communications psychology

Article

A Nature Portfolio journal

https://doi.org/10.1038/s44271-025-00220-x

Across six societies children engage in costly third-party punishment of unfair sharing

Katherine McAuliffe[1] ✉, Samantha Bangayan[1], Tara Callaghan[2], John Corbit[2], Henry G. W. Dixson[3,4], Yarrow Dunham[5], Ann Finkel[6], Emily Otali[7], Sophie Riddick[1], Patrick Tusiime[7] & Felix Warneken[8]

Third-party punishment of unfairness shows striking cross-societal variation in adults, yet we know little about where and when in development this variation starts to emerge. When do children across societies begin to pay a cost to prevent unfair sharing? We present an experimental study of third-party punishment of unfair sharing across $N = 535$ children aged 5–15 from communities in six diverse countries: Canada, India, Peru, Uganda, USA, and Vanuatu. We tested whether children were more likely to punish equal or selfish (maximally unequal) distributions between two absent peers. We also tested whether decisions depended on whether such punishment was costly—participants had to sacrifice their own rewards to punish—or free. Our study generated three main findings. First, children across societies engaged in third-party punishment of selfishness: they were more likely to punish selfish than equal distributions. Second, older children were more likely than younger children to punish selfish sharing in Canada, India, Peru, and the USA. Third, children in Canada and the USA punished more in general in the Free condition than in the Costly condition, whereas children in Uganda punished selfishness more in the Costly condition. These findings show that children from six diverse societal contexts consistently took a stance against unfair sharing, in some cases even sacrificing their own rewards to intervene against selfishness in their peers. We highlight and discuss similarities and differences in cross-societal patterns of age-related differences in third-party punishment and suggest potential explanations for these patterns.

When uninvolved third parties punish selfish individuals, it is regarded as a strong indicator that they care about fairness norms. Even though third-party punishers are not directly affected by the selfish act, they nevertheless feel compelled to intervene against selfishness and do so even when the act of punishment comes at a personal cost[1,2]. Conceptually, third-party punishment has been regarded as an important test of genuine fairness concerns: it involves a witness rather than a victim inflicting punishment on an unfair transgressor and, for this reason, cannot easily be explained by revenge motives or the pursuit of personal gain. Empirical work shows that when adult participants tested in a third-party punishment task witness a player violating an egalitarian fairness norm by allocating more money to themselves than to another person, they often pay their own money to inflict a penalty on the unfair person[1]. While third-party punishment of this kind has been observed in adults across diverse societies, the extent to which people are willing to punish shows striking cross-cultural variation[2]. For instance, adults in some societies show a strong willingness to pay to punish selfish offers, while those in other societies show a substantially weaker willingness to pay to punish those same offers.

Children in the USA exhibit third-party punishment (see[3] for a review). By the age of six, they intervene as third parties to punish unfair distributions, and do so even when punishment involves paying a personal cost[4]. With age, children become more likely to calibrate the amount of punishment to the degree of inequality, with the goal of equalizing shares among third parties[5]. Children in the USA preferentially elect to punish perpetrators of unfairness rather than compensating victims of unfairness, highlighting their willingness to invest in punishment specifically, even

[1]Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA, USA. [2]St. Francis Xavier University, Department of Psychology, Antigonish, Nova Scotia, Canada. [3]Research School of Social Sciences, Australian National University, Canberra, Australia. [4]Institute for Culture & Society, Western Sydney University, Sydney, Australia. [5]Department of Psychology, Yale University, New Haven, CT, USA. [6]Department of Psychology, Harvard University, Cambridge, MA, USA. [7]Kibale Forest Schools Program, Kabarole, Uganda. [8]Department of Psychology, University of Michigan, Ann Arbor, MI, USA. ✉e-mail: mcaulikg@bc.edu

when other options are on the table[6]. In line with this, recent work shows that children in the USA intervene because they are opposed to the perpetrator's unfair behavior, rather than because they want to encourage the perpetrator to treat the intervener better in the future[7]. Other work conducted with children in the USA suggests that investment in third-party punishment is worthwhile: children are increasingly likely to make fair (equal) resource distribution decisions after receiving third-party punishment in response to unfair (unequal) distributions[8].

Work on the developmental emergence of third-party punishment of unfairness (typically indexed through punishment of selfishness), like work in developmental science more broadly[9], has focused almost exclusively on samples from WEIRD populations (Western, Educated, Industrialized, Rich, Democratic[10]). Thus, although we know that children in the USA engage in third-party punishment, we do not know whether it generalizes across diverse cultures and, by extension, the extent to which third-party punishment of fairness norm violations is shaped by cultural input.

Although we do not have a clear sense of cross-cultural variation in third-party punishment of unfair behaviors, there has been work with adults and children on cultural variation in fairness behavior itself. Fairness in these contexts is typically studied in relation to a norm of equality or egalitarianism as a central distributive fairness principle (other principles would include, for example, merit-based sharing or procedural justice). Ethnographic observations show that the rules that govern how food is divided among community members vary widely across societies[11]. Moreover, experimental economic games in which people divide resources between themselves and others reveal a high degree of cross-societal variation in what resource divisions are considered acceptable by adults[12]. Adding to these findings, developmental work points to childhood as a key period for the development of fairness behavior[13] as well as for the acquisition of norms related to fairness[14–17]. Work with children shows that this variation extends into development, with children showing cross-cultural variation in their rejections of advantageously unfair resource distributions[18], and their merit- and equity-based fairness decisions[19,20]. Together, studies of fairness behavior and norm acquisition lay the groundwork for the research we present here: We ask whether children across societies show signatures of a strong normative stance on fairness, as indexed through their third-party punishment of selfish behavior. More specifically, we address open questions about the development and expression of third-party punishment of selfishness across diverse cultural contexts. When does third-party punishment of selfishness emerge and what are the developmental patterns of punishment across societies? Addressing these question will help paint a fuller picture of the foundations of third-party fairness norm enforcement in humans.

Two recent studies have importantly expanded our understanding of the cross-cultural development of third-party intervention, broadly conceived. First, Kanngiesser et al.[21] found that 5- to 8-year-olds in eight populations intervened—both with verbal protest and nonverbal intervention—against peers who broke conventional rules in games (sorting toy blocks by color or shape). Yet, the interventions themselves showed cross-cultural variation: for instance, children in small-scale societies were more likely to use imperative verbal protests (e.g., saying "no") than rule protests (e.g., mentioning the convention specifically and/or using normative language like "should"), whereas children in large-scale societies did not display this distinction. This work therefore shows that children promote adherence to conventional rule games. While not testing third-party punishment of fairness norm violations, it highlights third-party intervention as a widespread phenomenon in children across societies.

Second, cross-cultural work by House and colleagues[15,22] tested how children learn and enforce different norms about resource sharing. Specifically, across a set of two studies conducted in Argentina, Ecuador, Germany, India, and the USA, children first participated in a study examining resource distribution decisions in response to normative prompts[15]. Findings revealed that children calibrate their own distributions to the "right" kind of distribution across societies. Next, children participated in a second study[22] on third-party punishment where participants learned about an actor who made a choice between a prosocial (equal) split of resources (1 for actor, 1 for recipient) or a selfish split (2 for actor, 0 for recipient). Children then received one of three prompts: they were either told that (1) selfishness is bad and punishing selfishness is good, (2) prosociality (equality in this case) is bad and punishing prosociality is good, or (3) both selfishness and prosociality are ok and punishing either is ok. Then, children could pay one of their own resources to remove a resource from the actor. Results showed that third-party punishment emerged across societies, that norm prompts influenced children's punishment, and that children were more likely to follow the norm endorsing the punishment of selfishness than the norm endorsing the punishment of prosociality. Thus, when prompted to punish, children punish selfishness more than prosociality (as indexed by equality) and are sensitive to how normative information aligns with their existing concepts about how to divide resources.

Here we explore the developmental emergence of children's unprompted third-party punishment of selfishness across six diverse cultural contexts. We employ an established third-party punishment task[4] to establish when and where children punish selfish (maximally unequal) resource distributions relative to equal allocations. In a between-subject design, we test whether punishment is influenced by the costs of intervention. In the Costly condition, children must sacrifice their own resources to punish, while in the Free condition, punishment does not involve personal sacrifice. Using this design, we test three hypotheses: Hypothesis 1, children across societies will punish selfish distributions more than equal distributions; Hypothesis 2, punishment of selfishness will increase with age; and Hypothesis 3, rates of punishment will be higher in the Free condition than in the Costly condition.

Our cross-cultural developmental approach positions us to explore whether patterns of punishment and its developmental trajectory show general consistency or whether they show cross-societal variation. Based on past work, we expected to see third-party punishment in the USA and our aim was to examine whether this pattern would generalize to other populations. In exploratory analyses, we were also interested in capturing variation in punishment behavior across societies and asking where and when this variation—if it is observed—emerges in childhood.

Finally, in addition to third-party punishment decisions, we assess children's verbal endorsement of distributive norms (i.e., what they consider to be the 'right' way to distribute resources) and intervention norms (i.e., whether intervening against norm violators is the 'right' thing to do) to explore whether any observed variation in punishment decisions can be explained by variation in children's normative assumptions about what constitutes fair sharing and their attitude toward third parties stepping up to intervene against norm violators.

Methods

Participants

Participants were $N = 535$ children 5 through 15 years old across six countries: Canada, India, Peru, Uganda, USA, and Vanuatu (Fig. 1; see SOM for site information). The six sites varied across a number of dimensions including setting (urban versus rural), dominant religion, and economy (Table 1).

Our plan was to focus on children in the 6- to 11-year-old age range, recruited into predetermined bins for the purpose of counterbalancing (6–7, 8–9, 10–11). Our target was roughly 20 participants per cell (condition and age group). Ultimately, however, recruitment varied depending on site-specific recruitment opportunities and challenges, so age ranges varied slightly by country, with the youngest children ranging from 5 to 6 years and the oldest children ranging from 9 to 15 years (see Table S1). An additional 38 children (less than 7% of the total sample) were tested but excluded a priori (e.g., outside our age range, pilot sessions) or post-hoc (e.g., experimenter error, consistent difficulties following instructions; see SOM for exclusions by site).

