Cognition 134 (2015) 1–10

Contents lists available at ScienceDirect

Cognition

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

Costly third-party punishment in young children

Katherine McAuliffe [a,b,c,*], Jillian J. Jordan [b,c], Felix Warneken [b]

[a] Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, United States
[b] Department of Psychology, Harvard University, Cambridge, MA, United States
[c] Department of Psychology, Yale University, New Haven, CT, United States

ARTICLE INFO

Article history:
Received 20 October 2013
Revised 18 August 2014
Accepted 21 August 2014

Keywords:
Third-party punishment
Cooperation
Inequity aversion
Fairness
Social cognition

ABSTRACT

Human adults engage in costly third-party punishment of unfair behavior, but the developmental origins of this behavior are unknown. Here we investigate costly third-party punishment in 5- and 6-year-old children. Participants were asked to accept (enact) or reject (punish) proposed allocations of resources between a pair of absent, anonymous children. In addition, we manipulated whether subjects had to pay a cost to punish proposed allocations. Experiment 1 showed that 6-year-olds (but not 5-year-olds) punished unfair proposals more than fair proposals. However, children punished less when doing so was personally costly. Thus, while sensitive to cost, they were willing to sacrifice resources to intervene against unfairness. Experiment 2 showed that 6-year-olds were less sensitive to unequal allocations when they resulted from selfishness than generosity. These findings show that costly third-party punishment of unfair behavior is present in young children, suggesting that from early in development children show a sophisticated capacity to promote fair behavior.

1. Introduction

One striking feature of human social behavior is third-party punishment. A large body of work has shown that adults are willing to punish uncooperative individual seven when punishers are not personally affected by the uncooperative behavior and when punishment is costly (Buckholtz et al., 2008; Fehr & Fischbacher, 2004; Henrich et al., 2006, 2010; Raihani, Thornton, & Bshary, 2012). Adults intervene in three-party contexts to punish both distributional (dictator game) and cooperative (prisoner's dilemma) norms (Fehr & Fischbacher, 2004).

It has been argued that third-party punishment occurs because it plays an important role in promoting norms of cooperative behavior in human societies (Fehr & Fischbacher, 2004; Fehr, Fischbacher, & Gächter, 2003), though the mechanisms supporting third-party punishment are debated (Krasnow, Cosmides, Pedersen, & Tooby, 2012). The motivators of costly punishment have been studied extensively and psychological research with adults has shown that punishment is motivated by both inequality aversion (Dawes, Fowler, Johnson, McElreath, & Smirnov, 2007; Raihani & McAuliffe, 2012) and an aversion to the selfish intentions of defectors (Cushman, Dreber, Wang, & Costa, 2009; Falk, Fehr, & Fischbacher, 2003, 2008; Fehr et al., 2003; Nelissen & Zeelenberg, 2009; Nelson, 2002; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Despite its prevalence in humans, third-party punishment appears to be rare if not absent in non-human animals, including our closest ape relatives (Raihani et al., 2012; Riedl, Jensen, Call, & Tomasello, 2012), suggesting it may be a unique feature of human social interaction.

Costly third-party punishment is proposed to be of critical importance for the maintenance of cooperation and social order (Buckholtz & Marois, 2012; Fehr &

* Corresponding author at: Department of Psychology, Yale University, New Haven, CT, United States.
E-mail address: katherine.mcauliffe@yale.edu (K. McAuliffe).

http://dx.doi.org/10.1016/j.cognition.2014.08.013
0010-0277/

Fischbacher, 2004; Henrich et al., 2006). However, the developmental origins of costly third-party punishment have been unexplored. It is possible that costly third-party punishment emerges after childhood because it depends on factors such as the acquisition of a rich set of social norms; knowledge about the consequences of uncooperative behaviors on the social group and potentially the practice and responsibility that comes with adult social roles in the family, at work, and in the community at large (Rogoff, 2003). Alternatively, it is possible that third-party punishment emerges during childhood as an integral part of children's developing cooperativeness and fairness concerns (Baumard, Andre, & Sperber, 2013; Damon, 1977; Tomasello & Vaish, 2013). If third-party punishment does indeed emerge in early childhood it would suggest that only limited social experience is required for children to enact this important feature of human fairness behavior. Only by studying young children can we begin to understand the psychological and social factors that give rise to costly third-party punishment in humans.

Indirect evidence for the building blocks of costly third-party punishment comes from studies in which children are directly affected by unfair behavior. Specifically, a growing body of evidence suggests that children prefer equal shares (Hamann, Warneken, Greenberg, & Tomasello, 2011; LoBue, Nishida, Chiong, DeLoache, & Haidt, 2010; Schmidt & Sommerville, 2011; Warneken, Lohse, Melis, & Tomasello, 2011) and even pay a cost to prevent themselves from receiving unfair treatment that puts them at a disadvantage relative to a peer (Blake & McAuliffe, 2011; Fehr, Bernhard, & Rockenbach, 2008; Shaw & Olson, 2012). However, it is unknown when children begin to pay costs to prevent *others* from receiving unfair treatment in a third-party context. Do children care about fairness solely when they are affected or do they exhibit a motivation to enforce fairness even as third parties? One study found that five-year-old children paid costs to punish a puppet who had behaved selfishly towards both themselves and another puppet (Robbins & Rochat, 2011). While this may suggest that children intervene on behalf of others, this design confounds second and third-party punishment: children may have been motivated entirely by the fact that *they* were treated unfairly. Thus, there is currently no study that has addressed whether children pay personal costs to prevent unfair behavior that does not affect themselves.

