the two internet players. Since the players were both loyal and fair to each other, the refereeing child was expected to conclude that no misbehaviours had occurred.

In the *test trials*, the child watched four game bouts, representing a combination of norm transgressions varying in terms of moral domain and intentionality (Supplementary Information—section 1.6). Regarding the norm transgressions being shown in the videos, they could be either accidental transgressions or failed intentional transgressions, related either to the fairness or loyalty domain. Accidental transgressions were characterised by players having positive intentions, followed by negative outcomes. Conversely, failed intentional transgressions were characterised by negative intentions followed by positive outcomes. More specifically, in *accidental unfairness*, one player intended to split the bombs equally with the team member (five bombs each, out of 10) but, by mistake, ended up with more bombs (7/10) than the equal share (Figure 3a). In *failed intentional unfairness*, one player intended to take for themselves more bombs (7/10) than the equal share but inadvertently ended up allocating equal numbers of bombs (5/10) to themselves and the team member (Figure 3b). In *accidental disloyalty*, one player intended to cooperate with the teammate in the mega-gem collection but, due to a mistake, failed to free the trapped teammate from the mega-gem (Figure 3c). In *failed intentional disloyalty*, one player intended to leave the team-mate trapped in the mega-gem but inadvertently set them free (Figure 3d).

After having seen each of the five game bouts, the child had to answer for each of the two players in turn: "*Did this player behave badly?*" If a misbehaviour was identified, the child had to express their judgement of transgression severity of the norm transgression ("*How bad do you think that was?*"; answers provided by using the scale in Figure 2). The child was then asked to establish the punishment severity for the norm transgressor, operationalised as a time-out from the game ("*How long do you want the time out to be?*"; answers provided by using the scale in Figure 1). In addition, children were asked to rate their punishment affective states at three time points (answers provided by using the scale in Figure 2): (a) before their first 3PP decision by forecasting how punishment would feel (time point: before; question: "*How do you think it will feel to do that?*"); (b) once they had punished for the first time (time point: during; question: "*How did it feel to do that?*"); (c) after the last 3PP decision (time point: after; question: "*How did it make you feel?*"). The sentence preceding the question about punishment affective states was manipulated in order to highlight, or not, the cost suffered by the transgressors due to children's administration of punishment (between-subjects framing manipulation: focus on 3PP impact vs. no focus on 3PP impact). For example, with respect to the first time point, in the framing condition that did not emphasise the 3PP impact on transgressors, children were simply asked: "*So, you might punish some players. How do you think it will feel to do that?*" Instead, in the framing condition that emphasised the 3PP impact, children were asked: "*So, you might ban some players from the game so they can't play for quite a while. How do you think it will feel to do that?*" (for details of the framing manipulation for each time point, see Supplementary Information—sections 1.4, 1.7, 1.8).

At the end of the experiment, each child was asked a *manipulation check question* to evaluate the believability of the experimental setting. Specifically, children were questioned about whether they thought they had actually refereed real internet players during the trials ("*Do you think you really watched games with internet players now?*").

Analysis strategy and statistics

Linear mixed-effects models were used to examine the predictors of punishment severity and punishment

