The influence of age and experience of (un)fairness on third-party punishment in children

Young-eun Lee [1,2] and Felix Warneken [1]

[1]Department of Psychology, University of Michigan

East Hall, 530 Church St., Ann Arbor, MI, 48109, USA

[2]Department of Psychology, Columbia University

1190 Amsterdam Ave, New York, NY, 10027, USA

Conflict of Interest Statement

There is no conflict of interest.

Data Sharing and Data Accessibility

All data are available through the Open Science Framework: https://osf.io/upexj/?view_only=e3d5dea53d3c48e9b071580d82c2668a.

Acknowledgements

This work was funded by a National Science Foundation CAREER Grant [#1760238] to FW. We thank research assistants for their help with data collection. We also thank Michael Clark at CSCAR

at the University of Michigan for statistical advice, Sebastian Grüneisen for helpful comments on the manuscript, and Amy Nowack for proofreading the manuscript.

Author Note

Correspondence concerning this article should be addressed to Young-eun Lee at yl4898@columbia.edu.

Abstract

Third-party punishment is an important mechanism to enforce norm-following. However, the underlying process that explains the development of third-party punishment is understood poorly. Here we examine to what extent age-effects and contemporaneous experiences of receiving unfair offers influence third-party punishment. In two studies, a total of $N = 280$ 5- to 9-year-olds participated in a computer-based task in which they received either fair or unfair offers from another peer. In the subsequent test phase, children could punish unfair offers as an unaffected third-party. We found that with age, children become increasingly systematic in their decisions to punish unfair allocations. However, there was no strong evidence that an immediate experience of (un)fairness influenced children's punishment. Together, our results suggest that children develop a sophisticated application of fairness norms with age that is not easily swayed by their immediate experience of being treated unfairly.

Keywords: fairness, sharing, second-party experience, third-party punishment, development

The influence of age and experience of (un)fairness on third-party punishment in children

Fairness norms such as equality or merit guide people through how to divide resources (Deutsch, 1975). Despite these norms, sometimes people divide resources selfishly or take advantage of others' cooperativeness by free-riding (Fischbacher et al., 2001). How do people deter these uncooperative behaviors? Studies with adults suggest that punishment can discourage selfish behaviors and maintain cooperation (Boyd et al., 2003). Specifically, adults are often willing to pay a cost to punish those who engage in unfair resource allocations even when they are an unaffected third-party (Henrich et al., 2006; Krasnow et al., 2016). This phenomenon is called *costly third-party punishment* and is well-established in adults.

Third-party punishment often has been claimed to reflect one's concern for cooperative social norms because a third-party pays a personal cost to punish the perpetrator with no immediate benefits (Fehr & Fischbacher, 2004, but see Raihani & Bshary, 2019). In fact, it has been regarded as a test case for the emergence of fairness concerns both across phylogeny and ontogeny (McAuliffe et al., 2017). For these reasons, developmental psychologists have started to investigate the developmental origins of third-party punishment. Several studies have shown that 2- to 3-year-olds punish against certain forms of antisocial behavior, such as hindering of another agent's instrumental goal (Hamlin et al., 2011) or damage to one's property (Vaish et al., 2011; Yudkin et al., 2020). However, they did not assess how children react to violations of fairness norms, such as equal sharing.

Children's own sharing undergoes major changes over early childhood, moving toward more equitable sharing across different situations (Elenbaas et al., 2020; McAuliffe et al., 2017). For instance, young children tend to keep more for themselves and share resources equally not until around 7-8 years of age, displaying a discrepancy between their understanding of an equality norm and behavioral adherence to the norm (Smith et al., 2013 with US child samples). Similarly, 4- to 7-year-olds in the US avoid receiving fewer resources than a peer, but accept receiving more resources than the peer – whereas starting around age 7, children reject receiving more than someone else (Blake & McAuliffe, 2011). These studies suggest that children start applying equality norms in a

self-focused way and move toward more generalized egalitarian behavior by around 7 years. Interestingly, this is around the same time when children from these populations show a sophisticated enactment of third-party punishment (Gummerum & Chu, 2014; House et al., 2020; McAuliffe et al., 2015). For example, 6-year-olds in the US, but not 5-year-olds, show reliable costly third-party punishment, punishing more often in unequal allocations than in equal allocations at a personal cost (McAuliffe et al., 2015). Hence, even though they were uninvolved third parties, 6-year-olds reliably inflicted a cost on the selfish divider at their own cost whereas, 5-year-olds showed less reliable rates of punishment. For this reason, third-party punishment marks a developmental milestone in a sense of fairness in that children overcome their self-interest and apply the fairness norms even when their own interest is not at stake because the unfair behavior did not influence them directly (McAuliffe et al., 2017).

These results raise the question about the underlying process for this developmental trajectory. In our study, we focused on the potential role of children's experience on third party-punishment and tested three specific hypotheses that could explain its role in development. One hypothesis is that experiences of unfairness as a second-party victim promote third-party punishment. Specifically, by being exposed to unfair experience more often, children are more adept at imagining what the victim of unfairness is experiencing and feel compelled to set things straight after they put themselves in the victim's shoes. This hypothesis would be consistent with the notion that we rely on *simulation* to predict others' minds and behaviors (Gordon, 1986; Harris, 1992). That is, one's own mind can be used as a basis for understanding other people (e.g., "What would I do if I were in the same situation?"). The notion here is that an understanding of another person's psychological world starts with one's own experience. Therefore, having a similar experience in the past should be able to enrich the simulation process, leading to a better understanding of others in the same circumstance. Applied to the fairness context, children's personal experience of unfairness could increase their sensitivity to a third-party's unfair sharing by allowing them to apply their experience and to simulate others' minds (e.g., "What would I want to do if I were treated unfairly?"; Bloom, 2013).

Several recent studies indicate that children's personal experience might influence their sense of fairness. As reviewed above, resource allocations that disadvantage a child are rejected more often and earlier in development than allocations that put the child at an advantage in a second-party context. Furthermore, Bernhard et al. (2020) found that 5- and 7-year-olds in the US punished unfair individuals more often when they were directly affected by unfairness (second-party punishment) than when they were third parties (third-party punishment). These studies, therefore, suggest that the personal experience of receiving unfair treatment is perhaps a crucial aspect of fairness development and the developmental shift toward third-party punishment consists of becoming able to apply this first-hand experience to others. Building upon this idea, we hypothesized that children who experienced unfair allocations from others in the past would be more likely to intervene against unfairness as a third-party compared to those who did not experience unfair allocations. Further, this tendency would increase with age as children become better at simulating others' minds.

Another plausible hypothesis is that unfair experience could impact costly third-party punishment, but in the opposite direction. In other words, children who experienced unfair allocations from others would be *less* likely to intervene in a third-party context because children who experienced unfair resource sharing would possess fewer resources compared to those who experienced fair sharing. Therefore, the sense of resource scarcity and/or perceived unfairness might discourage children's costly third-party punishment, which requires a child's personal cost.

This possibility is consistent with findings from the literature that 5- to 6-year-old children tend to focus on having more resources than their peers, which reflects their competitive nature and less concern for fairness in young children (Sheskin et al., 2014). Furthermore, 4-year-olds donate less stickers to another person in need when there is high inequality in earnings between the self and peers (versus when there is low inequality; Kirkland et al., 2020). These findings suggest that high inequality between the self and a peer could decrease children's willingness to engage in altruistic behaviors. Thus, when applied to the context of costly third-party punishment, we hypothesized that children (especially, young children who tend to show less concern for fairness) would show less costly third-party punishment when they experienced unfair sharing than when they did not

experience unfair sharing potentially due to the sense of resource scarcity to enact punishment and/or because of perceived inequality between themselves and the peer.