While it is unknown when children begin to intervene against fairness norm violations, there is evidence that children intervene on behalf of others in other contexts. More specifically, young children at 3 years and younger respond to and intervene against antisocial behavior (Hamlin, Wynn, Bloom, & Mahajan, 2011; Kenward & Östh, 2012; Vaish, Missana, & Tomasello, 2011) and against moral and conventional norm violations (Schmidt, Rakoczy, & Tomasello, 2012). Indeed, from a young age, children appear to treat conventional and moral norm violations differently (Smetana, 1983, 1984) and, as children age, they begin to understand that norms can be specific to certain groups (Killen, Rutland, Abrams, Mulvey, & Hitti, 2013). Moreover, children have a nuanced understanding of norm violations, protesting when individuals violate conventional norms in rule games (Rakoczy, Warneken, & Tomasello, 2008) and pretend play (Wyman, Rakoczy, & Tomasello, 2009). In addition, there are results suggesting that when children are presented with stories of characters performing good and bad acts, children understand that good acts should be rewarded and bad acts should be punished (Karniol, 1978; Leman & Björnberg, 2010). There is also evidence to suggest that children occasionally spontaneously administer punishment to their peers in real-life situations (Masters & Furman, 1981). Thus, previous work shows that young children understand norms and can use punishment as a way to enforce those norms. However, past work has not investigated interventions about fairness or, critically, tested whether children will sacrifice their own resources to punish selfish behavior in others.

In this study we aimed to test whether children would intervene in an interaction between two unfamiliar individuals, in which one individual received fewer resources than another. We tested whether children would systematically punish unfair behavior and whether they would sacrifice their own resources to do so. Our goal was to document the age of emergence of a behavior that has not been demonstrated in children before.

On the basis of evidence that children will pay costs to prevent inequality in two-party contexts by age 5–6 (Blake & McAuliffe, 2011; Hamann et al., 2011), we tested 5- and 6-year-old children and hypothesized that third-party punishment of selfishness would emerge in this window. Furthermore, we manipulated whether punishment required subjects to pay personal costs, and hypothesized that children would be more likely to punish when it was cost-free than when it was costly. Finally, we investigated whether children would be equally likely to punish any unequal resource allocations, or whether they would show more sensitivity to inequity when it resulted from selfishness than generosity. On the basis of evidence that adults are sensitive to selfish intentions (Fehr & Fischbacher, 2004), we hypothesized that if children did engage in costly third-party punishment, they would punish selfish allocations more harshly than generous allocations.

2. Experiment 1: Will children pay a cost to punish selfish allocations?

Children played a game in which they were making decisions on behalf of two absent peer children. Children were told that one of the absent children (divider) had divided candies between him/herself and a partner (recipient). The participant (or subject) in our task could accept the divider's allocation, in which case candies would be saved for the absent children to collect. Alternatively, the subject could reject the divider's allocation (thus punishing the divider), in which case the candies would be thrown away. Rejections in this game can be defined as punishment because they inflict a cost on the divider (Gardner & West, 2004; Lehmann & Keller, 2006). In order to investigate costly punishment, we manipulated the personal cost of rejection to punisher. In the free condition, participants could reject allocations without giving up their own resources. In the costly condition, participants had to

sacrifice their own resources in order to reject allocations on behalf of the absent children.

2.1. Method

2.1.1. Participants

We tested 32 5-year-olds (M = 65.7 months, range = 61–70 months, 16 females); 32 6-year-olds (M = 77.3 months, range = 72–83 months, 16 females). Three additional children were excluded due to experimenter error (one), experimental interruption (one), or because the subject did not like the food resource used in the task (one).

2.1.2. Design

Participants were assigned to either costly or free conditions (between-subject) and presented with a total of 12 trials. On each trial, participants decided to accept or reject a proposed allocation of candy. Allocations were either equal (3–3, divider kept 3 and offered 3) or selfish (6–0, the divider kept 6 and offered 0). Each participant decided about 6 equal and 6 unequal allocation trials. Equal and unequal trials were presented in a pseudo-random order, with the constraint that no more than two of the same allocation type could be presented consecutively. Because unequal allocations were entirely selfish (all for divider, none for recipient), rejections of unequal allocations imposed a cost on the divider (who was prevented from receiving six candies), but not on the recipient (who received no candies either way, and was thus unaffected by punishment).

2.1.3. Procedure

An experimenter introduced children to the testing area and to Skittles, the food resource being used. Next, participants were familiarized with the apparatus (Fig. 1, also see Supplementary Material (SM), Figs. S1–S2 and videos): they were shown that Skittles could be moved back and forth across the apparatus, and that the handle could be moved in both the green (accept) and red (reject) directions. Subjects were then introduced to crayon drawings of the absent divider and recipient and were told that they had played the game on the previous day. Participants were told that the divider had proposed allocations of candy between him or herself and the recipient. The proposed allocations were illustrated on cards (see SM, Fig. S3). Children were led to believe that their decisions in the game would affect the payoffs of real peers who would be coming back to collect their respective allocations (see below for comprehension checks). Participants were instructed on how to use the handle to make decisions and asked to pull the handle in both green (accept) and red (reject) directions. They were also shown one equal and one unequal allocation on a card (shown in counterbalanced order).