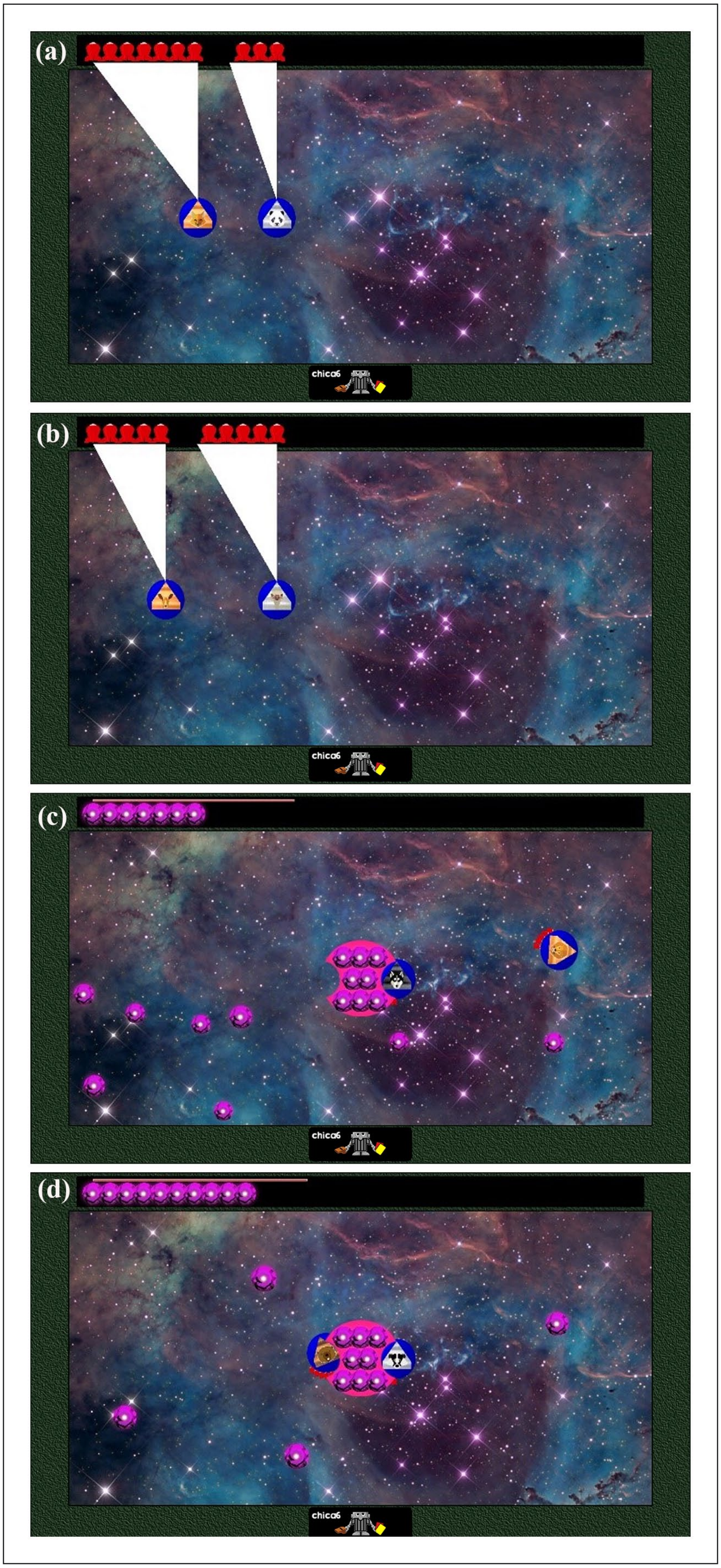

Figure 3. *(Continued)*

Figure 3. Moral violations shown during the test trials. (a) *Accidental Unfairness*: player Fox takes the majority of the bombs for themselves, leaving only a few for their teammate, player Panda. From the players' voice-overs, the participant can understand that player Fox wanted to divide the bombs equally but failed to do so because they inadvertently used the controls wrongly during the bomb distribution. (b) *Failed Intentional Unfairness*: player Kangaroo divides the bombs equally between themselves and their teammate, player Ostrich. From the players' voice-overs, the participant can understand that player Kangaroo intended to take the majority of the bombs for themselves. However, player Kangaroo did not succeed in their plan because they used the controls wrongly during the bomb distribution. (c) *Accidental Disloyalty*: player Lion fails to touch the mega-gem, thus leaving their teammate, player Dog, stuck in it. From the players' voice-overs, the participant can understand that player Lion tried to touch the mega-gem to free player Dog from it but did not succeed in doing so because of their poor spaceship piloting skills. (d) *Failed Intentional Disloyalty*: player Beaver touches the mega-gem, thus freeing their team-mate, player Badger, who was trapped inside. From the players' voice-overs, the participant can understand that player Beaver tried to leave player Badger trapped but, because of their poor spaceship piloting skills, ended up involuntarily touching the mega-gem.

affective states, with participants' ID included as a random factor because there were multiple data points per individual. All other IVs were included as fixed factors. Notably, we included country and gender in our models to explain variance in the DVs, not to test predictions. Our aim was not to detect differences between Colombian and Spanish children but to assess whether we would replicate findings previously obtained in Anglo-American or British samples (e.g., Bernhard et al., 2020; Cushman et al., 2013; Gummerum & Chu, 2014). Moreover, since there was no evidence of gender differences in 3PP in studies using similar paradigms with British children (Arini et al., 2021), we did not expect to find gender differences in Colombian and Spanish children.

Model fits were confirmed by examining diagnostic scatter plots of residuals. All analyses were conducted in the R programming environment (R version 4.0.2, R Core Team, 2020), with raw data and code openly available in the OSF repository (https://osf.io/c9w2a/). In our models, we included main effects and, where appropriate to answer our research questions, two-way interaction effects. However, we did not include any three-way interaction effects, due to concerns of insufficient power to detect effects because of the small sample size.

Results

Preliminary analyses

We first evaluated the believability of the experimental setting. Believability was high, with 90% of the children, 95% CI [83, 95], thinking they had actually refereed real internet players during the trials. Since there was almost no variability in this measure, we excluded believability from our statistical models.

Preliminary analyses of the control trial revealed that only 3 out of the 123 participants identified moral norm violations, when in fact they were shown cases of moral norm conformity (i.e., both outcomes and intentions were positive for both fairness and loyalty domains). Analyses presented below, therefore, exclude the control trial, which served its purpose by demonstrating that participants could generally distinguish moral norm violations from moral norm conformity. Therefore, in our statistical models about punishment severity, we considered only two within-subject levels for intentionality: failed intentional transgression and accidental transgression.

Punishment affective states

As shown in Table 2, linear mixed-effects analyses revealed that there were no significant temporal changes on children's punishment affective states. Overall, children did not much enjoy making 3PP decisions: across time points (before, during, and after punishment allocation), children's reported affective states were on average $M = -0.33$, $SD = 2.60$, which was not significantly different from 0, $t(122) = -1.42$, $p = .157$ (Q1 in Table 1; see Figure 4). In addition, whether children were prompted to focus on the impact of 3PP or not had no effect on their punishment affective states either (Q2 in Table 1). On the other hand, there was a significant effect of country, with Colombian children reporting more negative punishment affective states ($M = -0.92$, $SD = 3.55$) than Spanish children ($M = -0.03$, $SD = 2.80$).

Punishment severity

As shown in Table 3, linear mixed-effects analyses revealed a significant effect of judgement of transgression severity, indicating that the harsher the judgement, the more severe the punishment. In addition, there was a significant effect of age, meaning that the older the children, the more lenient their 3PP behaviour towards the transgressors. There was a significant effect of moral domain, with children punishing disloyalty transgressions ($M = 3.92$, $SD = 1.45$) more harshly than unfairness transgressions ($M = 3.32$, $SD = 1.54$). We also found a significant effect of intentionality, meaning that children on average punished failed intentional transgressions ($M = 3.77$, $SD = 1.45$) more harshly than accidental transgressions ($M = 3.50$, $SD = 1.60$).

Furthermore, we found a significant interaction between intentionality and age (Q3 in Table 1). Notably, this was not accompanied by an interaction between moral domain and intentionality (Q4 in Table 1). In other words, we

Table 2. Modulating factors of punishment affective states.

Factor	*b*	β	95% CI for β	χ^2	*p*
Judgement of transgression severity (average)	−.22	−.09	[−.25, .07]	1.25	.263
Age	−.11	−.06	[−.22, .10]	0.53	.467
Gender	.21	.07	[−.24, .37]	0.18	.669
Country	1.17	.38	[.05, .70]	5.07	.024*
Question focus	.19	.06	[−.23, .35]	0.16	.688
Question time				3.93	.140
During vs. Before	.38	.12	[−.05, .29]		
After vs. Before	.51	.16	[.00, .33]		