The two hypotheses illustrated above assume that children's immediate experience of being treated unfairly impacts their third-party punishment. A third hypothesis would posit that it is children's general development of a more generalized sense of fairness that drives third-party punishment, rather than children's experience in a first-party context. Therefore, the idea here is that in contrast to the first hypothesis, children do not necessarily rely on simulation by extending their own experiences. Instead, with age, they acquire and apply general fairness principles in a more unbiased, agent-neutral way, in which an interest or perspective of an agent (including children themselves) is not favored over that of another agent. This hypothesis would predict the absence of an effect of unfair experience, while predicting an effect of age on third-party punishment. This hypothesis would align with the theoretical claim that as children grow older, they employ a more principle-based and perhaps impartial fairness concept (Shaw, 2013).

Several pieces of evidence show that this impartiality hypothesis is plausible. For example, around age 8, children from the US give up their own resources to avoid getting more than others, showing an aversion to inequality advantageous to themselves (Blake & McAuliffe, 2011; Shaw & Olson, 2012). Further, with age, US children rectify inequality or punish selfishness not only when their ingroup members are disadvantaged but also when outgroup members are disadvantaged (Elenbaas et al., 2016; Jordan et al., 2014). Together, these studies suggest that with age, children apply fairness norms to the self and ingroup members as well as others and outgroup members, applying an equality norm in a more impartial way. This pattern of findings shows that children develop general principles that dictate their sense of distributional justice. Under this approach, the prediction would be that regardless of their immediate personal experience, as children grow older, they develop general principles that guide how resources should be distributed and what is the right thing to do in the face of unfairness.

To our knowledge, no research has examined whether experiencing fair versus unfair resource sharing as a second party influences subsequent third-party punishment. The current study was designed to address these three hypotheses on the potential role of experience on third-party punishment.

Current Study

The current study tested the impact of children's previous sharing experiences on their subsequent willingness to punish unfairness at a personal cost. We focused on fairness in terms of distributional equality as a norm that is particularly salient to young children. To manipulate children's sharing experiences with other people, we developed a novel computer-based task, in which child participants have live interactions with (computer-programmed) peers and watch how they divide coins that can be exchanged for prizes afterward. Children were assigned to one of three conditions: (a) unfair experience condition, in which they always received unequal offers (0 out of 6 coins) from another player, (b) fair experience condition, in which they always received equal offers (3 out of 3 coins) from another player, and (c) no experience condition, in which they did not play the role as a recipient at all. In the subsequent test phase, children played a third-party punishment game, in which children as a third-party could observe how one player divides coins with another player and decide whether to punish the offer or not. We presented 6:0 allocation as an unequal offer (as opposed to 5:1 or 4:2) with an expectation that receiving none would provide a stronger experience of unfair sharing in children. Also, children's third-party punishment against 6:0 allocations has been well-documented in previous studies (Jordan et al., 2014; McAuliffe et al., 2015).

Our study focused on testing the three main hypotheses articulated above. The simulation hypothesis would predict that children in the unfair experience condition will be more likely to later engage in third-party punishment against unfair offers than children in the no experience or fair experience conditions. The resource scarcity hypothesis would predict that children in the unfair experience condition will be less likely to engage in third-party punishment against unfair offers compared to those in the other conditions. The other main hypothesis purporting a developmental

change toward more principled, impartial fairness would predict that with age, children become more reliable in punishing unfair over fair offers, regardless of their own immediate prior experience. We conducted two experiments, with Experiment 1 focused on testing the effects of age and (un)fair experience on third-party punishment, and Experiment 2 replicating the main findings.

Experiment 1

Method

Participants. Our final sample were $N = 120$ 5- to 9-year-old children ($M = 89.32$ months, range = 60 - 119 months, $n = 24$ in each age group, 40 per condition, 60 male, and 60 female). Children were recruited and tested at museums or public parks in a city in the Midwest of the US. After parental consent, children were brought to a table with the study apparatus while the parents watched from a few steps away, instructed not to intervene. Demographic information such as race, education and income could not be obtained as per the rules of the museum.

Ten additional children were excluded because they failed to answer correctly in at least one of the comprehension check questions about the consequences and costs involved in each button (4), there was interference from a bystander (4), the child was not an English-speaker (1) or the parent provided no birth date (1).

Experimental design and procedure. A female experimenter introduced the computer game referred to as the "coin game". We first established that children could collect virtual coins that they could later exchange for prizes. Specifically, during a *prize introduction*, children saw an image of the prizes and an image of coins needed to purchase them later. Children learned that the more coins they have during the coin game, the more and the better prizes they could choose afterwards. The prize introduction concluded with comprehension check questions in which the experimenter asked children to identify prize(s) they could have with few versus many coins. Children were correct in these comprehension check questions (98%) confirming their understanding of the exchange value of coins

in the computer game. The experimenter provided corrective feedback to a minority of children who answered incorrectly.

In the subsequent *practice phase*, the experimenter introduced the two other players in the game by saying that they were children of the same age and gender as the participants but are at another location (at another museum, for those who tested at a museum or at another park, for those who tested at a park). In reality, the other players were computer-programmed. The experimenter then introduced the role of the *divider* and the *recipient*. The divider can decide how to divide 6 coins between the self and the recipient. The divider can make one of two offers: equal offer (3 for the self, 3 for the recipient) or unequal offer (6 for the self, 0 for the recipient). The recipient was a passive player who can only accept the offer made by the divider.

After introducing the roles, children saw the offers made by the divider enacted on the screen and practiced their role as a third-party punisher. The experimenter told children that after the divider makes an offer to the recipient, they can press either the green or the red button. If they push the green button (acceptance), the 6 coins will go into each player's basket just the way the divider allocated the coins (e.g., if the divider splits it up 3:3, each player's basket receives 3 coins), and the child's own coin goes back into their own basket. That is, acceptances incurred no cost to the child.

In contrast, if children push the red button (rejection), a vacuum will appear at the top of the screen and the 6 coins will be sucked up and disappear into the vacuum. To enact the rejection, children first had to pay their one coin into the vacuum. Therefore, pressing the red button serves as a costly third-party punishment.

There were four practice trials in total. Children practiced four possible outcomes of each button (accept vs. reject) in each offer type (equal vs. unequal). The experimenter asked comprehension check questions about the consequence of each button and whether each button required the payment of the child participant's coin or not. All participants included in the data analysis answered the comprehension check questions correctly. When the child answered incorrectly, the experimenter provided corrective feedback and asked the question(s) again in the next practice

trial. Across experiments, we excluded 6 out of 299 children (2%) who still answered incorrectly by the end of the practice phase.

After the four practice trials, to make children believed that the other players were real, the experimenter pretended to make a phone call to the other players on the speakerphone and check if they were ready to play the game. In reality, another experimenter (confederate) answered the phone call, and there were no other players. Children received 20 coins as their initial endowment (coins that dropped into their basket on the screen).

In the subsequent *experience phase*, children were assigned to one of three conditions (between-participant): *fair experience* condition, *unfair experience* condition, or *no experience* condition (see Figure 1). Children played the role of a passive recipient for four consecutive rounds during the second-party game with another child as the divider. In the fair experience condition, the divider always kept 3 coins for the self and gave 3 coins to the child participant. In contrast, in the unfair experience condition, the divider always kept 6 coins for the self and gave 0 coins to the child. Those in the no experience condition (baseline) were unaware of the existence of the second-party game and did not play the second-party game at all. Procedures in the no experience condition were identical to the two other experience conditions except that they skipped the experience phase before playing the test phase (i.e., third-party game). Because there was no second-party game in the no experience condition, the amount of time that children spent in a testing session was inevitably shorter compared to the two other conditions.