Design

In a 2×2 design, we manipulated both the cost of intervention (condition: Costly vs. Free) and the reward distribution (selfish vs. equal). Children were

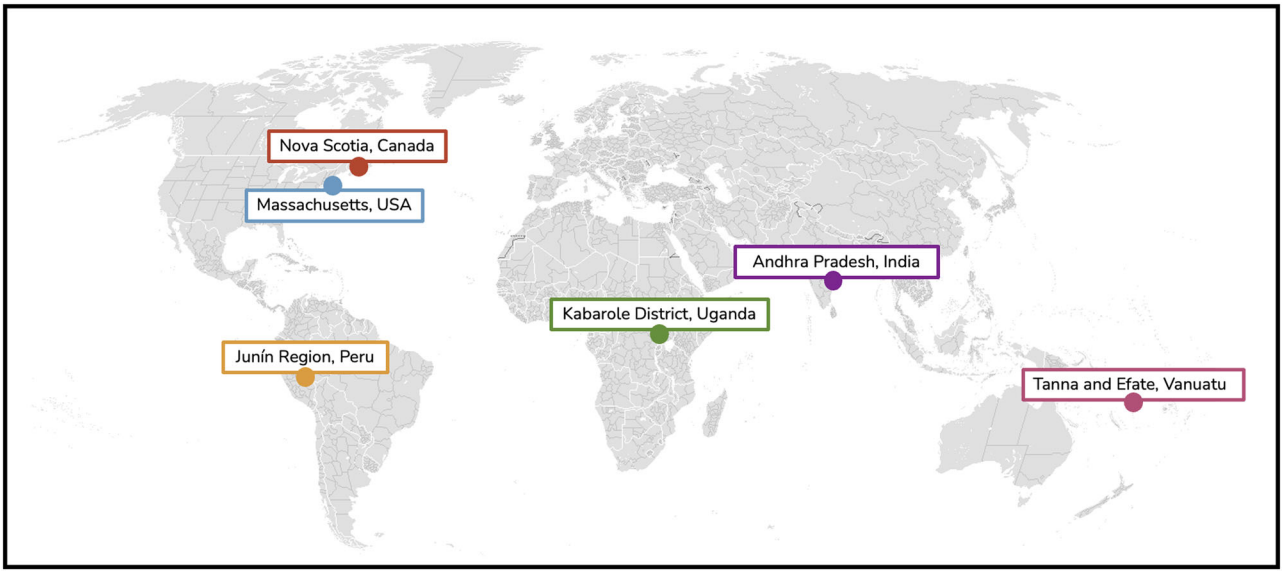

Fig. 1 | Map of data collection sites. Map image from Wikimedia commons.

assigned to either the Costly condition (where children could prevent a given resource distribution by paying one of their own candies or enact it without paying) or the Free condition (where children could prevent or enact resource distributions without paying). Assignment was not fully random because we endeavored to keep our sample in each site roughly balanced across condition and age group. See SOM for sample breakdown across conditions within sites.

In addition, we manipulated distribution within subject, with children responding to 6 equal trials (the divider allocated 3 rewards to the self and 3 rewards to the partner) and 6 selfish trials (the divider allocated all 6 rewards to the self and none to the partner). Trial order was randomized with the constraint that no more than two of the same trial type would be presented consecutively.

Materials and Procedure

Children indicated their choices by placing their own resources in containers representing "punish" or "enact" decisions. Following McAuliffe et al.[4], our setup for the Costly condition involved two boxes (Fig. 2), one red (punish) and one green (enact). Children knew that they could keep rewards they put in the green box but not the red box. The Free condition involved a single box, with one half colored red (punish) and the other half colored green (enact). Resources placed in either the red or green opening ended up in the same place, and children knew they could keep all rewards from the box (see Fig. S1). Decisions to place a reward into the red opening prevented the rewards from being distributed: i.e., these decisions removed six rewards from the divider in unfair trials and three rewards from the divider and three from the partner in fair trials. Because placing rewards into the red openings inflicted a cost on dividers, these decisions meet the definition of punishment.

Due to constraints in the availability of materials, our tests in Vanuatu involved plastic cups sitting on plates instead of boxes. The logic was the same: In the Free condition, both cups had missing bottoms and at the end of the task, participants would collect the rewards that had fallen through onto the plates. In the Costly condition, the green cup was open and the red cup had a bottom so that participants would only collect the resources invested in the green cup that fell onto the plates (see SOM for images).

Procedure

Third-party punishment task. The procedure was based closely on the task by McAuliffe et al.[4]. Children were told about two absent children who had allegedly participated in the task previously. The names of these two children were gender-matched to the participant and varied according to the local language. One of the children was the divider, whose responsibility was to divide rewards between themselves and the other child, the recipient. The participant then learned that they would be making a series of decisions by moving their resources into different container openings. Depending on condition, they were introduced either to the Costly condition containers or the Free condition container.

In the 'Costly' condition, children learned that they could enact the divider's distribution by putting one of their resources through the hole of the green container. Children knew that at the end of the session, they could take all resources from the green container with them, making this decision cost-free. On the other hand, children could punish the divider by putting one of their resources through the hole of the red container. Any resources moved into the red container were taken away at the end of the game, making this decision costly.

In the 'Free' condition, children learned that they could enact the divider's distribution decision by putting one of their resources through the hole of the green side of the single container and they could punish the divider by moving one of their resources through the hole of the red side of the container. Here, all resources ended up back in the same container and children knew that they could take them home at the end of the session, such that neither enacting nor punishing carried personal costs.

In each trial, the experimenter manipulated the six resources allocated by the divider in accordance with the participant's decision. If the participant chose to punish, all resources were removed from the area, meaning neither the divider nor the recipient received any rewards. If the participant decided to enact, the experimenter distributed the divider's distribution to the absent divider and recipient's bags.

We included several comprehension questions to assess children's understanding of the consequences of each decision, the distribution cards (Fig. 2), and the absent children. In general, children passed comprehension checks across sites. Children were not excluded for incorrect responses but were reminded of the correct answer if they got a question wrong. At the end of the task, children were asked to explain their decisions in two open-ended prompts. Due to transcription inconsistencies across sites, these open-ended responses were not examined further. Children were also asked whether they thought the absent children were real or just pictures (see SOM). Finally, across all sites except India, children were told that the absent children were not, in fact, real. Please see SOM for the full script along with details about comprehension checks and details regarding slight variations in the procedure that were required for testing across different sites.

Equality and intervention scenarios. At the end of the task, children were presented with two sets of scenarios that assessed their

Table 1 | Information about testing settings within each site

Site	Canada	India	Peru	Uganda	USA	Vanuatu	
	Antigonish, Nova Scotia	Challapalli, Andhra Pradesh	San Pedro de Sano, Junín Region	Area of Fort Portal, Kabarole	Boston, Massachusetts	Efate	Tanna
Setting	Rural	Rural	Rural	Rural	Urban	Urban	Rural
Language	English	Telugu	Spanish	Rutooro	English	Bislama	Bislama, Nafe
Dominant Religion	Christian	Hindu	Christian	Christian	Christian	Christian	Christian
Economy	Professional, trade/service, agriculture	Agriculture, labor	Agriculture, labor	Agriculture, labor	Professional, trade/ service, labor	Casual labor, tourism, service sector, agriculture	Subsistence Horticulture
Household structure	Nuclear	Extended	Extended	Extended	Nuclear	Extended	Extended
Adult literacy	High	Medium	High	Medium	High	High	Medium
Formal schooling	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child possessions	Many	Few	Few	Few	Many	Few	Few

Model output tables.

understanding of norms governing sharing and intervention: sharing norm scenarios (Set 1) and intervention norm scenarios (Set 2).

For Set 1, the sharing norm scenarios, children saw one scenario in which a set of six resources could be split between two agents and were asked to demonstrate how they thought the six resources should be divided. Children were free to distribute the resources in whatever way they wanted (*demonstration question*). We were specifically interested in the number of children who demonstrated that equality (3-3) was the 'right' way to divide resources. The other scenario in Set 1 asked children to evaluate selfishness. In this scenario, they were told about a situation in which an actor kept six resources and shared none. Children were asked if this was ok or not ok (*ok or not ok question*). We were interested in the number of children who said that selfishness was 'not ok'. Set 1 scenarios described either cookies or peanuts as the resources being divided. In India, Uganda, and the USA, the 'demonstration' question was paired either with the cookie or peanut scenario while the 'ok or not ok' question was paired with the opposite scenario. This meant that we had two data points per child for the Set 1 scenarios in India, Uganda, and the USA. In Canada and Peru, on the other hand, children were asked both question types for both the cookie and peanut scenarios, meaning we had four data points per child for the Set 1 scenarios in Canada and Peru.

For Set 2, the intervention norm scenarios, children saw two scenarios in which, after a transgression occurred, one witness intervened while the other was a passive bystander. Children were asked whether the transgression was ok or not ok to ensure that they interpreted the story as we had planned. Children were then asked whether the intervener or non-intervening bystander did the right thing. Set 2 stories involved a transgressor withholding a football from a game, thereby preventing others from playing with it, and a transgressor cutting the lunch queue. Note also that scenarios were implemented everywhere except Vanuatu due to time constraints. See SOM for script.

Across sites, we confirmed with translators that the language in these scenario questions conveyed that we were asking about the 'right' behaviors in the normative sense of the word.

Coding and analysis. Children's responses were live coded and video recorded where possible. A subset of videos was compared to live coding to provide an estimate of reliability for our distribution (equal, selfish) and decision (punish, did not punish) variables. Reliability was high with very few discrepancies between live and video coding (< 2% of trials across all sites; see SOM for reliability by site).

We conducted generalized linear mixed models (GLMMs) with the 'lme4' package[23] in R version 4.4.0. Our response term was binary (punish = 1, did not punish = 0) with participant identities (ID) fit as random intercepts to account for repeated measures. Because data were binomially distributed, it was assumed that data met assumptions of logistic regression, but this was not formally tested. Our modeling approach mirrored our three key hypotheses. We first explored whether punishment was predicted by distribution (Hypothesis 1). Second, we asked whether punishment increased with age (Hypothesis 2). Third, we examined whether punishment was more likely in the Free than Costly condition (Hypothesis 3). In testing our three hypotheses, we initially ran models on the full sample and then tested each hypothesis within the different sites. This study was not pre-registered.