Participants received twenty-five Skittles at the beginning of the game. Their Skittles were placed in a box on the table. Our main manipulation was that in the ***Costly condition***, children had to pay one of their Skittles to reject allocations, whereas acceptances were free. In the ***Free condition***, they did not have to pay for either. The participants' actions in these conditions were matched very closely, so that the only difference was whether rejections were costly or not.

In the Costly condition, subjects' Skittles were placed in a green box (see SM, Fig. S4). Before accepting allocations by pulling the handle in the green direction, children took one Skittle out of the green box and dropped it through the hole in the same green box. Before rejecting allocations by pulling the handle in the red direction, they dropped one Skittle in a red box. Children knew that they were allowed to take home the Skittles in the green, but not the red box, so that rejections required giving up Skittles. In the matched Free condition, subjects' Skittles were placed in a box that was green on one side and red on the other side (see SM, Fig. S4). Before accepting allocations, children took one Skittle out of the box and dropped it through a hole on the green side of the box. Before rejecting allocations, children took one Skittle out of the box and dropped it through a hole on the red side of the same box. Children knew that they could take home all the Skittles in the box and thus rejections came at no cost in this condition.

Prior to test trials, participants were asked several comprehension questions to ensure that they understood (1) that the actor and recipient had not taken Skittles home; (2) that the actor understood where to place a Skittle before pulling the handle to green or red zone (i.e., in the green or red box in the Costly condition, or in which side of the box in the Free condition); (3) who would get to take home the Skittles in the actor and recipient's bags; and (4) what would happen to the Skittles in each box at the end of the game. Virtually all participants answered comprehension questions either spontaneously correctly or correctly after additional questioning (98% to (1), 98% to (2), 100% to (3) and 97%, to (4)). The other children either never answered the questions correctly or were not asked due to experimenter error. Excluding children who did not pass comprehension checks did not change our results (see SM, Table S1).

On each test trial, six Skittles were pushed to the divider's side. The subject was then shown an allocation card and the Skittles were distributed accordingly. Children were reminded to put a Skittle through one of the holes before pulling the handle. The participant then had an opportunity to either accept or reject an allocation by pulling the handle in the green or red directions, respectively. If a subject decided to accept an allocation, Skittles were transferred to bags. If a subject decided to reject an allocation, Skittles disappeared underneath the apparatus.

At the end of the procedure, a second experimenter entered the room and asked a series of post-experimental questions including whether the child believed that the actor and recipient (1) were real and (2) would be coming to collect their Skittles. The majority of children answered that they believed the actor and the recipient to be real and will come to collect their Skittles (84% of 5-year-olds and 91% of 6-year-olds confirmed their belief in response to at least one of these questions). Analyses controlling for children's belief in the experimental manipulation lead to the same pattern of results reported below (see SM, Table S2 and Fig. S5). Following this check, Experimenter 1 debriefed the child and explained that the actor and recipient were in fact not real and were just pretend.

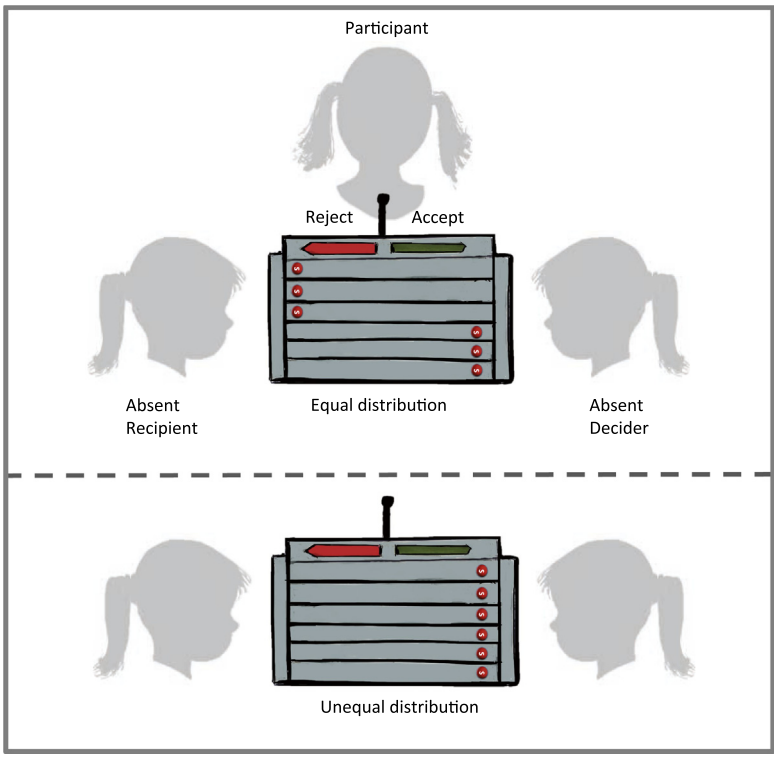

Fig. 1. Diagram of experimental set-up. The apparatus consisted of six 'Skittle pods' that could be moved back and forth across the surface of the apparatus and a handle that could be moved in both green (accept) and red (reject) directions. Participants made decisions about whether to accept or reject resource allocations between the divider and recipient. In the costly condition (depicted), participants moved one of their own Skittles from the green box into the red box before rejecting allocations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.1.4. Data coding and analysis