After the first and the fourth round of the second-party game, the experimenter asked children to recall the number of coins they received from the divider during the second-party game. Most children (99%) reported the number of coins they received correctly. Furthermore, at the end of the second-party game, the experimenter told children that a new divider, not the same divider from the second-party game, will play a third-party game with them. Subsequent comprehension checks confirmed that children correctly identified players from the second-party game and those from the

third-party game (96%), showing that children were aware of who had played each game. The experimenter provided corrective feedback to a minority of children who answered incorrectly.

During the subsequent *test phase*, children played 6 rounds of the third-party game where the child was in the role of the third-party and two other gender-matched children were the divider and the recipient identified with different names and different colors of stick figures (see Figure 1). Moreover, the divider and recipient in the third-party game differed from those in the practice trials or those in the second-party game to prevent potential carryover or retaliation towards the same divider. Within the test trials, the role of each player remained the same.

There were 3 rounds with equal offers (3:3) and 3 rounds with unequal offers (6:0) presented in a pseudo-random order with a restriction that no more than two identical allocation types were presented consecutively. Importantly, children were unaware of the total number of rounds in the third-party game to prevent them from calculating the number of coins they currently had and the number of remaining rounds, which could influence their decision whether to spend a coin in a given round. Our dependent measure was children's rate of rejection, with rejection (red button) coded as 1 and acceptance (green button) as 0.

After the test phase, the experimenter left, and a secondary experimenter asked children whether they thought the players in the game are real or pretend players. Across two experiments, 91% of children (106 out of 120 in Experiment 1, 149 out of 160 in Experiment 2) reported that the players were real. In both experiments, we counterbalanced the trial order of offer types during the third-party game, player's identity, left/right position of buttons, the order of practice trials, and the order of comprehension check questions. For an exploratory purpose, the experimenter asked children how close they feel to the recipient in the third-party game after the completion of the study (see Supplemental Material for more details about this task).

Data coding and statistical analyses. Children's responses were automatically recorded by the computer game program, GameMaker Studio (https://www.yoyogames.com/gamemaker), and later entered into a spreadsheet by independent coders. All statistical analyses were conducted with R

statistical software (R version 3.5.2; R Core Team, 2018). In both experiments, we analyzed children's rejection rate with Generalized Linear Mixed Models (GLMM) using the package *glmmTMB* (Brooks et al., 2017) and children's expectations about offers (Experiment 2) with linear models.

Our analysis procedure was as follows: (1) we examined a null model which included only participant ID in mixed models; (2) we created a full model which included our main predictors (e.g., age in months, condition, offer type) and all interactions among the predictors; (3) we compared the full model with the null model; (4) if the full model provided a significantly better fit to the data, we sequentially dropped single terms from the full model, testing whether their exclusion significantly reduced model fit and (5) stopped this process when the dropping of terms no longer reduced model fit significantly.

As a post-hoc analysis, we employed Bayesian statistics to provide more information about the robustness of our findings. Non-significant *p* values from traditional Null Hypothesis Significance Testing (NHST) could occur because the effect does not actually exist (i.e., true negative) or because there is lack of power to detect a true effect (i.e., false negative; Aczel et al., 2018), which cannot be distinguished in NHST. Due to this nature of NHST, Bayesian analysis is useful especially when the results are not significant and could provide more information on whether the non-significant effect is likely to be a true negative or false negative.

A Bayes factor (BF) quantifies the degree to which the data favor the null hypothesis over the alternative hypothesis, and vice versa (Aczel et al., 2018; Wagenmakers et al., 2016). Conventionally, a BF between 1 and 3 indicates anecdotal evidence, a BF greater than 3 suggests moderate evidence, and a BF greater than 10 provides strong evidence in favor of one hypothesis over the other (Jeffreys, 1961). BF_{10} refers to the Bayes factor in favor of the alternative hypothesis over the null hypothesis, whereas BF_{01}, which is the inverse number of BF_{10}, refers to the Bayes factor in favor of the null hypothesis over the alternative hypothesis.

We computed BFs by comparing a GLMM in which a predictor of interest was included with a GLMM in which the predictor was not included, using the package *brms* (Bürkner, 2017). As in main analyses, we included participant ID as a random intercept in these models. The population-level regression coefficients had a weakly informative Student's t distribution prior which was zero-centered with 3 degrees of freedom and a scale of 2.5 (Gelman et al., 2008; see pp. 6-8 of the Supplemental Material for exploring different prior distributions). All models were run with 10,000 iterations with the first half as burn-in. $\hat{R}$ was less than 1.01 for all parameters, suggesting convergence (Vehtari et al., 2021).

We pre-registered both experiments before data collection (https://aspredicted.org/blind.php?x=fy77e4 for Experiment 1 and https://aspredicted.org/blind.php?x=j33fi6 for Experiment 2). All data and protocols are available through the Open Science Framework: https://osf.io/upexj/?view_only=e3d5dea53d3c48e9b071580d82c2668a.

Results

Rejection rate. A full GLMM on children's punishment decision (0 = acceptance, 1 = rejection) with condition, offer type, age, and interactions among the predictors as fixed effects and participant ID as a random effect provided a significantly better fit to the data than the null model with only random intercepts (LRT, χ^2 (11) = 139.69, $p < .001$). There were no effects involving trial number, LRT, χ^2 (6) = 8.76, $p > .18$ (see pp. 2 of the Supplemental Material for the analysis from the first trials). Additionally, we found similar results in rejection rate even after excluding a minority of children ($n = 14$) who reported that the players were pretend players (see pp. 9 of the Supplemental Material for descriptive statistics of these children).

Our critical question was whether children's third-party punishment varied depending on their experiences of (un)fairness. Results revealed a significant main effect of condition, LRT, χ^2 (2) = 7.24, $p = .03$ (and no two- or three-way interactions involving condition; all ps > .28). Specifically, children in the unfair experience condition ($M = 0.23$, $SD = 0.42$) punished significantly *less often* than those

in the no experience condition (M = 0.33, SD = 0.47; b = 0.64, SE = 0.26, p = .01) or fair experience condition (M = 0.32, SD = 0.47; b = 0.55, SE = 0.26, p = .03; see Figure 2A). By contrast, children in the no experience and fair experience condition did not differ from each other in their punishment rate (b = 0.09, SE = 0.25, p > .72). These results suggest that the experience of unfairness decreased children's punishment, whereas the experience of fairness had no effect relative to baseline.

Another important question concerned whether children's third-party punishment changes with age. We found a significant interaction effect between age and offer type (LRT, χ^2 (1) = 18.69, p < .001). To unpack this interaction, we ran separate models for equal and unequal offers. The results indicated that with increasing age, children become less likely to punish equal offers (LRT, χ^2 (1) = 14.51, b = -0.72, SE = 0.20, p < .001; see Figure 2B). However, there was no such age-related change in children's punishment of unequal offers (LRT, χ^2 (1) = 1.98, b = 0.19, SE = 0.13, p > .15). That is, their punishment of unequal offers remained relatively high across our age-groups but that of equal offers declined with age. Inspection of confidence intervals revealed that from 69 months of age (5.75 years), the confidence interval of equal offers no longer overlapped with that of unequal offers.

In Experiment 1, to examine whether our results provide strong support for the effect of condition reported above, we computed Bayes factors (BF) for effects involving condition as a post-hoc analysis. This revealed moderate evidence in favor of an *absence* of the main effect of condition (BF_{01} = 4.47), suggesting that the data were about 4 times more likely to be observed under the hypothesis that children's prior experience has no effect than the hypothesis that prior experience has an effect on punishment. Furthermore, because we predicted that children would intervene more often in unequal offers than in equal offers after experiencing unfairness, we computed a BF of an interaction between condition and offer type, which provided evidence in support of an absence of the interaction effect (BF_{01} = 8.93). In contrast, we found very strong evidence for an interaction effect between age and offer type (BF_{10} = 805), indicating that the data were 805 times more likely to be observed under the hypothesis predicting an interaction effect than under the hypothesis that there is no such interaction (see pp. 4-5 of the Supplemental Material for more detailed results).