We first created a null model that consisted solely of gender and participant ID. This null model was then compared to a full model which contained all main effects and interactions of interest. This approach helps attenuate Type 1 error rates[24]. Predictors of interest were distribution (equal, selfish), condition (Costly, Free), age (years), and site (Canada, India, Peru, Uganda, USA, Vanuatu) and the two-way interactions between distribution x age as well as distribution x condition. In addition to this full model with interactions, we created a main effects model to provide an easily-interpretable test of Hypothesis 1. To facilitate comparison of effects across our full interactions model and the main effects model, we effect coded the distribution term (equal = −0.5, selfish = 0.5) and condition term

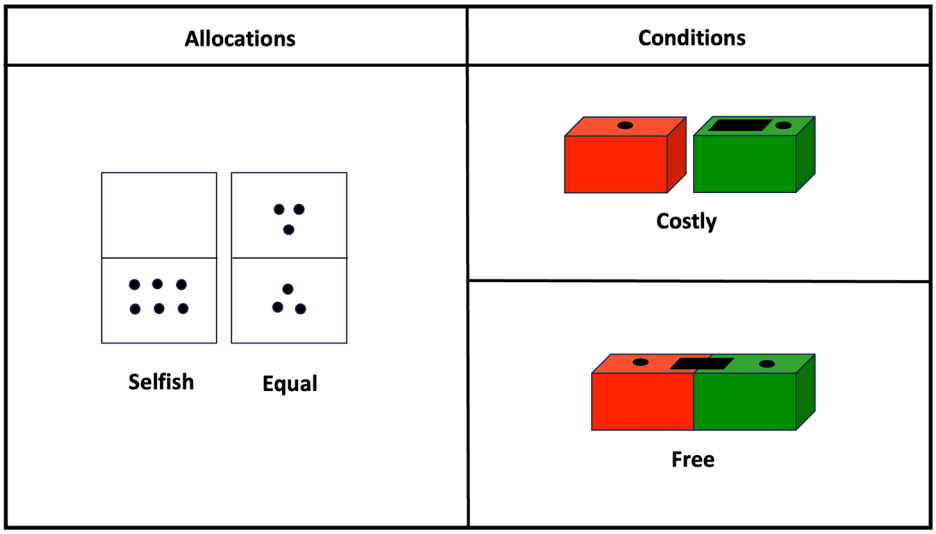

Fig. 2 | Distribution cards illustrating the absent divider's decision. Decisions were either selfish (divider kept six and shared none) or equal (divider kept three and shared three). Children were presented with 12 trials, six of which were selfish and six of which were equal. Between–subject, children were assigned to either the Costly or Free condition. In the Costly condition, children sacrificed one of their resources to punish (red box) but enacting decisions was free (green box). In the Free condition, both punishing (red side) and enacting (green side) were free.

(Free = −0.5, Costly = 0.5). We additionally scaled age using the 'scale' function, which mean-centered age values and divided them by the standard deviation. We present the main effects model under Hypothesis 1 but wish to note that the effect of distribution remains consistent in the full interactions model which is used to test Hypotheses 2 and 3.

The comparison between the full model and the null was significant for both the main effects model and the interactions model (using 'anova': likelihood ratio tests, $\chi^2_{11} = 880.43$, $p < 0.001$ and $\chi^2_{13} = 916.31$, $p < 0.001$, respectively). We tested for the significance of terms by removing them from a model and comparing the reduced model to a model that included the term of interest. Results from these model comparisons were conducted using the drop1 command and are reported as likelihood ratio tests (LRTs). Non-significant terms were dropped from models. Both full and reduced models are reported in the SOM. Note that recruitment in Vanuatu took place across two sites (see Table 1). Because we had no predictions that these sites would differ, our initial analyses include Vanuatu at the country-level in models. However, in an exploratory model described below, we model the two sites separately. Additionally, we ran supplementary models with the two Vanuatu sites included as separate levels of a fixed effect (Table S19). Reported confidence intervals (CIs) were computed using the Wald method. Data figures show predicted effects from models using 'ggpredict' from the ggeffects package[25] and ggplot[26]. Raw data figures are presented in the SOM, with confidence intervals adjusted using the Agresti-Coull method[27].

Note that age was not evenly distributed across years, and age ranges as well as age distributions varied by site. Models with interactions between age and distribution are run on restricted data sets that trimmed the age range by removing age cells with fewer than $N = 10$ participants per site. Because the effects vary slightly depending on how age was modeled, we privilege the findings based on narrower age ranges because they have more balanced coverage across years, even though this approach resulted in a slightly smaller sample overall ($N = 523$ rather than $N = 535$ for full sample). Following this same rule, we additionally restricted age ranges for within country analyses that included interactions between age and distribution which resulted in slightly smaller sample sizes in Canada ($N = 105$ rather than 106), India ($N = 47$ rather than 49), Uganda ($N = 88$ rather than 95), and Vanuatu ($N = 122$ rather than 130). Results from models with the complete data set are included in the SOM. The main effects model reported under Hypothesis 1 includes children of all ages to show generalizability of the distribution effect. Finally, for ease of interpretability, plots presented in the main text involving age are from models with age in years and non-effect-coded predictors, but plots with scaled age and effect-coded predictors can be found in the SOM (Figs. S10, S11).

For our three scenario measures—(1) what is the right distribution, (2) is a selfish distribution ok or not ok, and (3) who did the right thing: the intervener or the non-intervening bystander—measures were tallied by question within each site and then a proportion was calculated. We focus on proportions because, as described above, scenario data varied slightly by site (e.g., in Canada and Peru, children got both question types for the 'cookie' and 'peanuts' scenarios while in India, Uganda, and the USA, they got one question type for the 'cookie' scenario and the other for 'peanuts'). Our intervention tallies involved summing responses for the two scenarios. To compare proportions, we used proportion tests and conducted Bonferroni-corrected post-hoc tests where differences across populations were detected. See SOM for details.

Inclusion and ethics in global research. Research was conducted in collaboration with local researchers who were involved in implementation. Local researchers who were centrally involved in implementation and interpretation are authors on the paper. Local researchers who contributed to data collection are acknowledged. Research was approved by university ethics committees (Harvard University protocol #F18470; St. Francis Xavier protocol #22359; Australian National University protocol #2012/208) as well as local ethics committees where available (Uganda: REIRC IN-COM 111/2014). In Vanuatu, formal permission was granted by the Vanuatu Kaljoral Senta and the review process was conducted by the Malvatumauri Council of Chiefs. Research was conducted according to the same high standard across sites. Informed consent was given by participating families. Where available, we have integrated local research and expertise into this paper.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Hypothesis 1: children will punish selfishness more than equality

We found support for the hypothesis that children would be more likely to intervene against selfish than equal distributions (Fig. 3). Our main effects model showed that distribution was a significant predictor of children's punishment (LRT, $\chi^2_1 = 838.93$, $p < 0.001$, $\beta = 1.82$, CI = 1.69–1.95; Table 2): children were more likely to punish selfish distributions than equal distributions. Exploring the effect of distribution based on main effects models *within* each country showed that this effect was consistent across all sites: all within-country main effects models showed a significant effect of distribution (Canada: LRT, $\chi^2_1 = 586.7$, $p < 0.001$, $\beta = 3.91$, CI = 3.48–4.34; India: LRT, $\chi^2_1 = 48.56$, $p < 0.001$, $\beta = 1.29$, CI = 0.91–1.66; Peru: LRT, $\chi^2_1 = 96.13$, $p < 0.001$, $\beta = 1.37$, CI = 1.08–1.65; Uganda: LRT, $\chi^2_1 = 15.29$ $p < 0.001$, $\beta = 0.6$, CI = 0.3–0.9; USA: LRT, $\chi^2_1 = 143.72$ $p < 0.001$, $\beta = 2.11$,

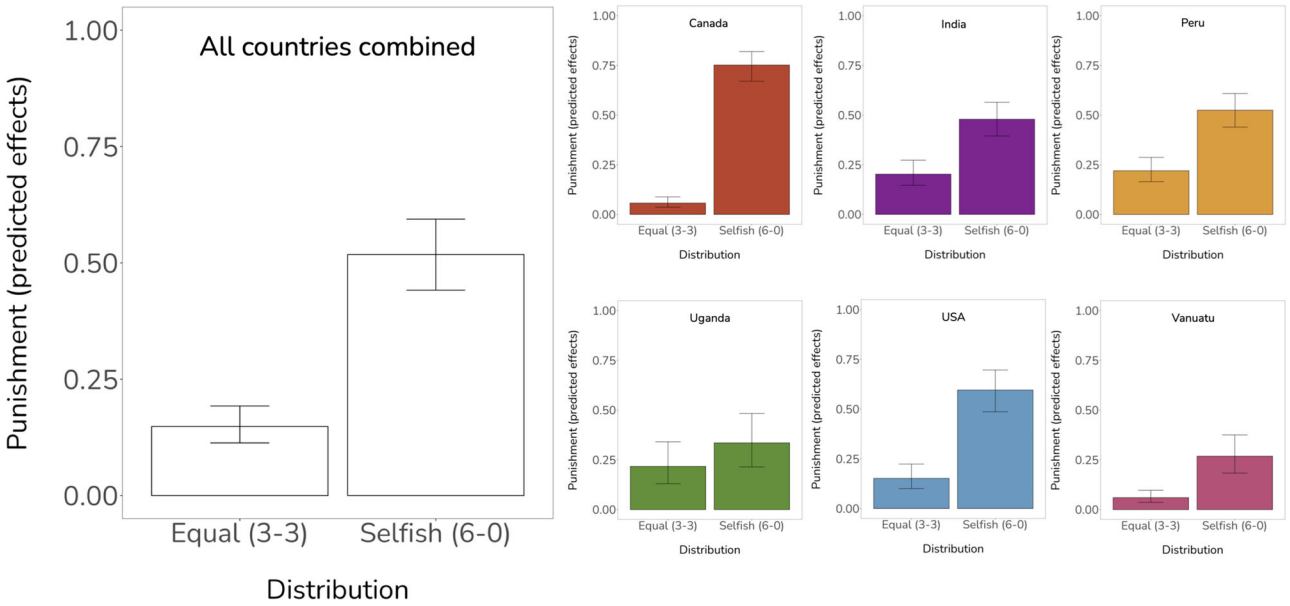

Fig. 3 | Plots of predicted effects (probabilities) from main effect models predicting punishment. The large plot shows the predicted effects from a model combining data from all sites ($N = 535$). This model included distribution, gender, age, site, and condition (Free, Costly). Individual plots show the effect from country-specific models that included distribution, gender, age (scaled), and condition (Canada, $N = 106$; India, $N = 49$; Peru, $N = 93$; Uganda, $N = 95$; USA, $N = 62$; Vanuatu, $N = 130$). Across all models shown here, gender, distribution, and condition were effect coded, age was scaled, and models included participant identity fit as random intercepts. Error bars show 95% confidence intervals.