All sessions were videotaped. Our main variable of interest was whether children decided to accept or reject allocations. Children's decisions were live coded by an observer and later recoded by an independent video coder. Disagreements between live and video coding were resolved by re-watching the video (disagreements were rare: 1% of all trials).

All statistical analyses were conducted with R statistical software (version R 3.0.1; R Development Core Team, 2013). Decision data were analyzed using Generalized Linear Mixed Models (GLMMs) with a binary response term (reject = 1, accept = 0) (Bolker et al., 2009). Mixed models were run using the package 'lme4' (Bates, Maechler, & Bolker, 2012). In all models subject identity (ID) was fit as a random effect to control for repeated measures.

Our procedure was as follows: (1) we examined a null model, which included subject ID as the only explanatory variable; (2) we created a full model, which included the predictor variables Distribution (equal, unequal), Cost (costly, free), Age group (5-year-olds, 6-year-olds) and Gender (male, female), as well as all two-way interactions between distribution and other predictors; (3) the full model was compared to the null model using a likelihood ratio test (LRT) to test whether the inclusion of predictors provided a better fit to the data than subject ID alone. Unless otherwise noted, full models provided a better fit to data than null models; (4) a minimal model was created from the full model by sequentially dropping single terms from the model and testing whether their inclusion improved the model fit using LRTs. In addition to conducting analyses addressing our hypotheses, we examined whether children's performance was influenced by the first distribution they saw (equal, unequal). This factor did not influence their rejections. We also explored children's behavior across trials by visually inspecting patterns of rejection. We found that trial number did not predict rejections.

All figures show raw data and binomial confidence intervals were calculated using the Agresti–Coull method (Agresti & Coull, 2010).

2.2. Results

Data from Experiment 1 are shown in Fig. 2. This figure suggests that there was an age difference in how participants responded to inequality, with 6-year-olds showing a stronger sensitivity to distributional inequity than 5-year-olds. To test for the effects of distribution and age, we ran a full GLMM with all predictors and two-way interactions with distribution and compared this model to a reduced model without an age group x distribution interaction. The model including the age group $\times$ distribution interaction provided a significantly better fit to the data (LRT, $\chi^2(1) = 12.10$, $p < 0.001$), showing a significant

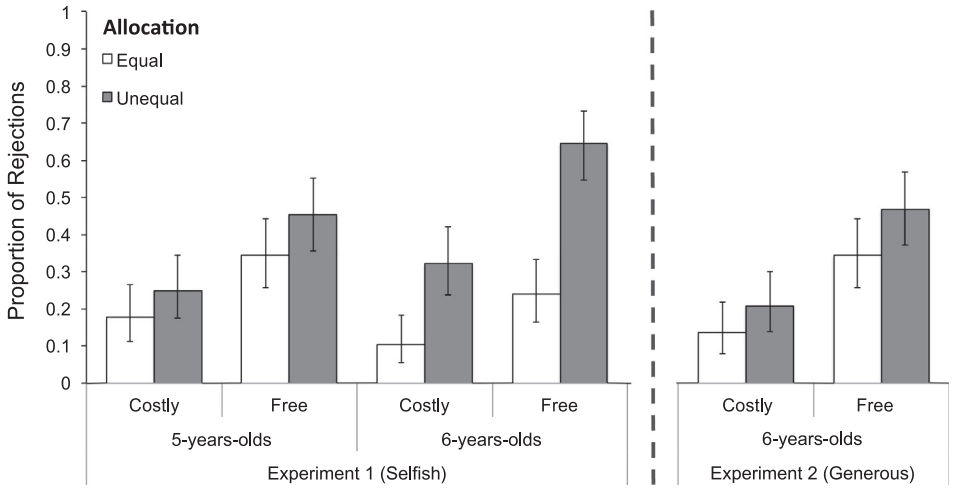

Fig. 2. Proportion of allocations rejected in Experiments 1 and 2. Subjects decided to accept or reject resource allocations on behalf of two absent children, a divider and a recipient. The divider had offered the recipient equal (3 for divider, 3 for recipient) and unequal allocations. Unequal allocations were either selfish (6 for divider, 0 for recipient, Experiment 1) or generous (0 for divider, 6 for recipient, Experiment 2). Participants made decisions about 12 trials. Error bars show confidence intervals.

interaction of age group and distribution. To unpack the interaction, we analyzed the two age groups separately.