Taken together, these findings suggest that as they grow older, children become more selective about the target of their punishment, understanding better when to punish and when not to punish. Although children in the unfair experience condition were slightly less likely to show punishment, the Bayesian analysis indicated that this result should be interpreted with caution.

Discussion of Experiment 1

In Experiment 1, older children became increasingly more selective in their enactment of third-party punishment. Also, we found that costly third-party punishment emerges around age 6, providing converging evidence to prior research (McAuliffe et al., 2015). These results suggest that children develop a sophisticated understanding of fairness norms and their application in a third-party context over development.

One important question was whether children's experience as a recipient would influence their third-party punishment. We found that if anything, receiving a series of unequal offers slightly discouraged children's subsequent punishment. This result would be consistent with the hypothesis that experience of unfairness would decrease costly third-party punishment, which opposes the prediction of the simulation account that the experience of unfairness would promote third-party punishment. However, further analysis using Bayes factors revealed that our data were moderately in favor of the absence of a condition effect, suggesting that the significant main effect of condition found in Experiment 1 is questionable. Therefore, we conducted Experiment 2 as a replication study.

Moreover, in Experiment 2, we aimed to asses two possible causes of the decrease in punishment in the unfair experience condition. One possibility is that children in the unfair experience condition show reduced third-party punishment because they feel "coin-deprived". That is, even though children across three conditions receive 20 coins as their initial endowment, receiving 0 coins for four consecutive trials in the second-party game might make children hold on to the remaining coins, leading to a decrease in costly third-party punishment because of the cost.

The alternative possibility is that during the second-party experience phase, children form an expectation of how individuals treat each other in this game. That is, when children themselves are treated unfairly by their social partner, this mistreatment might change their expectations about how people should treat each other (e.g., "It seems fine to treat each other unfairly in this game as I was treated unfairly"), and thus children do not feel compelled to intervene when someone acted unfairly.

Experiment 2 was designed to differentiate between these two possibilities by introducing non-social conditions in which a computer allocates coins in the second-party game. The number of coins children received in these non-social conditions were matched with those from the fair experience and unfair experience conditions except that children received coins based on the decision made by a computer. Additionally, we assessed how children's expectation about a divider's offer changes after their own experience of receiving either equal or unequal offers.

Experiment 2

The purpose of Experiment 2 was (a) to replicate the findings in Experiment 1 and (b) to examine the reason for the decrease in punishment observed in Experiment 1. In Experiment 2, in addition to the fair experience and unfair experience conditions, there were two additional conditions: winner condition and loser condition. The number of coins children received in these conditions were identical to those they received in the fair experience condition and unfair experience condition, respectively, except that a non-social agent (computer) made the resource allocations.

There are at least three possible outcomes in Experiment 2. The first possible outcome is that children in the unfair experience condition punish less often than those in the loser condition. Such a result would suggest that children who received unfair treatment from a social agent assume that selfish allocations are the norm in this game and therefore see no reason to intervene against unfairness.

A second possibility is that children with relatively fewer coins (i.e., unfair experience and loser conditions) will show less third-party punishment than those with relatively more coins (i.e., fair

experience and winner conditions). If obtained, this finding would imply that children's third-party punishment is affected by the number of coins they have or the perception that they are coin-deprived.

A third possibility is that regardless of their prior experience, children are focused on the third-party allocation and become more selective in punishing unequal over equal sharing with age. Such a finding would show that the developmental effect of children becoming more selective in their punishment is robust against any contemporaneous experiences.

Method

Participants. Our final sample were $N = 160$ 5- to 9-year-old children ($M = 89.64$ months, range = 60 - 119 months, $n = 32$ in each age group, 40 per condition, 80 male and 80 female). A new sample of children were recruited and tested at museums or public parks in a city in the Midwest of the US. Demographic information such as race, education and income could not be obtained as per the rules of the museum.

Nine additional children were excluded because they failed to answer correctly in at least one of the comprehension check questions about the consequences and costs involved in each button (2), the child wanted to stop the study before the test phase (3), there was an experimental error (3) or the parent provided no birth date (1).

We ran a sensitivity power analysis with G*Power (Faul et al., 2007) to determine the effect size that our sample would have 80% power to detect (alpha = .05). The result confirmed that our sample size was large enough to detect a medium-sized main effect ($d = .45$) of either coin endowment (rich vs. poor) or agent type (social vs. non-social) and to detect a medium-sized interaction effect ($f = .20$) between coin endowment and agent type.

Experimental design. In Experiment 2, a new feature was that there were two additional conditions: *winner* condition and *loser* condition (Figure 3). Children in winner and fair experience conditions received 3 coins for four consecutive rounds, whereas children in loser and unfair experience conditions received 0 coins for four consecutive rounds. 99% of the child participants

correctly reported the number of coins they received during the second-party game, confirming the effectiveness of our coin endowment manipulation.

The difference between social (fair experience and unfair experience) and non-social conditions (winner and loser) was that children in the non-social conditions were told that a computer decides how to divide coins between the child participant and another player. Concretely, in non-social conditions, children saw two boxes on the computer screen. In one box, all the cards showed equal offers, whereas all the cards in the other box showed unequal offers. Children were told that the computer will choose one of the boxes and will decide how many coins the child can get. Once the computer chose one of two boxes, the same box was used for drawing a card throughout the rounds during the second-party game. Children (98%) correctly identified the agent (another player vs. computer) that was making the offers in the second-party game. We counterbalanced the color of boxes used in the non-social conditions.

Procedure. We used the same procedure as in Experiment 1, except for the following modifications. First, before and after they played the second-party game, the experimenter asked children to predict how a *new* divider will share six coins with a *new* recipient who did not have any previous history of social interactions or resource allocations (i.e., "How many coins do you think [divider's name] will give to [recipient's name]?"). These questions were asked to see whether children's expectations about resource allocations change depending on their second-party experience. Therefore, we posed this question before the second-party game and once again after the second-party game. We predicted that those who received allocations made by a computer (i.e., non-social conditions) would not change their expectations about how other people divide coins in this game, whereas those who received offers made by another player (i.e., social conditions) would change their expectations depending on whether they experience fair or unfair sharing.

Second, unlike Experiment 1, children played the second-party game *before*, not after, the practice trials of the third-party game. During piloting, we found that children in the new non-social

conditions had difficulties when both games were introduced together and showed less confusion when they played the second-party game before practicing the third-party game.

Results

Change in children's expectations about offer. We first calculated the change in children's expectations about the divider's offer by subtracting their prediction *before* the second-party game from that *after* the second-party game. Then, we ran a full linear model on children's change in the expectations about the offer with agent type (social vs. nonsocial), coin endowment (poor vs. rich), age, and interactions among the predictors. The comparison between the full model and the null model with the only intercept indicated that the full model provided a better fit to the data, $F(7, 150) = 2.99$, $p = .006$. There was no main effect and no interaction effect involving age (all ps > .54). Importantly, we found a significant interaction between agent type and coin endowment on children's change in expectations, $F(1, 153) = 9.09$, $p = .003$ (Figure 4).

Specifically, in social conditions, children who received unequal offers in the unfair experience condition ($M = -1.73$, $SD = 2.12$) changed their expectations about the offer more dramatically and expected more selfish offers from another divider compared to those who received equal offers in the fair experience condition ($M = -0.24$, $SD = 1.88$; $F(1, 75) = 10.68$, $b = 1.49$, $SE = 0.46$, $p = .002$). In contrast, in non-social conditions in which allocations were decided by a computer, children's change in expectations about another divider's offer did not differ between the winner ($M = -0.33$, $SD = 1.82$) and loser conditions ($M = 0.00$, $SD = 1.69$; $F(1, 77) < 1$, $p > .41$).