CI = 1.73–2.49; Vanuatu: LRT, $\chi^2_1 = 151.71$ $p < 0.001$, $\beta = 1.75$, CI = 1.45–2.05; Table 2). These results suggest that third-party punishment of selfishness is widespread, providing evidence that children across countries show the signature of a strong response against unfair sharing—they are willing to intervene to prevent unfair treatment even when they are not affected by it.

Hypothesis 2: punishment will increase with age

When looking across countries, we found a significant interaction between distribution and age in both our full and reduced models (Full model: LRT, $\chi^2_1 = 30.23$ $p < 0.001$; $\beta = 0.41$, CI = 0.26–0.55; Reduced model: LRT, $\chi^2_1 = 30.27$, $P < 0.001$; $\beta = 0.41$, CI = 0.26–0.55; Table 3; Fig. 4), suggesting that children's punishment of selfish versus equal distributions varied across age. Inspection of Fig. 4 shows that across most sites—namely, Canada, India, Peru, and USA—older children were more likely to punish selfish (relative to equal) distributions than younger children. Indeed, our country-specific analyses (see Table 3) showed significant positive interactions between age and distribution across these sites (Canada: LRT, $\chi^2_1 = 42.59$ $p < 0.001$, $\beta = 1.64$, CI = 1.11–2.18; India: LRT, $\chi^2_1 = 27.94$ $p < 0.001$, $\beta = 1.24$, CI = 0.76–1.71; Peru: LRT, $\chi^2_1 = 28.98$, $p < 0.001$, $\beta = 0.86$, CI = 0.54–1.18; USA: LRT, $\chi^2_1 = 65.9$ $p < 0.001$, $\beta = 2.77$, CI = 2.76–2.77). By contrast, in Vanuatu we saw a negative interaction such that older children's punishment was less sensitive to distribution than younger children's punishment (LRT, $\chi^2_1 = 24.33$, $p < 0.001$, $\beta = -0.78$, CI = −1.09–−0.46; Table 3; see below for exploratory analyses within Vanuatu). In Uganda we saw a marginal interaction between age and distribution (LRT, $\chi^2_1 = 3.27$, $p = 0.07$, $\beta = 0.35$, CI = −0.02–0.71): older children were slightly more likely to distinguish between their punishment of selfish versus equal distributions (see below for exploratory analyses within Uganda) than were younger children.

Hypothesis 3: children will punish more in the Free than Costly condition

When looking across countries, we found no statistically significant evidence for an interaction between condition and distribution. When this term was removed from the model, we found a significant main effect of condition (LRT, $\chi^2_1 = 10.47$, $p = 0.001$, $\beta = -0.41$, CI = −0.67–−0.16; Table 3). This effect was also seen in our main effects model (LRT, $\chi^2_1 = 8.11$, $p = 0.004$, $\beta = -0.36$, CI = −0.61–−0.11; Table 2). Specifically, across both distribution types, children were more likely to intervene when doing so was free compared to when it was costly. In our within-country analyses, we found a main effect of condition in Canada (LRT, $\chi^2_1 = 10.99$, $p = 0.001$, $\beta = -0.94$, CI = −1.5–−0.39; Table 3) and the USA ($\chi^2_1 = 11.13$, $p = 0.001$, $\beta = -1.02$, CI = −1.03–−1.01; Table 3), but we found no statistically significant evidence for the effect of condition in India or Vanuatu. In Peru, we saw a marginal effect of condition (LRT, $\chi^2_1 = 3.11$, $p = 0.08$, $\beta = -0.42$, CI = −0.89–0.05). In Uganda, we found a significant interaction between condition and distribution (LRT, $\chi^2_1 = 6.12$, $p = 0.01$, $\beta = 0.27$, CI = 0.18–1.42). Inspection of Fig. 5 suggests that this interaction is due to children's increased willingness to punish selfish relative to equal distributions in the Costly condition (see below for exploratory model within Uganda). These data suggest that while there is a general effect of condition when examining the whole data set, such that children are more willing to intervene when doing so is free compared to when it is costly, this effect is likely driven by children in Canada and the USA. In three of our six sites we did not see a condition difference. Moreover, in Uganda, the condition effect was, surprisingly, in the opposite direction, with children showing a greater willingness to intervene against selfishness when doing so involved paying a personal cost.

Does the development of third-party punishment of selfishness vary across sites?

To explore whether the developmental trajectory of third-party punishment of selfishness shows cross-societal variation, we built an exploratory model predicting punishment as a function of the three-way interaction between distribution, age, and site (Fig. 6; Table S16), finding that this interaction was significant (LRT, $\chi^2_5 = 135.93$, $p < 0.001$; $\beta^{India} = 0.22$, CI = −0.35–0.79; $\beta^{Canada} = 0.76$, CI = 0.14–1.39; $\beta^{Uganda} = -0.53$, CI = −0.99–0.059; $\beta^{USA} = 1.97$, CI = 1.16–2.79; $\beta^{Vanuatu} = -1.57$, CI = −2.01–−1.12). Inspection of Fig. 6 indicates that punishment of unfair distributions is more likely in older than in younger children across Canada, India, Peru, and the USA. By contrast, in Uganda and Vanuatu, punishment is less likely in older than younger children.

To better understand this interaction, we conducted two additional analyses exploring the two-way interactions between age and distribution in

Table 2 | Estimate and standard errors of fixed effects in Generalized Linear Mixed Models predicting children's punishment

	Null	Full	Canada	India	Peru	Uganda	USA	Vanuatu
Intercept	−0.82***	−0.84***	−0.74***	−0.71***	−0.58***	−1.01**	−0.63*	−1.80***
	(0.08)	(0.16)	(0.20)	(0.16)	(0.17)	(0.31)	(0.25)	(0.28)
Gender (Boy)	0.08	0.09	−0.47	0.01	−0.22	0.78	0.18	0.34
	(0.11)	(0.13)	(0.27)	(0.23)	(0.23)	(0.45)	(0.29)	(0.33)
Distribution (EC)		1.82***	3.91***	1.29***	1.37***	0.60***	2.11***	1.75***
		(0.07)	(0.22)	(0.19)	(0.15)	(0.15)	(0.20)	(0.15)
Age (scaled)		−0.02	0.43**	0.18	0.19	−0.52*	0.14	−0.14
		(0.07)	(0.16)	(0.13)	(0.13)	(0.23)	(0.25)	(0.15)
Site: India		−0.08						
		(0.25)						
Site: Peru		−0.01						
		(0.21)						
Site: Uganda		0.21						
		(0.21)						
Site: USA		0.19						
		(0.23)						
Site: Vanuatu		−0.76***						
		(0.20)						
Condition (EC)		−0.36**	−0.83**	−0.11	−0.40	0.44	−0.90**	−0.51
		(0.13)	(0.27)	(0.23)	(0.23)	(0.45)	(0.29)	(0.33)
AIC	7777.82	6913.39	1046.69	720.09	1333.34	1239.92	835.95	1459.66
BIC	7798.12	6987.81	1077.58	746.35	1363.44	1270.14	863.61	1491.77
Log Likelihood	−3885.91	−3445.69	−517.35	−354.05	−660.67	−613.96	−411.98	−723.83
Num. obs.	6411	6411	1271	588	1115	1137	742	1558
Num. groups: new.id	535	535	106	49	93	95	62	130
Var: new.id (Intercept)	1.21	1.58	1.18	0.24	0.80	3.94	0.74	2.62

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

Baselines were set as follows: Gender = Girl. Distribution and Condition were effect coded (EC; equal = −0.5; selfish = 0.5; Free = - 0.5, Costly = 0.5). Table also shows goodness-of-fit statistics.

subsets of selfish and equal trials, respectively. Within selfish trials, we found that the two-way interaction between distribution and site was significantly different at the levels of Peru (baseline) and Uganda (β = −1.11, CI = −1.82–−0.40) and Vanuatu (β = −1.04, CI = −1.69–−0.4; Table S16). In Uganda and Vanuatu, punishment was less likely in older than younger children relative to Peru where punishment was more likely in older than younger children. When we subsetted to equal trials only, we found a difference at the level of Vanuatu relative to Peru (β = 0.71, CI = 0.17–1.24; Table S16): punishment of equality was more likely in older than younger children in Vanuatu relative to Peru (Fig. 6). We also found a difference at the level of USA relative to Peru (β = −1.27, CI = −2.21–−0.33; Table S16): punishment of equality was less likely in older versus younger children in the USA relative to Peru. We explore punishment in Uganda and Vanuatu below, but a main take-away from this exploratory analysis is that we did indeed see variation in the developmental pattern of third-party punishment across different societies.

Exploring punishment in the contexts of Uganda and Vanuatu

Based on our exploration of the developmental trajectory of punishment across sites, Uganda and Vanuatu stood out, revealing interesting—and surprising—patterns of punishment. Here, we explore these in more detail, starting with Uganda and then turning to Vanuatu.