For binary variables, the following categories are coded as 1 (and the others as 0): gender male, country Spain, question focused on impact, believed the game to be real. An additional categorical variable is question time, which is ternary rather than binary (categories are before, during and after, with before used as a reference). The continuous predicting factors are age and judgement of transgression severity averaged across trials. Raw model coefficients *b* are standardised to produce β and associated 95% confidence interval by normalising by standard deviation of the dependent variable in all cases and by the standard deviation of the predicting factor only when it is not categorical (age, judgement of transgression severity averaged across trials), meaning categorical β (gender, country, question focus, and question time) is analogous to Cohen's *d*. The scale used to measure judgement of transgression severity ranged from −5 to 0, meaning that the more negative the values, the harsher/more severe the judgements.
$^*p \leq .050$; $^{**}p \leq .010$; $^{***}p \leq .001$.

found evidence of an outcome-to-intent shift in 3PP behaviour, which occurred in both moral domains in parallel. From a visual interpretation of the data (see Figure 5), it appears that the outcome-to-intent shift occurred around 7 years of age in unfairness and disloyalty. Namely, children of 7 years of age or younger tended to punish failed intentional transgressions (disloyalty: M=4.31, SD=1.64; unfairness: M=4.39, SD=1.50) as severely as accidental transgressions (disloyalty: M=4.61, SD=1.46; unfairness: M=4.78, SD=1.00). In contrast, children older than 7 years tended to punish failed intentional transgressions (disloyalty: M=3.92, SD=1.31; unfairness: M=3.15, SD=1.35) more severely than accidental transgressions (disloyalty: M=3.50, SD=1.48; unfairness: M=2.65, SD=1.43). We additionally discovered a significant interaction between age and moral domain: punishment severity decreased with children's increasing age in cases of unfairness, whereas it remained more stable across ages in cases of disloyalty, see Figure 5.

Discussion

By testing 5- to 11-year-old children from Colombia and Spain, the present study has expanded knowledge about cognitive and emotional processes involved in the 3PP behaviour—topics that had been investigated so far mainly in samples from Anglo-America or Northwestern Europe (Marshall & McAuliffe, 2022; see also discussions about sampling bias in developmental psychology in the papers by Amir & McAuliffe, 2020, and Nielsen et al., 2017). We specifically focused on the emotional consequences of implementing 3PP decisions, and on the integration between outcome and intention information in 3PP decision-making, across different moral domains—disloyalty to the group (a group-focused moral domain) and unfairness in resource distribution (an individual-focused moral domain; Graham et al., 2011).

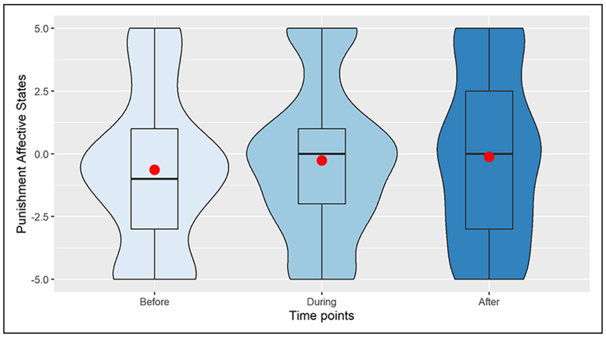

Figure 4. Punishment affective states across time points (before, during, and after punishment allocation). Violin plots wrapping boxplots; boxplots showing median and interquartile range, and a large dot for mean value.

Regarding punishment affective states, we replicated the result that Arini et al. (2021, Study 2) obtained in British children by demonstrating that Colombian and Spanish children tended not to derive enjoyment from punishing transgressors (although neither did they tend to find it deeply unpleasant). Interestingly, in the present study, children's affective states before 3PP allocation were not more positive than those reported during and after 3PP allocation. Consequently, we can rule out the hypothesis that children had hedonic expectations about 3PP that did not stand the test of reality. Rather, it is more likely that carrying out 3PP does not usually evoke much positive emotions in children, in line with Marshall et al. (2021, Supplementary Information). It is noteworthy that children in our study did not make the same forecasting error typically committed by adults. Indeed, adults who were asked to predict how they would feel if they could punish transgressors reported more positive feelings than their counterparts who actually enacted punishment (Carlsmith et al., 2008). We acknowledge the possibility that, after the children had responded to the first affective question, they

Table 3. Modulating factors of punishment severity.

Factor	b	β	95% CI for β	χ^2	p
Judgement of transgression severity	−.26	−.26	[−.36, −.15]	22.31	<.001***
Age	−.20	−.22	[−.40, −.05]	25.38	<.001***
Gender	−.01	−.01	[−.26, .24]	0.00	.947
Country	−.12	−.08	[−.35, .19]	0.34	.560
Moral domain	1.31	−.31	[−.56, −.07]	18.27	<.001***
Intentionality	−1.41	.10	[−.12, .32]	9.46	.024*
Moral domain × Intentionality	.12	.08	[−.24, .40]	0.23	.630
Age × Moral domain	−.21	−.24	[−.40, −.09]	9.07	.003**
Age × Intentionality	.19	.21	[.05, .37]	6.84	.009**

For binary variables, the following categories are coded as 1 (and the others as 0): gender male, country Spain, believed the game to be real, domain of unfairness, and failed intentional transgression. Raw model coefficients *b* are standardised to produce β and associated 95% confidence interval by normalising by standard deviation of the dependent variable in all cases and by the standard deviation of the predicting factor only when it is not categorical (age, judgement of transgression severity), meaning categorical β (gender, country, moral domain, and intentionality) is analogous to Cohen's *d*. The scale used to measure judgement of transgression severity ranged from −5 to 0 (the more negative the values, the harsher the judgements); therefore, negative *b* and β coefficients indicate a direct relationship between judgement of transgression severity and punishment severity (the harsher the judgement, the harsher the punishment).
*$p \leq .050$; **$p \leq .010$; ***$p \leq .001$.