These results confirmed that agent type was manipulated successfully given that children differentiated offers from social versus non-social agents. Interestingly, those who experienced unequal offers from a social agent generalized their own experience to predict how another divider would treat a third-party. In contrast, those who experienced the same, unequal offers from a computer did not generalize their own experience to predict a divider's offer. This result suggests that when forming expectations about how individuals will share coins in this game, children do not rely on mere observations of allocations (i.e., how frequently did the offer occur?), but consider whether

the allocation was an act of giving by a social agent, in which case they adjust their expectations on how new individuals will behave in this context.

Rejection rate. A full GLMM on children's punishment (0 = acceptance, 1 = rejection) with agent type (social vs. nonsocial), coin endowment (poor vs. rich), offer type (equal vs. unequal), age, and interactions among the predictors as fixed effects and participant ID as a random effect provided a significantly better fit to the data than the null model with only random intercepts (LRT, $\chi^2(15) = 207.51$, $p < .001$). There were no effects involving trial number, LRT, $\chi^2(6) = 1.85$, $p > .93$ (see pp. 3 of the Supplemental Material for the analysis from the first trials). Additionally, we found similar results in rejection rate even after excluding a minority of children ($n = 11$) who reported that the players were pretend players (see p. 9 in Supplemental Material for descriptive statistics of these children).

One question was whether children in the unfair experience condition would punish less often than those in the other conditions, which can be determined by an interaction between coin endowment and agent type. The results revealed that there was no two-way interaction between agent type and coin endowment (LRT, $\chi^2(1) = 0.05$, $p > .82$), suggesting that children showed a similar rate of third-party punishment across conditions (see Figure 5). Further, there were no main effects involving coin endowment (poor vs. rich) or agent type (social vs. nonsocial), LRT, $\chi^2(1) < 1$, ps > .42, and no interactions involving these variables (all ps > .08). These findings suggest that third-party punishment in children is affected neither by the number of coins they received nor by an agent who treated them fairly or unfairly. The Bayes factors also confirmed the non-significant effects. Specifically, we found strong evidence for an *absence* of an interaction between agent type and coin endowment ($BF_{01} = 10.18$), suggesting that these data were about 10 times more likely to be observed under the hypothesis predicting that there is no such interaction effect than the hypothesis predicting the interaction. Additionally, BFs for main effects of coin endowment or agent type ($BF_{01} = 15.72$ and $BF_{01} = 10.91$, respectively) suggest that the data were more consistent with the hypothesis that punishment was affected neither by the number of coins they had nor by the type of the agent who

made offers to the child. We also found an absence of a condition effect when we collapsed children's responses across two experiments (see pp. 3 of the Supplemental Material).

We found a significant interaction between age and offer type (LRT, χ^2 (1) = 23.62, b = 0.82, SE = 0.17, p < .001). Specifically, older children were less likely to punish equal offers than were younger children (LRT, χ^2 (1) = 9.00, b = -0.47, SE = 0.16, p = .003), replicating the results from Experiment 1. Furthermore, older children were more likely to punish unequal offers (LRT, χ^2 (1) = 11.55, b = 0.41, SE = 0.12, p < .001). In other words, compared to younger children, older children showed an increase in their tendency to punish unequal offers, while they showed a decrease in the tendency to punish equal offers. Also, we found very strong evidence in favor of the interaction between age and offer type (BF_{10} = 8290), indicating that the data were 8290 times more likely to be observed under the hypothesis predicting the interaction over the hypothesis predicting no interaction effect (see pp. 5-6 of the Supplemental Material for more detailed results about Bayes factors). Together, these findings confirm the results in Experiment 1 that children's punishment becomes more selective with age.

Discussion of Experiment 2

In Experiment 2, children showed a similar rate of costly punishment regardless of their experience as a second party. It is unlikely that children's third-party punishment is influenced by the number of coins they received, given that children in poor conditions and rich conditions showed a comparable rate of third-party punishment. Also, our findings are inconsistent with the hypothesis that children's inference on how individuals treat each other in the game affects their third-party punishment. Specifically, children in the unfair experience conditions did not show a significant difference in third-party punishment from those in other conditions. Interestingly, we found that children who received unequal offers from a social agent expected more selfish offers, whereas those who received unequal offers from a non-social agent did not show a dramatic change in their expectations. Yet, the change in children's expectation about offers in the unfair experience condition did not result in a significant decrease of their own punishment.

One major purpose of this study was to assess the robustness of the initial finding that children punish *less* after an unfair than fair experience. Although the effect found in Experiment 1 was statistically significant, our Bayes factor analyses had indicated that this effect might not be strong. In fact, Experiment 2 showed that this effect did not replicate. Moreover, in both experiments, the Bayes factors for the effect of condition were in favor of the *absence* of the effect. Based on these results across experiments, it is more reasonable to conclude that there is a lack of support for the influence of second-party experience on third-party punishment.

In contrast, we found a clear and robust developmental pattern in both Experiments 1 and 2: Children become more selective in their enactment of third-party punishment. Experiment 2 replicated the finding that older children were less likely to punish equal offers. In addition, Experiment 2 showed that older children were more likely to punish unequal offers, which was detected potentially by testing a larger sample compared to Experiment 1. These findings are consistent with the hypothesis that with age, children develop sophisticated general fairness principles that are not easily swayed by their own immediate experiences in the same context.

General Discussion

The current study examined the influence of experiencing (un)fairness on third-party punishment in 5- to 9-year-old children. This study was possible owing to our novel computer-based task in which children's experiences of (un)fairness were manipulated systematically. One main goal was to test the hypothesis that being exposed to unfairness personally would enable children to better simulate a third-party's perspective, leading to an increase in following third-party punishment. However, across two experiments, we found no evidence supporting this hypothesis. If anything, in Experiment 1, children showed a decrease in third-party punishment after receiving unfair offers. Further analyses revealed that this effect was overall weak and did not replicate in Experiment 2 where children showed similar rates of third-party punishment regardless of their experience. Together, these findings suggest that at least in the short-term, immediate experience of (un)fairness does not affect subsequent third-party punishment reliably in children ages 5 to 9.

Although there was no indication that prior experience affected children's punishment robustly, there was a very strong and consistent pattern regarding third-party punishment observed in both experiments. We found that with age, children became more selective about the enactment of third-party punishment: They became more likely to punish unequal offers consistently over equal offers. These results support the hypothesis that children develop more sophisticated fairness principles that guide how to divide resources and how to respond to a third-party's unfair act. Moreover, these results show that children punish for fairness reasons and not out of spite or some reason unrelated to fairness. Lastly, Experiment 2 showed that children punish unfairness regardless of the relative number of resources they have. For children, it is not a luxury to intervene against unfairness when they have a surplus of resources, but they punish even when they have been relatively deprived of resources.

Our findings contribute to the existing literature by showing that children employ fairness in an increasingly principled fashion over development (Blake & McAuliffe, 2011; Elenbaas et al., 2016; Jordan et al., 2014; Rizzo & Killen, 2016; Shaw & Olson, 2012). From around age 6, children were willing to pay a personal cost to enforce the fairness norms on another individual even when they were an unaffected third-party. Moreover, this tendency became increasingly selective and systematic with age. Our data suggest that children's enforcement of fairness norms is unlikely to be swayed by their immediate personal experiences.