One surprising result from our data from Uganda was the interaction between distribution and condition. Counter to Hypothesis 3, children in Uganda were more likely to punish selfishness when doing so entailed costs than when it was free. Additionally, punishment was lower in older children than in younger children, although older children displayed more sensitivity to distribution than younger children (they punished both equal and selfish distributions at younger ages and their punishment of equality decreased across our age variable at a steeper rate than their punishment of selfishness; Fig. 4). We were interested in exploring whether the surprising condition effect and the surprising age trend were related. To this end, we conducted an exploratory model in which we predicted punishment as a function of the three-way interaction between distribution, age, and condition. This model revealed a marginally significant three-way interaction (LRT, $\chi^2_1 = 3.69$, $p = 0.055$, β = 0.75, CI = 0.75–0.75; Fig. 7a; Table S17). In the Free condition, punishment of both equal and selfish distributions was lower in older versus younger children. By contrast, in the Costly condition, punishment of selfish distributions was higher in older versus younger children while punishment of equal distributions was lower in older versus younger children.

With respect to Vanuatu, it is important to note that data from two sites were combined to create our full sample for Vanuatu. These sites were Efate, an urban site (see Table 1 and SOM), and Tanna, a rural site. Because we had no a-priori predictions that these sites would be different, the analyses presented above combine data from both sites (though see Table S19 for models with these sites fit separately). However, plotting data by site revealed pronounced differences which can help us understand the surprising age effect noted in our exploration of cross-site differences in the development of punishment (Fig. 7b). To explore the potential difference between our two sites within Vanuatu, we conducted a model with punishment as a function of the interaction between distribution and site (Efate, Tanna). This model revealed a significant effect of this two-way interaction (LRT, $\chi^2_1 = 98.95$, $p < 0.001$, β = −3.2, CI = −3.88–−2.51; Table S17). Specifically, children in

Table 3 | Estimate and standard errors of fixed effects in Generalized Linear Mixed Models predicting children's punishment

	Null	Full	Reduced	Canada Full	Canada Reduced	India Full	India Reduced	Peru Full	Peru Reduced	Uganda	USA Full	USA Reduced	Vanuatu Full	Vanuatu Reduced
Intercept	−0.80***	−0.80***	−0.80***	−0.82***	−0.85***	−0.80***	−0.79***	−0.58***	−0.58***	−1.01**	−0.76**	−0.80***	−1.90***	−1.90***
	(0.08)	(0.15)	(0.16)	(0.22)	(0.22)	(0.18)	(0.18)	(0.17)	(0.17)	(0.32)	(0.28)	(0.00)	(0.29)	(0.29)
Gender (Boy)	0.06	0.07	0.07	−0.46	−0.45	0.05	0.04	−0.25	−0.25	0.75	0.23	0.23***	0.36	0.35
	(0.11)	(0.13)	(0.13)	(0.28)	(0.28)	(0.25)	(0.25)	(0.24)	(0.24)	(0.48)	(0.30)	(0.00)	(0.36)	(0.36)
Distribution (EC)		1.87***	1.88***	4.42***	4.45***	1.52***	1.50***	1.40***	1.41***	0.59***	3.96***	4.01***	2.27***	2.26***
		(0.07)	(0.07)	(0.27)	(0.27)	(0.21)	(0.21)	(0.15)	(0.15)	(0.16)	(0.36)	(0.00)	(0.19)	(0.19)
Age (scaled)		−0.03	−0.03	0.07	0.05	−0.02	−0.02	0.13	0.13	−0.46	−0.02	−0.04***	0.10	0.10
		(0.07)	(0.07)	(0.18)	(0.19)	(0.15)	(0.15)	(0.13)	(0.13)	(0.28)	(0.26)	(0.00)	(0.18)	(0.18)
Site: India		−0.08	−0.08											
		(0.25)	(0.25)											
Site: Peru		−0.03	−0.03											
		(0.20)	(0.21)											
Site: Uganda		0.10	0.10											
		(0.21)	(0.21)											
Site: USA		0.18	0.18											
		(0.23)	(0.23)											
Site: Vanuatu		−0.75***	−0.76***											
		(0.20)	(0.20)											
Condition (EC)		−0.41**	−0.41**	−0.83**	−0.94***	−0.23	−0.19	−0.40	−0.42	0.27	−0.96**	−1.02***	−0.55	−0.60
		(0.13)	(0.13)	(0.29)	(0.28)	(0.25)	(0.25)	(0.24)	(0.24)	(0.48)	(0.30)	(0.00)	(0.36)	(0.35)
Distribution x Age		0.41***	0.41***	1.63***	1.64***	1.25***	1.24***	0.87***	0.86***	0.35	2.77***	2.77***	−0.79***	−0.78***
		(0.07)	(0.08)	(0.27)	(0.27)	(0.24)	(0.24)	(0.16)	(0.16)	(0.19)	(0.37)	(0.00)	(0.16)	(0.16)
Distribution x Condition		−0.05		−0.53		0.41		−0.27		0.80*	−0.53		−0.30	
		(0.13)		(0.40)		(0.40)		(0.29)		(0.32)	(0.40)		(0.32)	
AIC	7641.01	6744.69	6742.75	998.86	998.62	664.65	663.71	1307.49	1306.36	1151.98	772.32	772.05	1355.77	1354.64
BIC	7661.24	6832.36	6823.67	1039.97	1034.59	699.33	694.06	1347.62	1341.47	1191.65	809.19	804.32	1398.08	1391.66
Log Likelihood	−3817.50	−3359.35	−3359.37	−491.43	−492.31	−324.33	−324.86	−645.74	−646.18	−567.99	−378.16	−379.03	−669.89	−670.32
Num. obs.	6268	6268	6268	1259	1259	564	564	1115	1115	1053	742	742	1463	1463
Num. groups: new.id	523	523	523	105	105	47	47	93	93	88	62	62	122	122
Var: new.id (Intercept)	1.16	1.53	1.54	1.19	1.21	0.26	0.26	0.83	0.83	4.13	0.74	0.76	2.89	2.86

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

Baselines were set as follows: Gender = Girl. Distribution and Condition were effect coded (EC; equal = −0.5; selfish = 0.5; Free = −0.5, Costly = 0.5). Table also shows goodness-of-fit statistics.

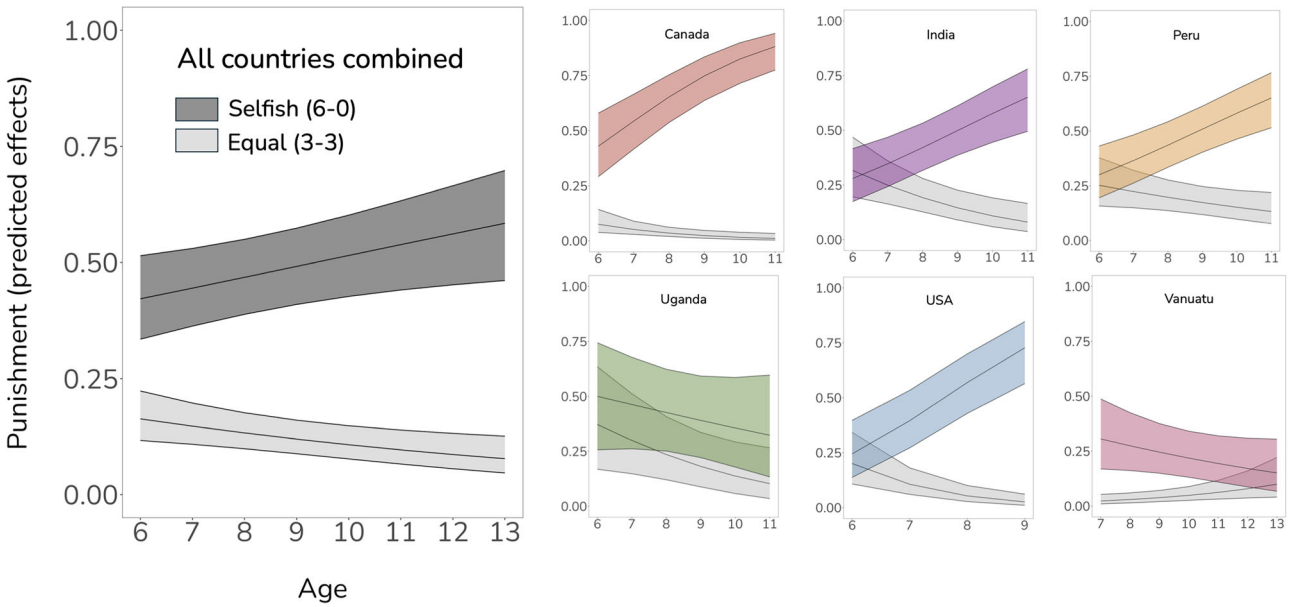

Fig. 4 | Plots of predicted effects (probabilities) from reduced interaction models predicting punishment. The large plot shows predicted effects from a model combining data from all sites ($N = 523$). This model included the interaction between distribution and age (years), gender, site, and condition. Individual plots show the effect from country-specific models that included the interaction between distribution and age, gender, and condition (Canada, $N = 105$; India, $N = 47$; Peru, $N = 93$; Uganda, $N = 88$; USA, $N = 62$; Vanuatu, $N = 122$). The Uganda model additionally included the interaction between distribution and condition. All models included participant identity fit as random intercepts. Ribbons show 95% confidence intervals.