A GLMM of 6-year-olds responses on the 12 trials showed two significant predictors of decisions: distribution (LRT, $\chi^2(1) = 46.9$, $p < .001$) and cost (LRT, $\chi^2(1) = 22.04$, $p < .001$; see Table 1 for model output). 6-year-olds were more willing to reject in the free condition ($M = 5.31$ trials, $SD = 0.87$) than the costly condition ($M = 2.56$ trials, $SD = 1.75$; $\beta = 1.22$, Odds Ratio = 3.39, Table 1). 6-year-olds also showed a strong sensitivity to inequality, rejecting more unequal allocations ($M = 2.91$ trials, $SD = 1.69$) than equal allocations ($M = 1.03$ trials, $SD = 1.03$; $\beta = 1.63$, Odd Ratio = 5.08, Table 1). Thus, 6-year-olds were sensitive both to cost and inequality, rejecting unequal allocations even when it was costly for them.

A GLMM of 5-year-olds participants' responses indicated that were two significant predictor's of children's decisions: cost (LRT, $\chi^2(1) = 4.39$, $p = 0.036$) and an interaction between distribution × gender (LRT, $\chi^2(1) = 7.13$, $p = 0.008$; see Table 1 for model output). These results suggest that 5-year-old participants were less likely to reject allocations when they had to sacrifice their own resources in order to reject (costly, $M = 2.56$ trials, $SD = 2.31$; free, $M = 4.75$ trials, $SD = 3.09$; $\beta = 1.16$, Odds Ratio = 3.21, Table 1). The interaction between distribution x gender was due to the finding that male participants in this age group showed stronger sensitivity to distribution, rejecting more selfish ($M = 2.38$ trials, $SD = 2.03$) than equal allocations ($M = 1.19$ trials, $SD = 1.42$), compared to female participants (equal, $M = 1.94$ trials, SD = 1.88; unequal,

Table 1
Estimate and standard error (s.e.) of fixed effects in Generalized Linear Mixed Models predicting children's rejection behavior in Experiments 1 and 2. Baselines were set as follows: Distribution = equal, Cost = costly, Gender = female, Inequity = generous). Table also shows goodness-of-fit statistics.

	Exp. 1		Exp. 2	Exp. 1 and 2
	Selfish, Age 5	Selfish, Age 6	Generous, Age 6	Selfish vs. Generous, Age 6
Intercept	−1.62 (0.50)**	−2.30 (0.26)***	−2.90 (0.56)***	−2.31 (0.32)***
Distribution	−0.09 (0.34)	1.63 (0.25)***	0.73 (0.28)**	0.64 (0.26)*
Cost	1.16 (0.53)*	1.22 (0.25)***	1.89 (0.71)**	1.46 (0.31)***
Gender	−0.89 (0.61)			
Distribution × Gender	1.38 (0.51)**			
Inequity				−0.30 (0.38)
Distribution × Inequity				1.13 (0.37)**
AIC	423.78	423.89	374.85	818.12
BIC	447.47	439.70	390.66	845.98
Log Likelihood	−205.89	−207.95	−183.43	−403.06
Deviance	411.78	415.89	366.85	806.12
Num. obs.	383	384	384	768
Num. groups: ID	32	32	32	64
Variance: ID (Intercept)	1.66	0.00	3.02	0.95

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

M = 1.81 trials, SD = 1.6; β = 1.38, Odds Ratio = 3.98, Table 1; see SM, Fig. S6).

When examining participant-level data, we found that a majority of participants engaged in punishment. Of our 64 participants, 54 punished at least once (84%). Of the 10 participants that never punished, 9 of them were 5-year-olds. Examining individual-level proportional data in 6-year-olds by cost and allocation type showed that all 16 children in the free condition punished at least one unequal allocation. Three of those 16 punished only unequal allocations, and never punished equal allocations. In the costly condition, 12 of the 16 participants punished at least one unequal allocation and eight of those 12 punished only unequal allocations.

2.3. Discussion

Results from Experiment 1 show both 5- and 6-year-old children were sensitive to the costs of intervention: children in both age groups were less likely to intervene when intervention required sacrificing one's own resources. Additionally, 6-year-olds were more likely to intervene when the divider proposed a selfish, unfair allocation than when the decided proposed an equal allocation. Indeed, 6-year-olds intervened regularly, rejecting almost half of all unequal allocations on average.

Unlike 6-year-olds, 5-year-olds did not show a clear sensitivity to distribution. This pattern of behavior, which we observed during pilot testing, contrasts with previous work that has demonstrated that 5-year-olds show a preference for equality in a two-party setting, even when enacting fairness comes at a personal cost (Blake & McAuliffe, 2011). As such, we wanted to ensure that 5-year-olds were not insensitive to distribution in our three-party task because they were confused about how to use our experimental apparatus to express their preferences. We addressed this concern in a follow-up to Experiment 1.

3. Follow-up to Experiment 1: Will 5-year-olds reject unequal allocations in a two-party game?

To ensure that participants did understand how to use the apparatus, we ran a short follow-up task with 5-year-old participants. In this task, participants used the apparatus to make decisions about reward allocations between themselves and an absent partner. In this task, the experimenter presented proposed allocations of candies between the participant and the absent partner. Our main question of interest was: would 5-year-olds show sensitivity to distributional inequality when their own resources were at stake?

3.1. Method

3.1.1. Participants

Participants in this follow-up task were the same 5-year-old children that participated in Experiment 1: 32 5-year-olds (M = 65.7 months, range = 61–70 months, 16 females, 16 males).