One remaining question is why children's second-party experience did not influence their third-party punishment. One possibility is that the second-party experience of (un)fairness was too short and harmless to influence third-party punishment. Being exposed to unequal offers briefly may not be sufficient to induce such a change. In fact, studies with adults using more extreme moral harms suggest that victims who experienced extreme suffering (e.g., abuse, violence) were more likely to report feelings of empathy for victims who are undergoing suffering and were more likely to participate in volunteer activities to help victims compared to those who did not experience such suffering (Staub & Vollhardt, 2008). Conversely, it also is known that people who were exposed to abuse or violence are more at risk of being a perpetrator than those who were not exposed to these

events (Craig & Sprang, 2007). In either case, it is possible that more extreme, long-term experiences of unfair treatment could have induced a change in children's third-party punishment. However, our manipulation included only short-term, mild experiences of unfairness and not extreme or long-term unfair experiences — an experimental manipulation that would not be possible for ethical reasons. One viable approach for future research would be to collect data on children's experiences of unfairness in their daily life and relate it to their willingness to intervene in a third-party context (with the limitation that data would only be correlational). Also, a recent study found that children's socioeconomic status predicted their fairness perceptions about an offer they received (Peretz-Lange et al., 2022). Thus, it would be interesting to examine how children's socioeconomic status as a proxy for their resource deprivation in daily life shapes their willingness to enact costly third-party punishment.

Another consideration to explain our results is that children might not have made a connection between their own experience and the other peer's experience. That is, children failed to take the perspective of the recipient in the third-party game. Prior research with adults has shown that adult participants, who were instructed to take the perspective of the recipient of an unfair offer as much as they can, showed a higher third-party punishment than those instructed to take an objective perspective and remained detached from the recipient (Pfattheicher et al., 2019). This study implies the importance of taking the victim's perspective in third-party punishment which might not be a skill automatically employed by children. Future research could examine if children would show an increase in third-party punishment when they are given more explicit instruction to take the recipient's perspective before making their own decisions.

We would like to address a few limitations in the current research. One potential concern is that the non-significant effect across the experience conditions reflects a lack of understanding of the task or conditions. However, this possibility is unlikely given the ease with which children passed manipulation checks. Specifically, virtually all children (99%) in Experiments 1 and 2 correctly reported the number of coins they received during the second-party game, suggesting that they paid attention to the information. Also, most children (98%) in Experiment 2 correctly identified which

agent had shared or not shared with them, suggesting that they paid attention to whether the offers were made by a peer or a computer. Furthermore, children formed expectations of more selfish offers in the unfair experience condition, but not the non-social loser condition, showing that children differentiated offers made by a social agent from those made by a computer and made adequate inferences about future resource allocations. Based on the results, it would be reasonable to conclude that the lack of a condition difference is not due to a lack of task understanding.

Another concern is the extent to which children's responses in the current computer-based task can be generalized to their everyday life. We used the computer-based task because it allowed for better experimental control, systematic manipulation of (un)fairness experience from peer dividers, and measurements of children's reaction times (see pp. 2-4 of the Supplemental Material for results from reaction times). However, it only simulates social interactions. Therefore, it is possible that children might respond differently when they were treated unfairly by a peer in a face-to-face interaction which mirrors their everyday interactions. There has been a similar question about the external validity of computer-based punishment tasks in adult literature (Guala, 2012). Moreover, researchers have begun to examine how adults respond to real-life conflicts outside of lab experiments. For instance, Molho and colleagues (2020) have shown that adults use direct punishment and indirect punishment differently depending on contexts. Also, punishment of a transgressor is relatively rare when adults are asked to recall how they responded to past real-life conflicts (Pedersen et al., 2020). Thus, it would be important for future research to examine the external validity of children's punishment in a computer task. One way to address this question is to investigate if there is a correlation between children's third-party punishment in lab experiments and their intervention in a naturalistic setting. For example, it is possible that children who showed more punishment in a lab study are more likely to intervene against a transgression (e.g., bullying) in classroom setting. Also, it would be interesting to examine how children intervene differently (e.g., verbally sticking up for the victim, reporting the bully to the teacher) in a naturalistic setting where their option of intervention is not limited to punishment.

Finally, we would like to address other potential mechanisms for third-party punishment development that were not tested in our study. Although the current study examined the potential role of (un)fairness experience on the development of punishment, it is by no means the only candidate that could explain its development. One possibility is that children showed increased third-party punishment over development because older children begin to consider their reputation. Studies with adults suggest that third-party punishment signals one's trustworthiness, and thus generates positive reputations (Barclay, 2006; Jordan et al., 2016; Kurzban et al., 2007; Nowak & Sigmund, 2005; Raihani & Bshary, 2015). Based on this account, it is possible that older children in the current study showed increased punishment against unfairness because they become more sensitive to reputational concerns with age (see Engelmann & Rapp, 2018 for a review). Thus, future research should examine how children's reputational concerns and norms of fairness might interact in the emergence of third-party punishment.

Another mechanism that potentially could contribute to the development of punishment is children's increasing understanding of the communicative function of punishment over development. A recent study (Marshall et al., 2020) has shown that 5- to 7-year-olds are more likely to enact punishment toward a transgressor when the transgressor could know the reason for receiving punishment than when the person could not know the reason. Therefore, it is plausible that older children in the current study applied fairness principles and punished more often potentially because they expected a future encounter with the same unfair divider and intended to communicate with the person (e.g., teaching a lesson). Although the current study is not designed to address this possibility, it would be important to examine how children's expectations of future interactions impact subsequent third-party punishment.

In summary, across two experiments, our data do not support the notion that the immediate experience of (un)fairness influences children's third-party punishment. Rather, our results provide converging evidence across two experiments that with age, children develop a sophisticated

application of fairness norms that is not easily swayed by immediate personal experiences of how other children treat them in the same context.

References

Aczel, B., Palfi, B., Szollosi, A., Kovacs, M., Szaszi, B., Szecsi, P., Zrubka, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2018). Quantifying support for the null hypothesis in psychology: An empirical investigation. *Advances in Methods and Practices in Psychological Science, 1*(3), 357–366. https://doi.org/10.1177/2515245918773742

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

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

Bernhard, R., Martin, J., & 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., & McAuliffe, K. (2011). "I had so much it didn't seem fair" Eight-year-olds reject two forms of inequity. *Cognition*, *120*(2), 215–224. doi:10.1016/j.cognition.2011.04.006

Bloom, P. (2013). *Just babies: The origins of good and evil.* Broadway Books.

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

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378-400. https://doi.org/10.3929/ethz-b-000240890

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

Craig, C. D., & Sprang, G. (2007). Trauma exposure and child abuse potential: Investigating the cycle of violence. *The American Journal of Orthopsychiatry*, *77*(2), 296–305. https://doi.org/10.1037/0002-9432.77.2.296

Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, *31*(3), 137–149. https://doi.org/10.1111/j.1540-4560.1975.tb01000.x

Elenbaas, L., Rizzo, M. T., Cooley, S., & Killen, M. (2016). Rectifying social inequalities in a resource allocation task. *Cognition*, *155*, 176–187. https://doi.org/10.1016/j.cognition.2016.07.002

Elenbaas, L., Rizzo, M. T., & Killen, M. (2020). A developmental-science perspective on social inequality. *Current Directions in Psychological Science*, *29*(6), 610–616. https://doi.org/10.1177/0963721420964147

Engelmann, J. M., & Rapp, D. J. (2018). The influence of reputational concerns on children's prosociality. *Current Opinion in Psychology*, *20*, 92–95. https://doi.org/10.1016/j.copsyc.2017.08.024

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. doi: 10.3758/bf03193146

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language*, *1*(2), 158–171. https://doi.org/10.1111/j.1468-0017.1986.tb00324.x

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, *35*(01), 1–15. https://doi.org/10.1017/S0140525X11000069

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. doi:10.1016/j.cognition.2014.06.001

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, 108*(50), *19931-19936.* doi:10.1073/pnas.1110306108

Harris, P. L. (1992). From simulation to folk psychology: The case for development. *Mind & Language*, *7*(1-2), 120–144. https://doi.org/10.1111/j.1468-0017.1992.tb00201.x

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Henrich., N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–1770. doi:10.1126/science.1127333

House, B. R., Kanngiesser, P., Barrett, H. C., Yilmaz, S., Smith, A. M., Sebastian-Enesco, C., Alejandro, E., & 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

Jeffreys, H. (1961). *Theory of probability*. Oxford, England: Oxford University Press.