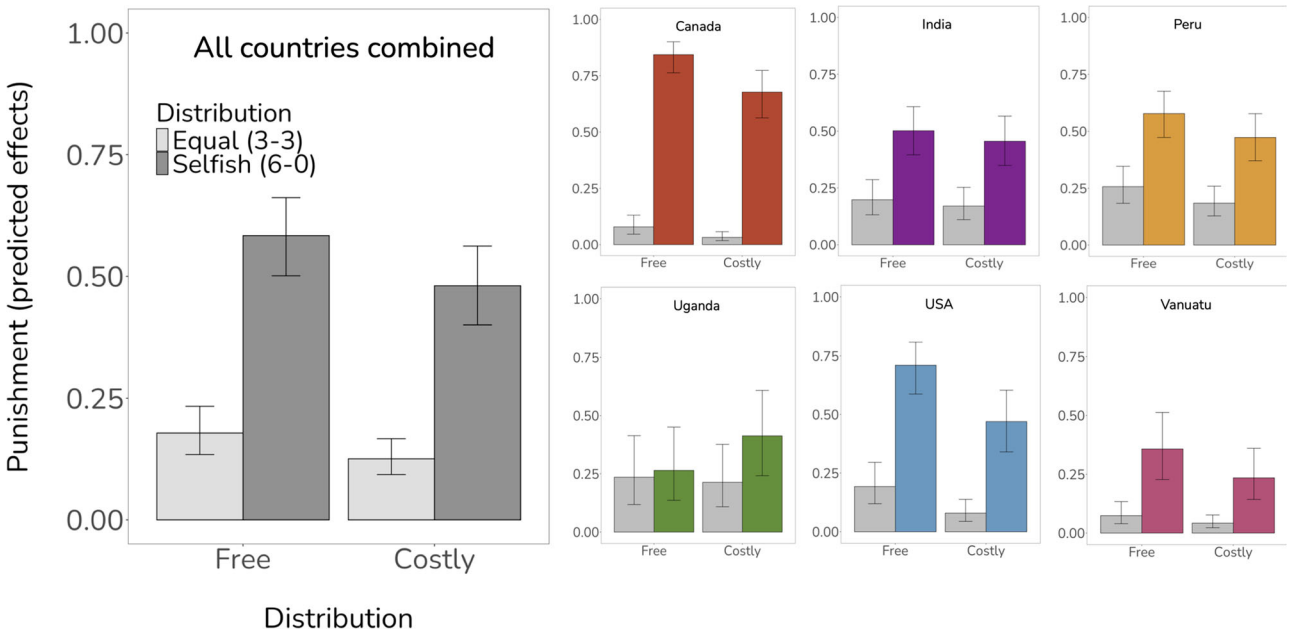

Fig. 5 | Plots of predicted effects (probabilities) from reduced interaction models predicting punishment. The large plot shows predicted effects from a model combining data from all sites ($N = 523$). This model included the interaction between distribution and age (years), gender, site, and condition. Individual plots show the effect from country-specific models that included the interaction between distribution and age, gender, and condition (Canada, $N = 105$; India, $N = 47$; Peru, $N = 93$; Uganda, $N = 88$; USA, $N = 62$; Vanuatu, $N = 122$). The Uganda model additionally included the interaction between distribution and condition. Across all models shown here, gender, distribution, and condition were effect coded, age was scaled, and models included participant identity fit as random intercepts. Error bars show 95% confidence intervals.

Efate punished selfish distributions at high rates and considerably more than equal distributions (mvt-corrected post-hoc test comparing punishment of selfish and equal distributions, estimate = −3.73, SE = 0.3, $p < 0.001$), whereas we saw much lower rates of punishment in Tanna. Despite low rates of punishment overall, children in Tanna were nonetheless more likely to punish selfish than equal distributions (mvt-corrected post-hoc test of selfish versus equal distributions, estimate = -0.54, SE = 0.18, $p = 0.017$). This urban-rural difference could account for the age trend revealed in our cross-site analyses (Fig. 6). Our younger sample was disproportionately urban and our older sample was disproportionality rural (Efate, mean = 8.4 years, SD = 0.68, restricted range = 7 (min) to 10 (max); Tanna mean = 10.5 years, SD = 2.26, restricted range = 6 (min) to 15 (max)).

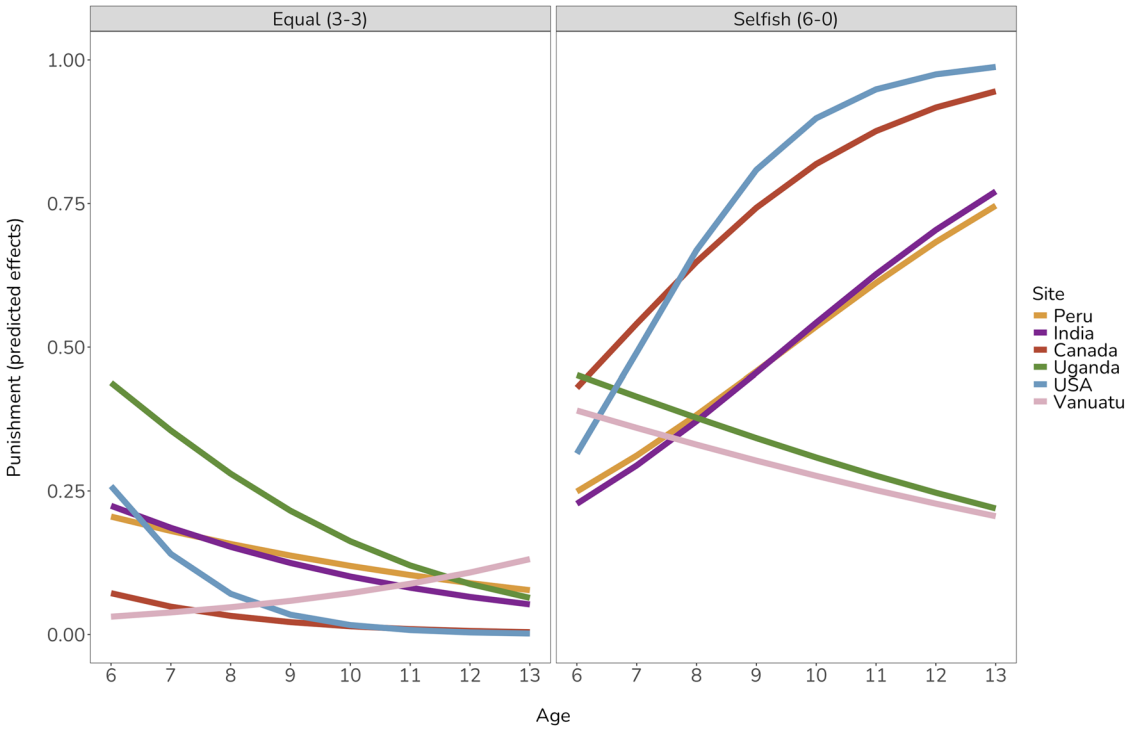

Fig. 6 | Plots of predicted effects (probabilities) from exploratory model predicting punishment as a function of distribution (equal versus selfish), age (years), and site (*N* = 523). This model additionally included gender as well as participant identity, which was fit as random intercepts.

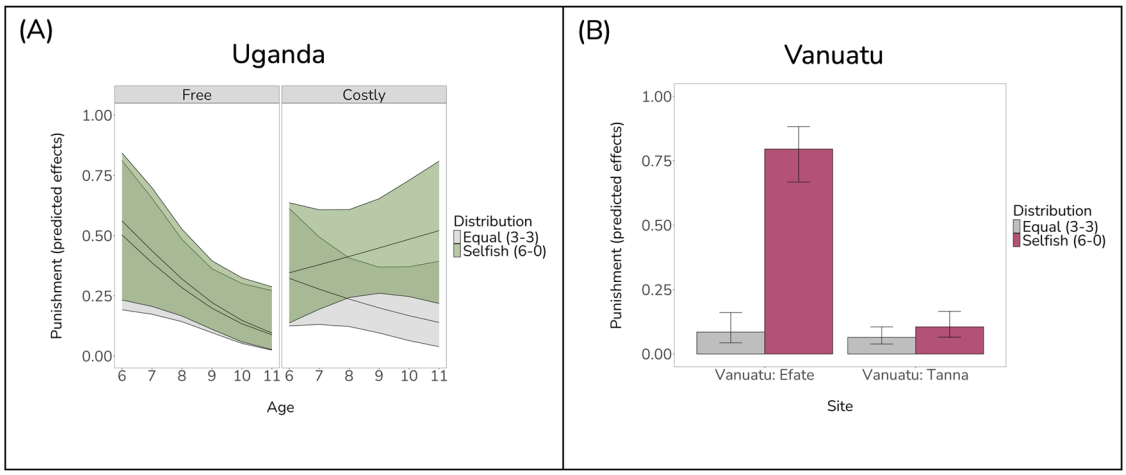

Fig. 7 | Predicted effects from models exploring punishment in Uganda and Vanuatu. A Predicted effects (probabilities) from a model predicting punishment in Ugandan children (*N* = 88). Model predicted punishment as a function of the interaction between distribution, age, and condition and additionally included gender and participant identity, which was fit as random intercepts. **B** Predicted effects (probabilities) from a model predicting punishment in Ni-Vanuatu children (*N* = 130). Model predicted punishment as a function of the interaction between distribution and site (Efate (*N* = 42) vs. Tanna (*N* = 88)). The model also included gender, condition, and participant identity, which was fit as random intercepts. Ribbons and errors bars represent 95% confidence intervals.