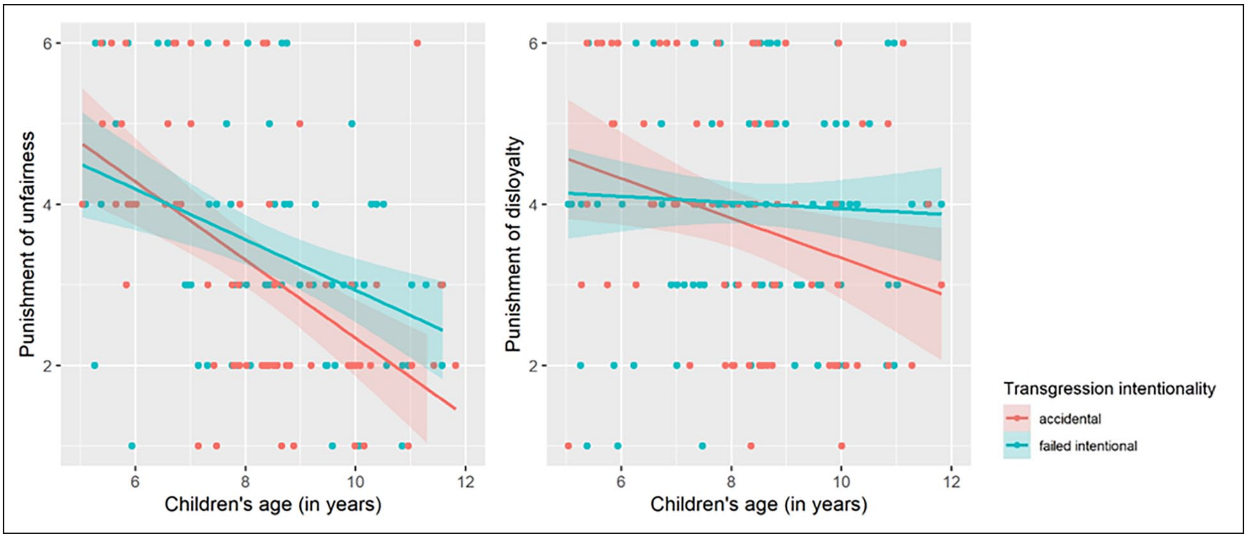

Figure 5. Punishment severity by moral domain (disloyalty, unfairness) and intentionality (accidental transgression, failed intentional transgression). Punishment severity is measured on a scale from 1 (no punishment) to 6 (1-day ban).

simply responded in similar ways on the two subsequent questions for the sake of consistency, rather than because their affective states were really the same throughout the experiment. However, changes in punishment affective states over time have been recorded using similar experimental methods, even though children were being repeatedly asked the same question at different time points (Arini et al., 2023). Therefore, it is unlikely that our finding (i.e., consistent lack of much enjoyment over time) is a mere artefact due to the protocol we adopted. Nevertheless, future studies should complement self-reported measures of punishment affective states with implicit measures of emotional arousal (e.g., skin conductance) to validate the findings (as in the work of Gummerum et al., 2020). Future research could also test whether children would be more likely to enjoy 3PP if they were presented with evidence that transgressors suffered and/or changed moral attitude after punishment, as it has been demonstrated in adults (Aharoni et al., 2022; Eder et al., 2020; Funk et al., 2014; Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011).

Furthermore, since past research found that children were more likely to report lack of enjoyment when their 3PP decisions had real rather than pretend consequences (Arini et al., 2021, Study 2), we decided to investigate whether children's affective states are sensitive to the impact of 3PP on the transgressors. We predicted that inducing children to focus on the impact the 3PP has on the transgressors while questioning them about their punishment affective states would make them feel worse, due to feeling responsible for the transgressors' suffering. In fact, it was found that question focus (focus on 3PP impact vs. no focus on 3PP impact) was not a significant predictor of children's punishment affective states in our experiment. However, since we did not include any manipulation check to verify whether the wording of the question about punishment affective states was effective in activating punishment impact representations, this null result is difficult to interpret. At this stage, we cannot know if our manipulation did not work or if children did not feel responsible for the transgressor's suffering.

We speculate that these affective findings may shed light on children's 3PP motives. Potential motives guiding punishment are retribution (i.e., the desire to make the transgressors suffer in proportion to the damage they caused as a means to righting past wrongs) and deterrence (i.e., the desire to make the transgressors learn a moral lesson to prevent them from misbehaving again in the future; Aharoni et al., 2022). In adults, it has been found that manipulating retribution-relevant information increased participants' punitive tendencies, yet manipulating deterrence-relevant information did not (Carlsmith et al., 2002; Molho et al., 2022), which suggests that adults are primarily motivated by retribution. Notably, a piece of information connected to retribution is punishment severity (Keller et al., 2010, Study 3), which is akin to how we framed 3PP impact in our experiment. Therefore, the fact that we did not find an effect of 3PP impact on children's affective states is suggestive that children's 3PP behaviour may not be motivated by retribution. This would be in accordance with research by Arini et al. (2023), which showed that children not only endorsed deterrence over retribution (explicit measure of punishment motivation) but also recalled deterrence messages at higher rates than retribution messages (as an implicit measure of punishment motivation). Furthermore, there is evidence that children punished transgressors at higher rates and invested more resources into 3PP when doing so satisfied deterrent goals, in addition to retributive ones (Marshall et al., 2021; Twardawski & Hilbig, 2020).

In our study, we additionally investigated the cognitive integration of outcome and intention information in actual punishment behaviour in response to ostensibly real moral violations (rather than in verbal punishment recommendations for hypothetical moral violations, as in most previous studies). We found that children began to punish failed intentional transgressions more severely than accidental transgressions from around the age of 7 years, when the outcome-to-intent shift became noticeable, with similar patterns across moral domains (unfairness and disloyalty). In contrast, previous behavioural studies investigating the outcome-to-intent shift in 3PP either found no evidence of sensitivity to intentions in 4- to 7-year-old children (Bernhard et al., 2020) or found evidence of this sensitivity only in adulthood (Gummerum & Chu, 2014; Hechler & Kessler, 2022). Therefore, ours appears to be the first behavioural study to provide evidence of the capability to integrate outcomes and intentions into 3PP decisions already in childhood. To note, the different degree of processing demands and stimuli salience between our study and that of Gummerum and Chu (2014) is likely responsible for the striking age difference in the onset of the outcome-to-intent shift (7 years of age vs. late adolescence). Indeed, the methodology employed by Gummerum and Chu (2014) may have taxed children's cognitive resources, impeding their capability to integrate outcomes and intentions into their 3PP decisions (Hilton & Kuhlmeier, 2019; Margoni & Surian, 2020). If confirmed, this would represent further evidence in support of the hypothesis that the outcome-to-intent shift is affected by the development of cognitive skills (Killen et al., 2011; Zelazo et al., 1996; see the review by Margoni & Surian, 2016).