3.1.2. Design

Each participant was paired with an absent child (gender and age matched) and presented with the following three trials: equal (3 for child, 3 for partner); disadvantageously unequal (0 for child, 6 for partner) and advantageously unequal (6 for child, 0 for partner). The order in which trials were presented was counterbalanced across subjects and children were randomly assigned to receive one of the 6 possible trial orders. Children did not have to sacrifice candy before pulling the handle in the red direction. Thus, it was only costly for children to reject the equal and advantageously unequal allocations, for which they stood to gain something by accepting.

3.1.3. Procedure

After completing Experiment 1, five-year-old participants were told that they would play another game, in which they would make decisions for a third absent child (i.e. not one of the absent children from Experiment 1). As in Experiment 1, children were shown a paper bag on which the absent partner had allegedly drawn his or her face. Children were led to believe that their decisions would affect both their own allocations and the absent partner's allocations. Children were told that they would make decisions by pulling the handle in the green and red directions, as they had been doing in Experiment 1. Following this explanation, children received one of each of the three reward allocations. The procedure for test trials was similar to that used in Experiment 1 except that in this task, the experimenter made clear that the offers came from her, not the absent child. After the last trial, children were told that the game was over, and the post-decision comprehension questions described in Experiment 1 were administered.

3.1.4. Data coding and analysis

We used the same coding and analyses procedures as were used in Experiment 1.

3.2. Results

Participants in this task were more likely to reject disadvantageously unequal allocations than advantageously unequal and equal allocations (Fig. 3).

A GLMM of subjects' decisions showed a main effect of distribution (LRT, χ^2 (2) = 18.65, $p < .001$). Children were less likely to accept disadvantageous allocations than equal allocations but there was no difference between their likelihood of accepting advantageous allocations and equal allocations (β = 1.69, Odds Ratio = 5.44; see SM, Table S3 for model output).

3.3. Discussion

Findings from this follow-up task suggest that 5-year-old participants did understand how to use the apparatus used in Experiment 1 and were capable of using it to express fairness preferences. When we modified Experiment 1 such that the experimenter proposed allocations and the allocations directly affected the participant, 5-year-olds used our apparatus to systematically reject

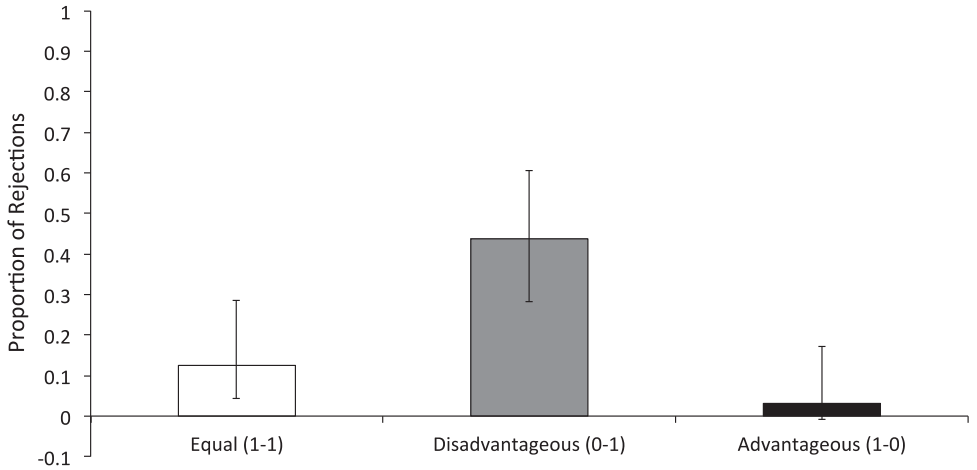

Fig. 3. Proportion of allocations rejected by 5-year-olds participants in the follow-up to Experiment 1. Subjects decided whether to accept or reject reward allocations that were to be divided between themselves and an absent partner. The allocations were equal (3 for subject, 3 for partner), disadvantageous (0 for subject, 6 for partner) and advantageous (6 for subject, 0 for partner). Error bars show confidence intervals.

disadvantageous allocations, but not equal or advantageous allocations. Thus, this follow-up experiment, together with the comprehension checks from Experiment 1, indicates that 5-year-olds' weak sensitivity to unequal resource allocations in third-party contexts does not result from a lack of understanding of the apparatus.

Taken together, results from Experiment 1 and the follow-up suggests that around the age of six, children begin to systematically punish unequal allocations, even when they are not personally affected by the inequity. Furthermore, children were sensitive to the costs of rejection, suggesting that the cost was sufficient to deter rejection, and that costly punishment in our task can be interpreted as a meaningful sacrifice.

One possible explanation for these results is that 6-year-olds' rejections were motivated by a specific desire to punish selfish actions or intentions (e.g. Cushman et al., 2009; Sanfey et al., 2003). Alternatively, 6-year-olds' rejections may have been motivated by a general inequality aversion (Dawes et al., 2007; Raihani & McAuliffe, 2012). If this were the case, they may have been averse to any form of distributional inequity and would reject any unequal offers, regardless of whether they are selfish or generous. In Experiment 2, we sought to investigate the extent to which punishment observed in Experiment 1 reflected each of these two motivations.