Children's endorsement of norms governing sharing and intervention

We next analyzed children's responses to our hypothetical scenarios assessing endorsement of norms governing resource distributions and intervention by third parties (Fig. 8). Children overwhelmingly endorsed equal sharing when asked to demonstrate the "right" way to divide six resources (probability set at 0.14 because equality was one of seven possible distributions: $\chi^2_1 = 3039.3$, $p < 0.001$, estimate = 0.94, CI = 0.92–0.96). Similarly, children overwhelmingly said that a selfish division of resources was not ok (probability = 0.5: $\chi^2_1 = 414.88$, $p < 0.001$, estimate = 0.96, CI = 0.93–0.97). See SOM for children's other choices by site. Finally, when asked who did the right thing, the vast majority of children chose a witness who intervened over a witness who did not intervene against a transgression (probability = 0.5: $\chi^2_1 = 674.14$, $p < 0.001$, estimate = 0.96, CI = 0.94–0.97). These findings suggest that, in general, children endorse the norm of fairness as equality and support intervention as an action that should be taken in the face of a transgression. However, we did see differences in the degree of endorsement for two of the three scenario measures: we found population differences in the proportion of children endorsing equal sharing ($\chi^2_4 = 57.85$, $p < 0.001$) and an intervening bystander ($\chi^2_4 = 22.08$, $p < 0.001$), but not in the proportion of children who endorsed unfair sharing as 'not ok' ($\chi^2_4 = 3.33$, $p = 0.5$). On our equal sharing measure, Bonferroni-corrected post-hoc tests revealed several differences: endorsement was higher in Canada than India ($\chi^2_1 = 36.8$, $p < 0.001$, estimate = 0.23;

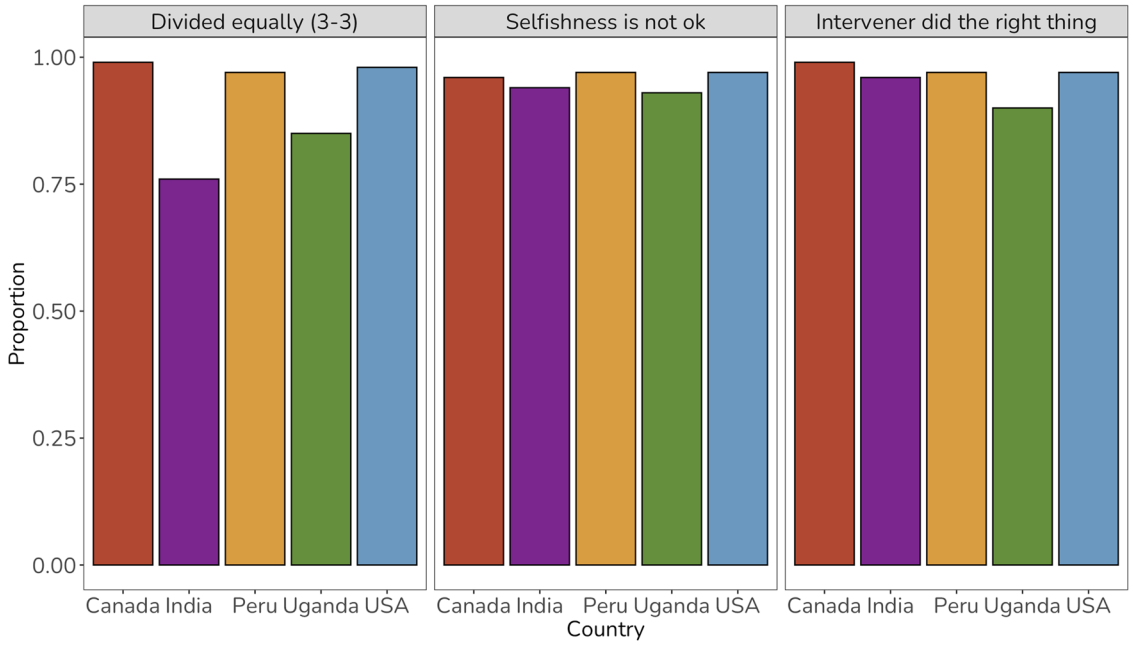

Fig. 8 | Children's endorsement of fair sharing (left), that unfairness is not ok (middle), and that the intervener (rather than non-intervening bystander) did the right thing (right).

CI = 0.09–0.37), higher in Canada than Uganda ($\chi^2_1 = 22.39$, $p < 0.001$, estimate = 0.14; CI = 0.06–0.22), higher in Peru than India ($\chi^2_1 = 34.65$, $p < 0.001$, estimate = 0.24; CI = 0.1–0.38), higher in the USA than India ($\chi^2_1 = 11.03$, $p = 0.009$, estimate = 0.23; CI = 0.08–0.38), and higher in Peru than Uganda ($\chi^2_1 = 12.13$, $p = 0.005$, estimate = 0.12, CI = 0.04–0.21). On our intervener measure, Bonferroni-corrected post-hoc tests revealed that endorsement was higher in Canada than Uganda ($\chi^2_1 = 13.89$, $p = 0.002$, estimate = 0.087, CI = 0.04–0.14). Despite these differences, however, the overall picture is that children across sites generally endorsed fairness and intervention.

Discussion

In this cross-cultural, developmental study, we tested three hypotheses about the development of third-party punishment of unfair sharing in children across diverse societal contexts. We found that children across sites were more likely to punish selfish distributions than equal distributions (Hypothesis 1), pointing to a widespread normative concern for fairness as indexed by their willingness to prevent unfair sharing amongst peers, even at personal cost. Second, we found some support for the prediction that punishment of unfair behavior increases with age (Hypothesis 2). The punishment behavior of children tested in Canada, India, Peru, and the USA conformed to this prediction, but we saw different patterns of punishment with age in Uganda and Vanuatu. These data, supported by our exploratory cross-site analyses, provide evidence for cross-societal variation in the developmental trajectory of third-party punishment. Third, we found some support for the prediction that children would punish more in the Free condition than in the Costly condition (Hypothesis 3). Punishment in Canada and the USA was influenced by condition in the predicted direction, yet we did not see condition differences in India, Peru, or Vanuatu.

In Uganda, we saw an unexpected interaction between condition and distribution such that children were more likely to punish selfishness when doing so was costly than when it was free. This was qualified by age, such that younger children in Uganda punished both selfish and equal distributions regardless of cost, while older children began to preferentially punish selfish distributions and were especially likely to punish when doing so was costly. In Vanuatu, we saw a pronounced difference between children tested in the rural site, Tanna, and the urban site, Efate. Specifically, children in Tanna were overall unlikely to punish, whereas children in Efate punished selfish behaviors at high rates.

Additional measures using hypothetical scenarios showed that children across all sites generally endorsed fairness and intervention. However, children in India and Uganda were less likely to demonstrate that an equal division was the "right" way to divide 6 resources. While we find these data informative in that they shed light on children's normative reasoning in contexts relating to fairness and intervention, these questions were asked following the third-party punishment task. Given this, it is reasonable to expect that children's exposure to repeated selfish (6-0) and equal (3-3) divisions may have influenced their responses to our hypothetical scenario questions. This design decision was made because we privileged the behavioral data over the hypothetical data in this study. However, we view the scenario responses as pointing to exciting directions for future work, particularly work exploring the space between what children actually do and what they think ought to be done—the so-called knowledge-behavior gap[28].

One clear finding from this study is that children across all six sites were more likely to punish selfishly unfair sharing than equal sharing. Moreover, punishment of unfair sharing emerged around middle childhood across sites (Figs. 4 and 6; see Fig S11 for site-by-age plot). These findings suggest that the developmental trajectory of third-party punishment of unfairness shows a striking degree of similarity across diverse societal settings. Another key finding from this work is that children's punishment of unfairness occurred even when punishment was personally costly. Indeed, costs only influenced punishment in Canada and the USA, suggesting that children generally punish regardless of costs. In Canada and the USA, costs mattered, but even in the costly condition punishment was frequent (Fig. 5). This finding suggests that children across diverse cultural contexts respond negatively to unfair treatment, even when they are uninvolved bystanders rather than the victims of unfairness. More broadly, these findings are relevant to discussions of the third-party punishment of unfairness as a signature of a normative stance on fairness: Children not only act to prevent unfair treatment when they are affected themselves, but also apply equality norms to sharing behaviors among third parties.

These findings offer an important developmental extension of prior cross-cultural work documenting that adults across the world engage in third-party norm enforcement, although to varying degrees[29,30]. Specifically, we show that across diverse societies, children as young as six years

spontaneously intervene against unfair sharing as indexed by selfishness. While we discover some variation in patterns of emergence, third-party punishment is present in middle childhood across all sites tested here. We therefore conclude that these behaviors are characteristic of children and not just adults, with children punishing unfair sharing among peers before they assume roles of authority or are tasked with socializing others as parents or teachers.

The overall conclusion that third-party punishment is present in middle childhood across societies aligns with other work that has provided cross-cultural data on children's third-party punishment[22]. House and colleagues[22] showed that when prompted to make punishment decisions across three conditions (punish antisocial, punish prosocial, punish either), children punished selfish sharing. Our data echo these results. Replications in the area of cross-cultural, developmental work are rare due to the challenges associated with collecting data with children across multiple sites. We thus see this conceptual replication as a major strength of the current study. Additionally, our findings add to those of House and colleagues in three main ways. First, we tested children at six new sites, providing an important expansion of populations represented in this area of research. Taken together with their results, we now have evidence for third-party punishment in children across 12 different sites. Second, we show that children engage in third-party punishment spontaneously, without any communication or prompt about what norm should or should not be followed. Third, we directly manipulated the costs of punishment, showing that while costs influenced punishment in two of our sites, children were generally willing to punish regardless of whether or not it was costly.

Our data also inform our understanding of the cultural distribution of fairness in childhood. It is interesting that we saw third-party punishment of unfair offers across sites while other work has shown that fairness behavior itself varies in children across countries. Specifically, Blake et al.[18] found that children's rejections of unfair distributions that placed them at a disadvantage relative to a peer (so-called disadvantageous inequity aversion) showed relatively less cross-site variation than children's rejections of unfair distributions that placed them at an advantage relative to a peer (so-called advantageous inequity aversion). Why would we see more cross-cultural variation in children's fairness behavior—advantageous inequity aversion specifically—than third-party punishment? While we cannot answer this question directly, one possibility is that occupying the role of a third party encourages children to take a principled stance on the behavior they are witnessing. This stance can be harder to take when one is directly involved in an interaction. Indeed, our norms data align with the idea that children are taking a principled stance: across both distribution scenarios, children demonstrated their stance that equality (3-3) is the "right" way to divide 6 resources and that unequal divisions are "not ok." Given this general pattern of endorsement and children's punishment of unequal (selfish) sharing, we are seeing alignment between their understanding of fairness norms and their punishment behavior. In other studies of fairness—particularly ones where the child must incur a higher cost in order to be fair—there is often more of a gap in their knowledge and behavior[28]. Thus, third-party punishment may be a particularly good marker of a strong normative sense of fairness because it involves less of a trade-off between the self and others. Consistent with this, it is important to note that third-party punishment in the USA develops earlier than advantageous inequity aversion, again potentially because the former is less costly to the self than the latter[13].