In addition, children in our experiment manifested the outcome-to-intent shift in their 3PP behaviour within the same age range previously observed in the vignette studies on punishment recommendations (between 5 and 8 years; Baird & Astington, 2004; Cushman et al., 2013; Killen et al., 2011; Martin et al., 2022; Nobes et al., 2016). This is noteworthy given that individuals often do not carry out the behaviour they judge appropriate (Blake, 2018; see also the discussion in the study by Kenward & Östh, 2015). A final relevant consideration is that, if 3PP evolved as a mechanism to enforce group cooperation (Boyd & Richerson, 1992), it makes sense for children to start taking intentions into account in their 3PP behaviour during middle childhood. It is indeed during this developmental period that children increasingly engage in social interactions with their peers and face their first coordination and bargaining problems (Grueneisen & Tomasello, 2020, 2022). Thus, becoming watchful about clues indicating someone's intention to disregard cooperative norms is likely to be adaptive, as it would allow the avoidance of apparently unreliable social partners. Consequently, the cost of developing this ability only in late adolescence would probably be too high (Margoni & Surian, 2016).

Regarding the effect of moral domain on intention sensitivity, children in our experiment assigned equal weight to intentions in their 3PP decisions across moral domains (unfairness vs. disloyalty). In contrast, previous studies have shown that adults assign different weights to intentions depending on the moral domain (with intentions having greater influence in harm than purity transgressions) (Barrett et al., 2016; Chakroff et al., 2016; Sweetman & Newman, 2020a, 2020b; Young & Saxe, 2011; Young & Tsoi, 2013). Therefore, our findings do not support the hypothesis that people tend to attribute more importance to intentions in individual-focused domains (i.e., harm and fairness) than in group-focused domains (i.e., loyalty, authority, and purity; Graham et al., 2011). These contrasting results could be due to differences in the details of the moral scenarios or to developmental differences.

Differences in intention sensitivity may only be detectable when comparing harm and purity transgressions and may not extend to other moral domains, or all exemplars within them. Alternatively, children might not have fully developed the ability to differentiate the weight of intentions across different moral domains. If confirmed, this would suggest that moral decision-making becomes more domain-specific with age. Future studies should therefore discern between alternative explanations by investigating the developmental trajectory of intention sensitivity from childhood to adulthood across a broader range of moral domains and scenarios within them.

However, it is worth noting our exploratory analyses about how punishment of different moral transgressions changes across development. In this experiment, Colombian and Spanish children punished disloyalty more harshly than unfairness. Moreover, their 3PP severity of disloyalty tended to remain stable across ages, while 3PP severity of unfairness decreased as children got older. It could be argued that 3PP severity of disloyalty remained stable throughout development because the disloyalty scenario was less cognitively demanding than the unfairness scenario. However, this would not explain why British children in the same age range reacted to the view of the same moral scenarios in quite different ways (Arini et al., 2021). When British children were tested on a paradigm that closely resembled the one Colombian and Spanish children were confronted with, their 3PP severity was comparable across moral domains and decreased with an age-dependent pattern for disloyalty and unfairness alike (Arini et al., 2021, Study 2). When instead British children were tested on a paradigm that allowed them to use 3PP not only to make the transgressor pay for their action but also to equalise the resource imbalance between the victim and the transgressor (an arguably cognitively demanding task), 3PP severity of unfairness remained steadily high across ages, while 3PP severity of disloyalty decreased as children got older (Arini et al., 2021, Study 1). Considering this evidence, even though we cannot rule out that differences in cognitive demands between moral scenarios played a role in children's 3PP decisions, we deem them unlikely to explain our pattern of results.

To sum up, if we consider punishment severity as a proxy of the importance attributed to a specific moral domain, Colombian and Spanish children were more concerned about disloyalty than unfairness, whereas British children were either equally concerned about the two or more concerned about unfairness than disloyalty (Arini et al., 2021). Given that the culture in Spain and Colombia is more collectivist than that in the UK (Hofstede, 2001; see also the works by Krys et al., 2022; Uskul et al., 2023), these findings are in line with research conducted in adults suggesting that collectivism may be associated with higher concerns about group- than individual-focused moral domains (Graham et al., 2011; Triandis, 1989). Crucially, when differences between moral concerns were detected within a sample, either in the present experiment or in the one by Arini et al. (2021), they tended to increase with development. In other words, the longer children were exposed to the specific moral system of their own sociocultural environment, the more their moral concerns became selective towards the moral domain deemed central in said environment, thus mirroring adults' moral concerns—a pattern consistent with cultural learning processes. This is an important testing ground for the moral foundations theory's claim that moral development is driven by cultural learning (Graham et al., 2013). Although the results are suggestive, this interpretation warrants caution. The samples of children in both the present study and in that of Arini et al. (2021) were not necessarily representative of the respective national populations. It follows that their punishment behaviour may reflect local norms in their specific environment (e.g., school, neighbourhood) rather than collectivist or individualistic tendencies in their countries (Colombia, Spain, UK). However, if future research confirmed this preliminary evidence, it would provide insight into the complex relationship between culturally-salient moral norms and the development and variation of children's 3PP behaviour across societies. Such studies would also benefit from taking a gender perspective since there is evidence of gender differences in individualism and collectivism (Dabiriyan Tehrani & Yamini, 2022), which may affect moral concerns towards specific norm violations.