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

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, 111(35), 12710-12715.* doi:10.1073/pnas.1402280111

Kirkland, K., Jetten, J., & Nielsen, M. (2020). The effect of economic inequality on young children's prosocial decision-making. *British Journal of Developmental Psychology*, *38*(4), 512–528. https://doi.org/10.1111/bjdp.12334

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the hood of third-party punishment reveals design for personal benefit. *Psychological Science*, *27*(3), 405-418. https://doi.org/10.1177/0956797615624469

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, *28*(2), 75–84. https://doi.org/10.1016/j.evolhumbehav.2006.06.001

Marshall, J., Yudkin, D. A., & Crockett, M. J. (2020). Children punish third parties to satisfy both consequentialist and retributive motives. *Nature Human Behaviour*. *5*(3), 361-368. https://doi.org/10.1038/s41562-020-00975-9

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, *1*(2), 1-9. https://doi.org/10.1038/s41562-016-0042

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

Molho, C., Tybur, J. M., Van Lange, P. A. M., & Balliet, D. (2020). Direct and indirect punishment in daily life. *Nature Communications*, *11*(3432). https://doi.org/10.1038/s41467-020-17286-2

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity by image scoring. *Nature*, *393*(6685), 573-577. https://doi.org/Doi 10.1038/31225

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

Peretz-Lange, R., Harvey, T., & Blake, P. R. (2022). Socioeconomic status predicts children's moral judgments of novel resource distributions. *Developmental science*, e13230. Advance online publication. https://doi.org/10.1111/desc.13230

Pfattheicher, S., Sassenrath, C., & Keller, J. (2019). Compassion magnifies third-party punishment. *Journal of Personality and Social Psychology*, *117*(1), 124–141. https://doi.org/10.1037/pspi0000165

Raihani, N. J., & Bshary, R. (2015). The reputation of punishers. *Trends in Ecology and Evolution*, *30*(2), 98–103. https://doi.org/10.1016/j.tree.2014.12.003

Raihani, N. J., & Bshary, R. (2019). Punishment: one tool, many uses. *Evolutionary Human Sciences*, *1*(e12), 1–26.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rizzo, M. T., & Killen, M. (2016). Children's understanding of equity in the context of inequality. *British Journal of Developmental Psychology*, *34*(4), 569–581. https://doi.org/10.1111/bjdp.12150

Shaw, A. (2013). Beyond "to share or not to share": The impartiality account of fairness. *Current Directions in Psychological Science*, *22*(5), 413–417. https://doi.org/10.1177/0963721413484467

Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General*, *141*(2), 382–395. https://doi.org/10.1037/a0025907

Sheskin, M., Bloom, P., & Wynn, K. (2014). Anti-equality: Social comparison in young children. *Cognition*, *130*(2), 152–156. https://doi.org/10.1016/j.cognition.2013.10.008

Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PLoS ONE*, *8*(3): e59510. https://doi.org/10.1371/journal.pone.0059510

Staub, E., & Vollhardt, J. (2008). Altruism born of suffering: The roots of caring and helping after victimization and other trauma. *The American Journal of Orthopsychiatry*, *78*(3), 267–280. https://doi.org/10.1037/a0014223

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology*, *29*(1), 124–130. doi:10.1348/026151010X532888

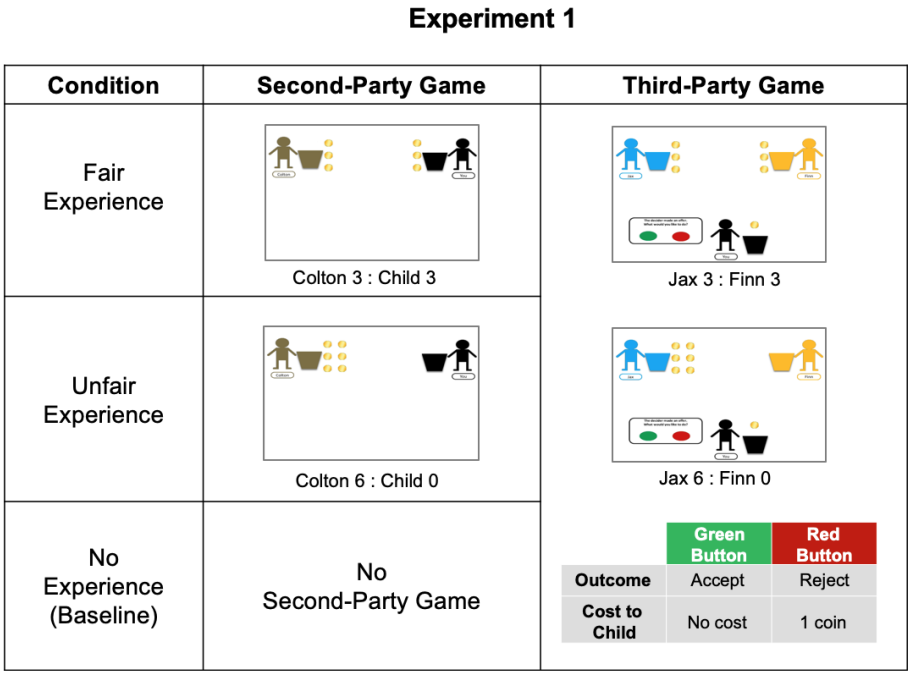
Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian analysis, 16*(2), 667-718.

Wagenmakers, E. J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, *25*(3), 169–176. https://doi.org/10.1177/0963721416643289

Yudkin, D. A., Van Bavel, J. J., & Rhodes, M. (2020). Young children police group members at personal cost. *Journal of Experimental Psychology: General, 149*(1), 182-191. https://doi.org/10.1037/xge000061

Figure 1

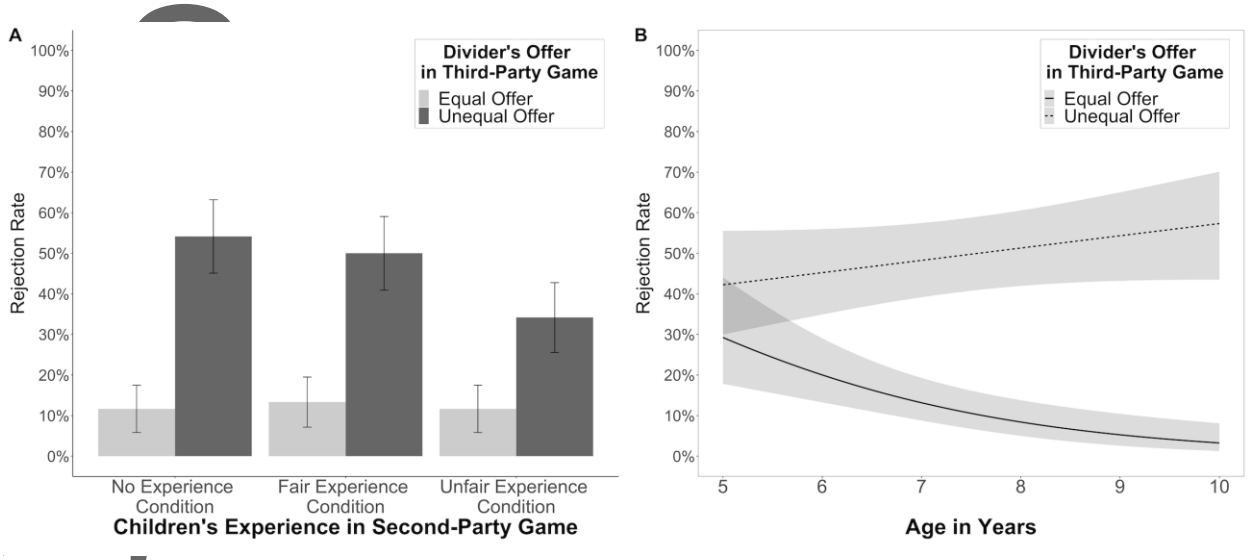
Schematic Representation of the Design in Experiment 1



Note. In this example, Colton (left) was a divider and the child participant (right) was a recipient during the second-party game (i.e., experience phase). In the subsequent third-party game (test phase), Jax (top left) was a divider and Finn (top right) was a recipient. The child participant (bottom) was a third-party observer. Regardless of conditions, every child saw three equal and three unequal offers during the third-party game. The divider and recipient in the third-party game differed from those in the practice trials or those in the second-party game.