4. Experiment 2: Do 6-year-olds reject all unequal allocations or do they specifically reject selfish allocations?

In Experiment 1 we showed that children were willing to intervene on behalf of another child when resources were distributed unequally. However, this behavior could reflect a general aversion to unequal allocations, regardless of whether they resulted from selfishness or generosity. In Experiment 2 we tested this by reversing the direction of inequality: in this version of the game, 6-year-olds were given an opportunity to accept or reject equal allocations of candy or generously unequal allocations. The allocations were identical to those used in Experiment 1 except that the direction of inequality was reversed (0–6 offers; divider kept 0 and gave 6). Again, equal and unequal allocation trials were randomized with the constraint that no more than two of the same allocations could be presented consecutively. We only tested 6-year-olds because 5-year-olds had not shown a strong sensitivity to distribution in Experiment 1.

4.1. Methods

4.1.1. Participants

We tested 32 6-year-olds (M = 76.4 months, range 72–82 months, 16 females). None of these children participated in Experiment 1. Experiment 2 was run in parallel with Experiment 1 (May 2012 – September 2012). One additional subject was excluded because of an apparatus malfunction.

4.1.2. Design, procedure, data coding and analysis

The methodology for Experiment 2 was identical to that used in Experiment 1 with the exception that participants were tested with 6 trials of generous unequal allocations, where the divider gave six candies to the recipient and kept none for him or herself. We asked the same comprehension questions as in Experiment 1 (questions 1–4, see above). Virtually all participants answered comprehension questions either spontaneously correctly or correctly after additional questioning (100% to (1), 97% to (2), 100% to (3) and 100%, to (4)). Additionally, the majority of participants (66%) answered that they believed the other children were real and would be coming back to collect their Skittles (they confirmed their belief on one or both of the belief checks). We obtained the same results when controlling for children's answers to these questions (see SM). As in Experiment 1, decision data from Experiment 2 were live coded and double-checked with video coding. There were no disagreements between live and video coding.

4.2. Results

Results from Experiment 2 are shown in Fig. 2. As in Experiment 1, participants were more likely to reject allocations when doing so was cost-free ($M = 4.88$ trials, $SD = 3.30$) than when it was costly ($M = 2.06$ trials, $SD = 2.93$; $\beta = 1.89$, Odds Ratio = 6.60, Table 1). Additionally, they were more likely to reject unequal allocations ($M = 2.03$ trials, $SD = 1.99$) than equal allocations ($M = 1.44$ trials, $SD = 1.9$; $\beta = 0.73$, Odds Ratio = 2.09, Table 1), even though unequal allocations were generous. Our GLMM showed that participants' decisions were predicted by distribution (LRT, $\chi^2(1) = 6.91$, $p = 0.009$) and cost (LRT, $\chi^2(1) = 6.57$, $p = 0.01$; see Table 1 for model output). This suggests that 6-year-olds were motivated to reject unequal outcomes in a three-party context, even when they resulted from generosity rather than selfishness.

To test whether children were sensitive to the direction of inequality in addition to inequitable outcomes, we compared 6-year-olds' decisions in Experiments 1 and 2. To validate pooling of data from Experiment 1 and Experiment 2, we ensured that 6-year-old participants in Experiment 1 and 2 did not differ in their response to equal trials, which were the same (3 for divider, 3 for recipient) across both experiments. Participants were no more likely to reject equal trials in Experiment 2 than in Experiment 1 (LRT, $\chi^2(1) = 0.46$, $p = 0.5$), which suggests that the two groups of 6-year-old participants were comparable.

As Fig. 2 shows, children were less sensitive to distributional inequity in Experiment 2 (generous inequality) than Experiment 1 (selfish inequality; $\beta = 1.13$, Odds Ratio = 3.11, Table 1). Indeed, a GLMM of children's decisions as a function of distribution, condition and inequity type (selfish vs. generous) and all two-way interactions with distribution revealed that two terms were significant predictors of behavior: cost (LRT, $\chi^2(1) = 19.72$, $p < .001$) and an interaction between distribution x inequity type (LRT, $\chi^2(1) = 9.36$, $p = .002$; see Table 1 for model output).

As in Experiment 1, we examined individual-level proportional data in Experiment 2. We found that 13 of the 16 participants in the free condition punished at least one unequal allocation and four of those 13 punished only unequal allocations. In the costly condition, six of the 16 participants punished at least one unequal allocation and one of those six punished only unequal allocations.

4.3. Discussion

Across Experiments 1 and 2, children showed a main effect of cost: they were more likely to reject when it was free. Furthermore, children in both experiments were more likely to reject unequal than equal allocations, suggesting that children were averse to unequal outcomes even as third-party observers. However, children were more sensitive to inequity when it resulted from selfishness (Experiment 1) than when it resulted from generosity (Experiment 2), suggesting that punishment in Experiment 1did not result entirely from a straightforward aversion to all unequal outcomes. Rather, unequal outcomes were more objectionable when they resulted from selfishness than from generosity.