Finally, it is important to acknowledge that a limitation of our study due to logistic constraints is the relatively small size, which might have prevented the detection of effects when they were in fact present because of the lack of statistical power. This shortcoming might have also created issues of reliability for the effects that were indeed detected. Therefore, the current evidence should be regarded as preliminary, and future studies should aim at replicating our results in a larger sample. Moreover, there were differences in size, gender distribution, mean age, and representativeness between the Colombian and Spanish samples. The Colombian sample was smaller, with a higher proportion of male children, and its mean age was a whole year younger compared to the Spanish sample. In addition, all Colombian participants came from the same city and attended the same school, whereas Spanish participants were recruited from four different schools located in two different cities. However, our choice to recruit children from Colombia and Spain was not motivated by the desire to detect cultural differences between these two samples. Rather, we wanted to broaden representation in developmental psychology (Amir & McAuliffe, 2020; Nielsen et al., 2017) and test the generalisability of findings about children's 3PP previously obtained in Anglo-American or Northwestern European samples (reviewed in the paper by Marshall & McAuliffe, 2022). Another limitation of the current study is that it was conducted in one specific experimental setting—a computer-mediated paradigm—whose generalisability to real-world situations has only recently been tested (Arini et al., 2023). However, computer games represent a real social world that children already inhabit, experience, and react to norm violations within (Kou et al., 2017); thus, the ecological validity of the present experimental paradigm is expected to be high. A further weakness of our study is that we employed only one behavioural exemplar for each moral domain; therefore, future studies would benefit from using more than one example of behaviour per type of moral domain. Finally, we acknowledge the lack of order balancing for the control trial (only test trials were counterbalanced). This design choice was motivated by the need to initiate the refereeing sessions with the same baseline condition for all the participants (similar to what has been done in, e.g., the studies by Twardawski & Hilbig, 2020 and Arini et al., 2021), but it would be beneficial if future studies adopted a fully counterbalanced design as a robustness check.

In conclusion, the present study has deepened the understanding of cognitive and emotional processes playing a crucial role in children's moral development. To our knowledge, this has been the first study to provide evidence of the outcome-to-intent shift in 3PP behaviour during middle childhood. More specifically, children began to attribute higher importance to intentions over outcomes in 3PP behaviour, across different moral domains, around 7 years of age, in line with findings about the outcome-to-intent shift in punishment recommendations (Baird & Astington, 2004; Cushman et al., 2013; Killen et al., 2011; Martin et al., 2022; Nobes et al., 2016). We also found that children in our study did not derive much enjoyment from enacting 3PP, in accordance with previous literature (Arini et al., 2021, Study 2; Marshall et al., 2021, Supplementary Information), nor did they anticipate to feel much enjoyment. We also discovered interesting cross-cultural differences: Colombian and Spanish children punished disloyalty more severely than unfairness, in contrast with the behavioural patterns observed in British children, whose 3PP severity of unfairness was either higher or equal to that of disloyalty (Arini et al., 2021). Since different cultures privilege different moral domains (Graham et al., 2011; Triandis, 1989), further studies are needed at the intersection between developmental psychology and cognitive anthropology in order to shed light on moral development from a cross-cultural perspective. This would enable a more fine-grained distinction between universal and culture-specific developmental patterns of punishment behaviour and affective states, ultimately enriching understanding about proximate and evolutionary causes of our socio-moral behaviour.

Acknowledgements

The authors thank Morag MacLean and Michaela Gummerum for their precious feedback on an early draft of the manuscript, all the school staff who supported this project (particularly Gina Piñeros Romero), and the families who took part in it. In addition, the authors thank Daniela Rosas Moreno for assistance in the development of the methodology, Juan Sebastian Nassar Pereira and Juan Jaccobo Garzon Martelo for assistance in data collection, Vanessa Anahi Merlos-Fuentes for subtitling the video rendition of the refereeing sessions, and all the students of the Universidad de los Andes who kindly agreed to have their voices recorded to create the voice-overs for the experimental stimuli.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: 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).

ORCID iD

Rhea L. Arini https://orcid.org/0000-0002-9856-6859
Estrella Fernández Alba https://orcid.org/0000-0003-1226-5615

Data accessibility statement

The data and materials from the present experiment are publicly available at the Open Science Framework website: osf.io/c9w2a/

Supplementary material

The supplementary material is available at qjep.sagepub.com.

References

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*, 138–151. https://doi.org/10.1027/2151-2604/a000460

Amir, D., & McAuliffe, K. (2020). Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behavior*, *41*(5), 430–444. https://doi.org/10.1016/j.evolhumbehav.2020.06.006

Arini, R. L., Mahmood, M., Aljure, J. B., Ingram, G. P. D., Wiggs, L., & Kenward, B. (2023). Children endorse deterrence motivations for third-party punishment but derive higher enjoyment from compensating victims. *Journal of Experimental Child Psychology*, *230*, 105630. https://doi.org/10.1016/j.jecp.2023.105630

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

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, *2004*(103), 37–49. https://doi.org/10.1002/cd.96

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., Henrich, J., Kanovsky, M., Kushnick, G., Pisor, A., Scelza, B. A., Stich, S., von Rueden, C., Zhao, W., & Pisor, A. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4688–4693. https://doi.org/10.1073/pnas.1522070113

Bernhard, R. M., Martin, J. W., & Warneken, F. (2020). Why do children punish? Fair outcomes matter more than intent in children's second- and third-party punishment. *Journal of Experimental Child Psychology*, *200*, 104909. https://doi.org/10.1016/j.jecp.2020.104909

Blake, P. R. (2018). Giving what one should: Explanations for the knowledge-behavior gap for altruistic giving. *Current Opinion in Psychology*, *20*, 1–5. https://doi.org/10.1016/j.copsyc.2017.07.041

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*(3), 171–195. https://doi.org/10.1016/0162-3095(92)90032-Y

Bueno-Guerra, N., Leiva, D., Colell, M., & Call, J. (2016). Do sex and age affect strategic behavior and inequity aversion in children? *Journal of Experimental Child Psychology*, *150*, 285–300. https://doi.org/10.1016/j.jecp.2016.05.011

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. https://doi.org/10.1037/a0012165

Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., & Young, L. (2016). When minds matter for moral judgment: Intent information is neurally encoded for harmful but not impure acts. *Social Cognitive and Affective Neuroscience*, *11*(3), 476–484. https://doi.org/10.1093/scan/nsv131

Chernyak, N., & Sobel, D. M. (2016). "But he didn't mean to do it": Preschoolers correct punishments imposed on accidental transgressors. *Cognitive Development*, *39*, 13–20. https://doi.org/10.1016/j.cogdev.2016.03.002

Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of "morality-as-cooperation" with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

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

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, *127*(1), 6–21. https://doi.org/10.1016/j.cognition.2012.11.008

Dabiriyan Tehrani, H., & Yamini, S. (2022). Gender differences concerning the horizontal and vertical individualism and collectivism: A meta-analysis. *Psychological Studies*, *67*(1), 11–27. https://doi.org/10.1007/s12646-022-00638-x

De Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, *305*(5688), 1254–1258. https://doi.org/10.1126/science.1100735

Eder, A. B., Mitschke, V., & Gollwitzer, M. (2020). What stops revenge taking? Effects of observed emotional reactions on revenge seeking. *Aggressive Behavior*, *46*(4), 305–316. https://doi.org/10.1002/ab.21890

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140. https://doi.org/10.1038/415137a

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

Ginther, M. R., Hartsough, L. E., & Marois, R. (2022). Moral outrage drives the interaction of harm and culpable intent in third-party punishment decisions. *Emotion*, *22*(4), 795. https://doi.org/10.1037/emo0000950

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

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In P. Devine & A. Plant (Eds.), *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Academic Press.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366. https://doi.org/10.1037/a0021847

Grueneisen, S., & Tomasello, M. (2020). The development of coordination via joint expectations for shared benefits. *Developmental Psychology*, *56*(6), 1149–1156. https://doi.org/10.1037/dev0000936

Grueneisen, S., & Tomasello, M. (2022). How fairness and dominance guide young children's bargaining decisions. *Child Development*, *93*(5), 1318–1333. https://doi.org/10.1111/cdev.13757

Gummerum, M., & Chu, M. T. (2014). Outcomes and intentions in children's, adolescents', and adults' second-and third-party punishment behavior. *Cognition*, *133*(1), 97–103. https://doi.org/10.1016/j.cognition.2014.06.001

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*, 105376. https://doi.org/10.1016/j.jecp.2022.105376

Gummerum, M., Takezawa, M., & Keller, M. (2009). The influence of social category and reciprocity on adults' and children's altruistic behavior. *Evolutionary Psychology*, *7*(2), 147470490900700. https://doi.org/10.1177/147470490900700212

Güroğlu, B., van den Bos, W., & Crone, E. A. (2009). Fairness considerations: Increasing understanding of intentionality during adolescence. *Journal of Experimental Child Psychology*, *104*(4), 398–409. https://doi.org/10.1016/j.jecp.2009.07.002

Güroğlu, B., van den Bos, W., van Dijk, E., Rombouts, S. A., & Crone, E. A. (2011). Dissociable brain networks involved in development of fairness considerations: Understanding intentionality behind unfairness. *NeuroImage*, *57*(2), 634–641. https://doi.org/10.1016/j.neuroimage.2011.04.032

Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *128*(3), 451–474. https://doi.org/10.1016/j.cognition.2013.04.004

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

Hartsough, L. E., Ginther, M. R., & Marois, R. (2020). Distinct affective responses to second-and third-party norm violations. *Acta Psychologica*, *205*, 103060. https://doi.org/10.1016/j.actpsy.2020.103060

Hechler, S., & Kessler, T. (2022). The importance of unfair intentions and outcome inequality for punishment by third parties and victims. *Zeitschrift für Psychologie*, *230*, 114–126. https://doi.org/10.1027/2151-2604/a000458

Helwig, C. C., Zelazo, P. D., & Wilson, M. (2001). Children's judgments of psychological harm in normal and noncanonical situations. *Child Development*, *72*(1), 66–81. https://doi.org/10.1111/1467-8624.00266

Hilton, B. C., & Kuhlmeier, V. A. (2019). Intention attribution and the development of moral evaluation. *Frontiers in Psychology*, *9*, 2663. https://doi.org/10.3389/fpsyg.2018.02663

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage Publications.

House, B. R., Kanngiesser, P., Barrett, H. C., Yilmaz, S., Smith, A. M., Sebastian-Enesco, C., Erut, A., & Silk, J. B. (2020). Social norms and cultural diversity in the development of third-party punishment. *Proceedings of the Royal Society B*, *287*(1925), 20192794. https://doi.org/10.1098/rspb.2019.2794

Ingram, G. P. D., & Moreno-Romero, C. (2021). Dual-process theories, cognitive decoupling and the outcome-to-intent shift: A developmental perspective on evolutionary ethics. In J. De Smedt & H. De Cruz (Eds.), *Empirically engaged evolutionary ethics. Vol. 437: Synthese library* (pp. 17–40). Springer. https://doi.org/10.1007/978-3-030-68802-8_2

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology*, *96*(3), 521. https://doi.org/10.1037/a0013779

Jaroslawska, A. J., McCormack, T., Burns, P., & Caruso, E. M. (2020). Outcomes versus intentions in fairness-related decision making: School-aged children's decisions are just like those of adults. *Journal of Experimental Child Psychology*, *189*, 104704. https://doi.org/10.1016/j.jecp.2019.104704

Jensen, K. (2010). Punishment and spite, the dark side of cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2635–2650. https://doi.org/10.1098/rstb.2010.0146

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(35), 12710–12715. https://doi.org/10.1073/pnas.1402280111

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. (2015). Five-year-olds punish antisocial adults. *Aggressive Behavior*, *41*(5). https://doi.org/10.1002/AB.21568

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Testing theory of mind and morality knowledge in young children. *Cognition*, *119*, 197–215. https://doi.org/10.1016/j.cognition.2011.01.006

Kou, Y., Johansson, M., & Verhagen, H. (2017). Prosocial behavior in an online game community: An ethnographic study. In *FDG '17: Proceedings of the 12th International Conference on the Foundations of Digital Games* (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3102071.3102078

Krys, K., Vignoles, V. L., de Almeida, I., & Uchida, Y. (2022). Outside the "cultural binary": Understanding why Latin American collectivist societies foster independent selves. *Perspectives on Psychological Science*, *17*(4), 1166–1187. https://doi.org/10.1177/17456916211029632

Lee, Y. E., & Warneken, F. (2022). Does third-party punishment in children aim at equality? *Developmental Psychology*, *58*(5), 866. https://doi.org/10.1037/dev0001331

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

Margoni, F., & Surian, L. (2016). Explaining the U-shaped development of intent-based moral judgments. *Frontiers in Psychology*, *7*, 171613. https://doi.org/10.3389/fpsyg.2016.00219

Margoni, F., & Surian, L. (2017). Children's intention-based moral judgments of helping agents. *Cognitive Development*, *41*, 46–64. https://doi.org/10.1016/j.cogdev.2016.12.001