Figure 2

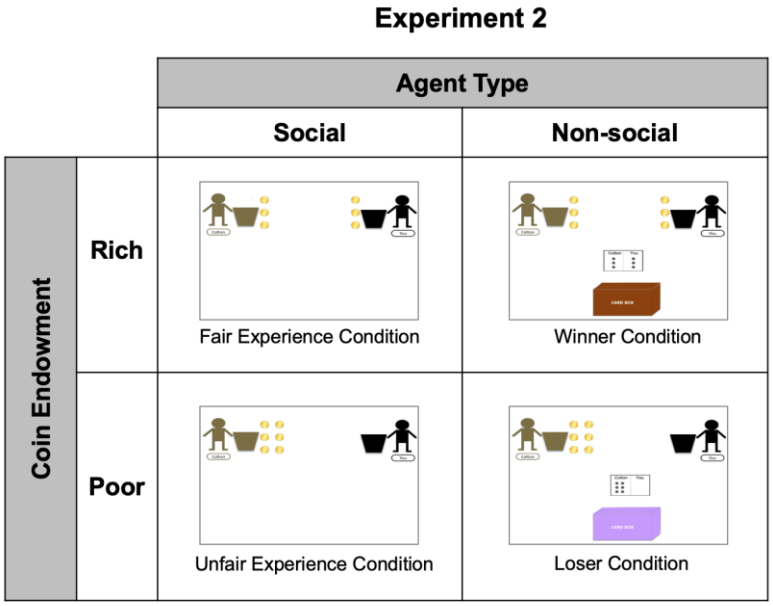
Children's Rejection Rate in Experiment 1



Note. (A) Children's rejection rate by condition in Experiment 1. (B) Estimates of rejection rate based on the final model in Experiment 1 (collapsed across conditions). The final model included main effects of condition, offer type, age and an interaction effect between age and offer type. Error bars and confidence bands represent 95% CI.

Figure 3

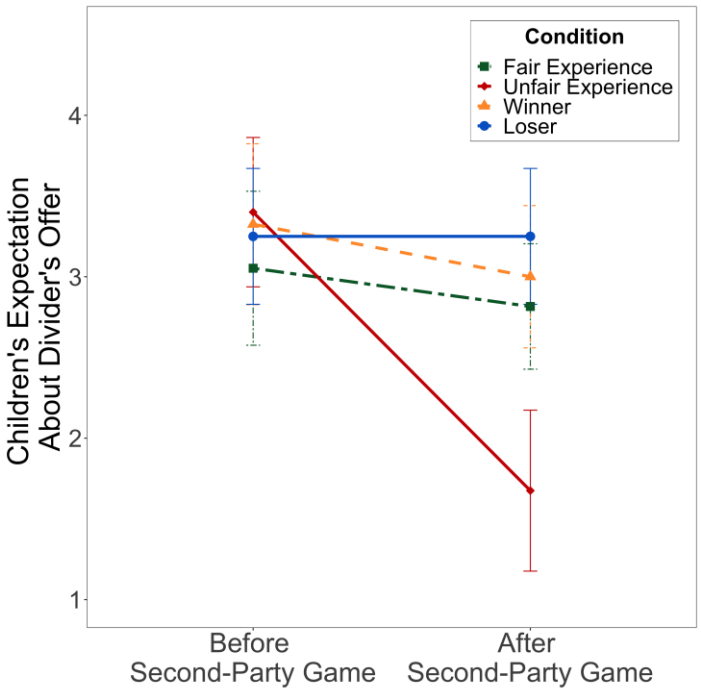
Representation of the Design in Experiment 2



Note. The figure illustrates how the second-party game in each condition differed by coin endowment (rich vs. poor) and agent type (social vs. nonsocial). In this example, 6 coins were divided between Colton (left) and the child participant (right). The labels "Fair", "Unfair", "Winner", "Loser", "Rich", "Poor" etc. were added here for representation of the experimental design. Children never heard these words during the study.

Figure 4

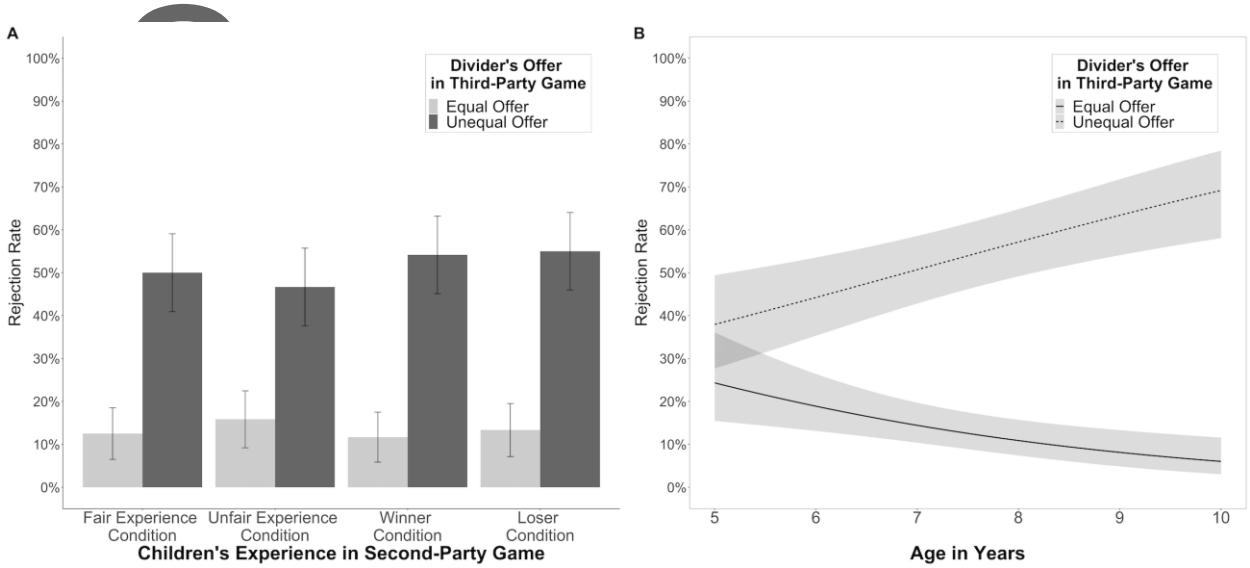
Children's Expectation About a Divider's Offer Before and After the Second-Party Game



Note. A 3 on the Y-axis indicates an expectation of an equal offer (sharing 3 out of 6 coins). Error bars represent 95% CI.

Figure 5

Children's Rejection Rate in Experiment 2



Note. (A) Children's rejection rate by condition in Experiment 2. (B) Estimates of rejection rate based on the final model in Experiment 2 (collapsed across conditions). The final model included main effects of agent type, coin endowment, offer type, age, and an interaction effect between offer type and age. Error bars and confidence bands represent 95% CI.