In Uganda, we found that older children were particularly likely to punish unfairness when doing so was costly, a finding that opposed our prediction that punishment would be more common in the Free condition. Why did Ugandan children punish in this way? One possibility that emerged in our discussions about this finding is that costs legitimize involvement in the interaction. That is, there may be norms governing involvement in others' conflicts in Uganda that preclude intervention when there is nothing at stake for the intervener. Our costly manipulation may have addressed this by allowing punishers to 'put their money where their mouths are', legitimizing their involvement in the intervention. Of course, at this stage, this explanation is purely speculative and future work in this community should seek to capture more detailed information about the norms that govern both intervention and fairness norms in both children and adults.

In Vanuatu, we saw a striking difference in punishment behavior between children living in Tanna—a rural site—and children living in Efate—an urban site. Namely, children in the urban site were more likely to punish than children in the rural site. This finding is in alignment with work on third-party punishment in adults showing that punishment scales positively with community size and market integration[29]. The central argument is that the enforcement of norms is crucial in larger communities where other mechanisms for sustaining cooperation (e.g., reputation) will be less successful. That this difference is potentially reflected in the punishment behavior of children suggests that social and ecological factors may shape this behavior from its emergence in development. Similar findings of stark contrasts between urban and rural ni-Vanuatu children have been shown in a battery of theory of mind tasks, warranting further investigation between these two contexts[31]. Future work should seek to better understand this difference, drawing on ethnographic work with ni-Vanuatu children (see SOM for a more in-depth discussion of this finding).

Limitations

There are a number of limitations that warrant discussion. First, experimental paradigms designed to measure third-party intervention[1]—while useful in that they can be deployed in a standardized way across diverse cultural contexts[29]—do not map 1:1 to instances of third-party punishment in people's daily lives. Future work should integrate findings from tasks like ours and observational data on children's third-party intervention behavior in daily life[32]. A second limitation involves our operationalization of fairness as equality. While equality is an important fairness principle, fairness is, of course, much richer than equality alone. Further work should extend the range of fairness principles and test situations that require the integration of information about contextual cues such as deservedness (e.g., merit or need), pre-existing inequalities (e.g., structural factors), and fairness concerns extending beyond resource sharing (e.g., procedural fairness). All of these features likely feed into children's understanding of fairness and, by extension, their willingness to intervene against fairness norm violations. Even within the operationalization of fairness as equality, we tested an extreme version of unfairness (fully selfish sharing). Future work could titrate out the degree of inequality to explore whether children's punishment is calibrated to the level of unfairness. Finally, while our study focused on third-party punishment, there are other ways in which children could intervene against unfairness. Specifically, verbal admonishments, protest, or compensating the victim of unfair treatment are other ways third parties can intervene to express fairness concerns. Therefore, future work on interventions against unfairness should explore a broader space of interventions to understand the extent to which children value punishment when other choices are available.

Conclusion

In conclusion, we found that children across six diverse societal contexts—Canda, India, Peru, Uganda, USA, and Vanuatu—punished unfair sharing. They did so as third parties—i.e., when they were not direct victims of unfair sharing but rather witnessed unfair sharing between peers. In some cases, children were willing to intervene at personal cost, sacrificing their own rewards to punish selfishness in others. These findings point to a strong normative stance against unfairness. More broadly, our results suggest that the third-party punishment of unfairness is widespread and seen relatively early in development, emerging prior to extensive experience with locally-specific justice systems or extensive exposure to theories of justice.

Data availability

Data from all sites are available at: https://osf.io/6nm2s/?view_only=83989fdb987d428dabf796df541b7829.

Code availability

R Code is available at: https://osf.io/6nm2s/?view_only=83989fdb987d428dabf796df541b7829.

Received: 24 July 2024; Accepted: 24 February 2025;
Published online: 17 March 2025

References

1. Fehr, E. & Fischbacher, U. Third-party punishment and social norms. *Evolution Hum. Behav.* **25**, 63–87 (2004).
2. Henrich, J. et al. Costly punishment across human societies. *Sci. (N. Y., N. Y.)* **312**, 1767–1770 (2006).
3. Marshall, J. & McAuliffe, K. Children as assessors and agents of third-party punishment. *Nat. Rev. Psychol.* **1**, 334–344 (2022).
4. McAuliffe, K., Jordan, J. J. & Warneken, F. Costly third-party punishment in young children. *Cognition* **134**, 1–10 (2015).
5. Lee, Y. & Warneken, F. Does third-party punishment in children aim at equality? *Developmental Psychol.* **58**, 866–873 (2022).
6. McAuliffe, K. & Dunham, Y. Children favor punishment over restoration. *Developmental Sci.* **24**, e13093 (2021).
7. Lee, Y., He, S. & Warneken, F. Children's third-party punishment does not change depending on the prospect of future interaction. *J. Exp. Psychol.: Gen.* https://doi.org/10.1037/xge0001515 (2023).
8. Martin, J. W., Martin, S. & McAuliffe, K. Third-party punishment promotes fairness in children. *Developmental Psychol.* **57**, 927–939 (2021).
9. Nielsen, M., Haun, D., Kärtner, J. & Legare, C. H. The persistent sampling bias in developmental psychology: a call to action. *J. Exp. Child Psychol.* **162**, 31–38 (2017).
10. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–135 (2010).
11. Gurven, M. To give and to give not: The behavioral ecology of human food transfers. *Behav. Brain Sci.* **27**, 543–559 (2004).
12. Henrich, J. et al. 'Economic man' in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* **28**, 795–815 (2005).
13. McAuliffe, K., Blake, P. R., Steinbeis, N. & Warneken, F. The developmental foundations of human fairness. *Nat. Hum. Behav.* **1**, 0042 (2017).
14. House, B. R. et al. Ontogeny of prosocial behavior across diverse societies. *Proc. Natl Acad. Sci. USA* **110**, 14586–14591 (2013).
15. House, B. R. et al. Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nat. Hum. Behav.* **4**, 36–44 (2020).
16. House, B. R. & Tomasello, M. Modeling social norms increasingly influences costly sharing in middle childhood. *J. Exp. Child Psychol.* **171**, 84–98 (2018).
17. McAuliffe, K., Raihani, N. J. & Dunham, Y. Children are sensitive to norms of giving. *Cognition* **167**, 151–159 (2017).
18. Blake, P. R. et al. The ontogeny of fairness in seven societies. *Nature* **528**, 258–261 (2015).
19. Huppert, E. et al. The development of children's preferences for equality and equity across 13 individualistic and collectivist cultures. *Dev. Sci.* **22**, 1–15 (2019).
20. Schäfer, M., Haun, D. B. M. & Tomasello, M. Fair Is Not Fair Everywhere. *Psychological Sci.* **26**, 1252–1260 (2015).
21. Kanngiesser, P. et al. Children across societies enforce conventional norms but in culturally variable ways. *Proc. Natl. Acad. Sci. USA* **119**, e2112521118 (2022).
22. House, B. R. et al. Social norms and cultural diversity in the development of third-party punishment. *Proc. R. Soc. B.* **287**, 20192794 (2020).
23. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* **67**, 1–48 (2015).
24. Forstmeier, W. & Schielzeth, H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav. Ecol. Sociobiol.* **65**, 47–55 (2011).
25. Lüdecke, D. ggeffects: tidy data frames of marginal effects from regression models. *JOSS* **3**, 772 (2018).
26. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer International Publishing: Imprint: Springer, Cham, 2016). https://doi.org/10.1007/978-3-319-24277-4.
27. Agresti, A. & Coull, B. A. Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *Am. Statistician* **52**, 119–126 (1998).
28. Blake, P. R., McAuliffe, K. & Warneken, F. The developmental origins of fairness: The knowledge-behavior gap. *Trends Cogn. Sci.* **18**, 559–561 (2014).
29. Henrich, J. et al. Markets, religion, community size, and the evolution of fairness and punishment. *Science* **327**, 1480–1484 (2010).
30. Molho, C., De Petrillo, F., Garfield, Z. H. & Slewe, S. Cross-societal variation in norm enforcement systems. *Philos. Trans. R. Soc. B* **379**, 20230034 (2024).
31. Dixson, H. G. W., Komugabe-Dixson, A. F., Dixson, B. J. & Low, J. Scaling theory of mind in a small-scale society: a case study from vanuatu. *Child Dev.* **89**, 2157–2175 (2018).
32. Dixson, H. G. W. Generosity, Third Party Punishment & Theory of Mind Among Ni-Vanuatu Children.

Acknowledgements

We are grateful to the communities that allowed us to conduct this research. We are also grateful to the local research assistants and assistants based at our home institutions for their work supporting this research. Specifically, we relay our deep thanks to Natalie Benjamin, Michael Bogese, Hannah Bolotin, Jessie Doyle, Alissa Fiedler, Sindhu Madala, Oriella Mansur, Molly McDowell, Lillian Nsimiire, and Audrey Todd. We are grateful to Julia Marshall for comments on an earlier version of this paper. HD is grateful to the Vanuatu Kaljoral Senta and for the cooperation of ni-Vanuatu friends throughout Port Vila and Niprainitata, Tanna. AF thanks a grant from the Herchel Smith Program that supported data collection in Uganda. FW was supported by an NSF CAREER award. Finally, we are grateful to Ehri Ryu for providing statistical consultation during the revision process. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

K.M.A.: Study design, General project administration; Data processing and analysis, Interpretation, Writing—original draft, Funding acquisition; S.B.: Project administration in Peru, Interpretation; T.C.: Project administration in Canada, India, and Peru, Interpretation, Writing—review & editing; J.C., Project administration in Canada and India, Interpretation, Writing—review & editing; H.G.W.D.: Project administration in Vanuatu, Interpretation, Writing—review & editing; Y.D.: Interpretation, Writing—review & editing, Funding acquisition; A.F.: Project administration in Uganda and the USA, Interpretation, Writing—review & editing, Funding acquisition; E.O.: Project administration in Uganda, Interpretation; S.R.: Data processing, Interpretation, Writing—review & editing; P.T.: Project administration in Uganda; F.W.: Study design, General project administration, Interpretation, Writing—review & editing, Funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44271-025-00220-x.

Correspondence and requests for materials should be addressed to Katherine McAuliffe.

Peer review information *Communications Psychology* thanks Patricia Kanngiesser and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Jennifer Bellingtier. A peer review file is available.