Margoni, F., & Surian, L. (2020). Conceptual continuity in the development of intent-based moral judgment. *Journal of Experimental Child Psychology*, *194*, 104812. https://doi.org/10.1016/j.jecp.2020.104812

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Ensminger, J., Gurven, M., Gwako, E., Henrich, J., Henrich, N., Lesorogol, C., McElreath, R., & 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., & McAuliffe, K. (2022). Children as assessors and agents of third-party punishment. *Nature Reviews Psychology*, *1*(6), 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

Martin, J. W., Leddy, K., Young, L., & McAuliffe, K. (2022). An earlier role for intent in children's partner choice versus punishment. *Journal of Experimental Psychology: General*, *151*(3), 597. https://doi.org/10.1037/xge0001093

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, *134*, 1–10. https://doi.org/10.1016/j.cognition.2014.08.013

Mendes, N., Steinbeis, N., Bueno-Guerra, N., Call, J., & Singer, T. (2018). Preschool children and chimpanzees incur costs to watch punishment of antisocial others. *Nature Human Behaviour*, *2*(1), 45–51. https://doi.org/10.1038/s41562-017-0264-5

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

Molho, C., Tybur, J. M., Güler, E., Balliet, D., & Hofmann, W. (2017). Disgust and anger relate to different aggressive responses to moral violations. *Psychological Science*, *28*(5), 609–619. https://doi.org/10.1177/0956797617692000

Nelson, S. A. (1980). Factors influencing young children's use of motives and outcomes as moral criteria. *Child Development*, *51*, 823–829. https://doi.org/10.2307/1129470

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Nobes, G., Panagiotaki, G., & Bartholomew, K. J. (2016). The influence of intention, outcome and question-wording on children's and adults' moral judgments. *Cognition*, *157*, 190–204. https://doi.org/10.1016/j.cognition.2016.08.019

Nobes, G., Panagiotaki, G., & Pawson, C. (2009). The influence of negligence, intention, and outcome on children's moral judgments. *Journal of Experimental Child Psychology*, *104*(4), 382–397. https://doi.org/10.1016/j.jecp.2009.08.001

Pelligra, V., Isoni, A., Fadda, R., & Doneddu, G. (2015). Theory of mind, perceived intentions and reciprocal behaviour: Evidence from individuals with Autism Spectrum Disorder. *Journal of Economic Psychology*, *49*, 95–107. https://doi.org/10.1016/j.joep.2015.05.001

Pfattheicher, S., Sassenrath, C., & Keller, J. (2019). Compassion magnifies third-party punishment. *Journal of Personality and Social Psychology*, *117*(1), 124. https://doi.org/10.1037/pspi0000165

Raihani, N. J., & Bshary, R. (2019). Punishment: One tool, many uses. *Evolutionary Human Sciences*, *1*, e12. https://doi.org/10.1017/ehs.2019.12

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

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

Salali, G. D., Juda, M., & Henrich, J. (2015). Transmission and development of costly punishment in children. *Evolution and Human Behavior*, *36*(2), 86–94. https://doi.org/10.1016/j.evolhumbehav.2014.09.004

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

Sutter, M. (2007). Outcomes versus intentions: On the nature of fair behavior and its development with age. *Journal of Economic Psychology*, *28*(1), 69–78. https://doi.org/10.1016/j.joep.2006.09.001

Sweetman, J., & Newman, G. A. (2020a). Attentional efficiency does not explain the mental state× domain effect. *PLOS ONE*, *15*(6), Article e0234500. https://doi.org/10.1371/journal.pone.0234500

Sweetman, J., & Newman, G. A. (2020b). Replicating different roles of intent across moral domains. *Royal Society Open Science*, *7*(5), 190808. https://doi.org/10.1098/rsos.190808

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, *96*(3), 506. https://doi.org/10.1037/0033-295X.96.3.506

Twardawski, M., & Hilbig, B. E. (2020). The motivational basis of third-party punishment in children. *PLOS ONE*, *15*(11), e0241919. https://doi.org/10.1371/journal.pone.0241919

Tybur, J. M., Molho, C., Cakmak, B., Cruz, T. D., Singh, G. D., & Zwicker, M. (2020). Disgust, anger, and aggression: Further tests of the equivalence of moral emotions. *Collabra: Psychology*, *6*(1), 34. https://doi.org/10.1525/collabra.349

Uskul, A. K., Kirchner-Häusler, A., Vignoles, V. L., Rodriguez-Bailón, R., Castillo, V. A., Cross, S. E., Yalçın, M. G., Harb, C., Husnu, S., Ishii, K., Jin, S., Karamaouna, P., Kafetsios, K., Kateri, E., Matamoros-Lima, J., Liu, D., Miniesy, R., Na, J., Özkan, Z., & . . .Uchida, Y. (2023). Neither Eastern nor Western: Patterns of independence and interdependence in Mediterranean societies. *Journal of Personality and Social Psychology*, *125*(3), 471–495. https://doi.org/10.1037/pspa0000342 van den Bos, W., van Dijk, E., & Crone, E. A. (2012). Learning whom to trust in repeated social interactions: A developmental perspective. *Group Processes & Intergroup Relations*, *15*(2), 243–256. https://doi.org/10.1177/1368430211418698

Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Preschoolers focus on others' intentions when forming sociomoral judgments. *Frontiers in Psychology*, *9*, Article 1851. https://doi.org/10.3389/fpsyg.2018.01851

Wittig, M., Jensen, K., & Tomasello, M. (2013). Five-year-olds understand fair as equal in a mini-ultimatum game. *Journal of Experimental Child Psychology*, *116*(2), 324–337. https://doi.org/10.1016/j.jecp.2013.06.004

Yang, F., Choi, Y. J., Misch, A., Yang, X., & Dunham, Y. (2018). In defense of the commons: Young children negatively evaluate and sanction free riders. *Psychological Science*, *29*(10), 1598–1611. https://doi.org/10.1177/0956797618779061

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*(2), 202–214. https://doi.org/10.1016/j.cognition.2011.04.005

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

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. https://doi.org/10.1037/xge0000613

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development*, *67*(5), 2478–2492. https://doi.org/10.1111/j.1467-8624.1996.tb01869.x