5. Conclusion

This is the first study to demonstrate that children engage in costly third-party punishment of unfair behavior. Children are sensitive to costs, as they punish less when punishment entails personal sacrifice. Importantly, despite the personal sacrifice, 6-year-old children systematically punish unfair behavior. Moreover, these children are sensitive to the direction of inequity: they are more likely to reject selfish unequal allocations (6–0) compared to generous unequal allocations (0–6). Thus, our results are the first to demonstrate the developmental origin of what has previously been demonstrated only in adults, namely costly enforcement of fairness in others.

Our results show that between the ages of 5 and 6, children develop a willingness to punish unequal allocations in a third-party context. A sensitivity to cost is present in 5-year-olds and appears to be stable across 5- and 6-year-olds. Moreover, our data show that 6-year-old children intervene to punish unfairness *despite* their clear sensitivity to costs. In conclusion, 6-year-olds are paying a cost to intervene against unfair behavior between two individuals, and are thus engaging in costly third-party punishment.

Results from Experiments 1 and 2 indicate that 6-year-olds show pluralistic motivations for third-party punishment. In line with work on adults (Dawes et al., 2007; Raihani & McAuliffe, 2012), 6-year-olds show an aversion to inequality. Participants were more likely to reject unequal allocations than equal allocations, regardless of whether the unequal outcomes resulted from generosity or selfishness. Thus, inequality aversion seems to account for some of their punishment behavior. However, children were also more sensitive to inequality that resulted from selfishness than generosity, suggesting inequity aversion and an aversion to selfishness as two separate motivations that work together to promote punishment.

One possible explanation for the finding that 6-year-oldsare sensitive to the direction of inequality is that 6-year-olds are sensitive to inferred intentions behind unequal offers and want to prevent selfish acts. This interpretation would be consistent with work on adults that shows that subjects are more likely to punish intentional unfairness than unintentional unfairness (e.g. Cushman et al., 2009; Sanfey et al., 2003). Future work could explore the extent to which children show sensitivity to intentionality by manipulating whether allocations were intentionally or randomly generated. However, until this work is done, we minimally claim that 6-year-old children show context sensitivity in their punishment behavior.

Results from these experiments show that children care about fairness even when they are not affected by unfair behavior. Thus, by the age of 6, children's show a desire to promote fairness in other children, even when their partners are anonymous, absent others. This desire emerges around the same time as a desire for fairness in a two-party context (Blake & McAuliffe, 2011; Shaw & Olson, 2012).

These results add to a growing body of work that suggests that children understand and intervene against norm violations (Killen et al., 2013; Schmidt and Sommerville, 2011; Smetana, 1983, 1984) by showing that they are also willing to pay a cost to punish unfair behavior.

Critically, our results cannot be explained by social comparison between the subject and the other children because subjects always had more Skittles in their box than either absent child. Moreover, rejections in our task could not benefit the subject through direct or indirect gains since the divider's allocations did not affect the subject (thus no direct benefits) and the children were anonymous others (thus no indirect, reputational benefits).

Unlike 6-year-olds, 5-year-olds did not reject more unequal allocations than equal allocations. However, they did reject some of both types of allocations. Rejections of equal allocations are slightly puzzling and were seen in both 5- and 6-year-olds participants. We suspect that rejections of equal allocations were due to motivations other than inequity aversion or distaste for selfishness. For example, children may have wanted to try accepting and rejecting to explore the different functions of the handle. Another possible motivation for their rejections is spite: they may have been incurring a cost to inflict a cost on another individual (Hamilton, 1970). However, while 5-year-olds did not systematically differentiate between equal and unequal allocations, 6-year-olds *were* more likely to punish unequal allocations. This clearly indicates that by age 6, punishment is preferentially directed at unfair behavior, an effect that cannot be attributed to spite, or a desire to explore the handles.

One feature of our study was that the unequal allocations used in Experiment 1 were entirely selfish (6 for actor, 0 for recipient). The rationale for this allocation was that punishment of selfishness would target only the allocator and not negatively affect the recipient. Future work could manipulate the degree of inequality of unfair allocations to investigate how unequal a proposed offer needs to be in order to be punished and whether children reject unequal allocations even if this involves negatively affecting the recipient. Additionally, it is worth noting that punishment in our task – even in the costly condition –did not require a large investment of the punisher's resources. In order to punish, children had to sacrifice one piece of candy each time. Our results suggest that even this relatively low cost was salient enough to reduce punishment in the costly condition compared to the free condition. However, future work could titrate the costs children are willing to pay to punish in this task. Doing so would shed light on the limits of children's willingness to invest in third-party punishment of fairness norm violations.

Children in the 5-year-old age group were sensitive to the costs of intervention but did now show a systematic willingness to intervene against unfairness. In this group, we did observe a gender effect, with boys being more likely to punish inequality than equality, while girls did not show this difference. This gender difference merits further research, but given mixed results on gender and social preferences (Balliet, Li, Macfarlan, & Van Vugt, 2011), we think it is first necessary to assess the robustness of this effect. Overall, we can conclude that 5-year-olds children did not show the same degree of third-party punishment of unfairness as 6-year-olds.

Our results suggest that costly third-party punishment is not clearly present until the age of six. This finding opens up the possibility to investigate this developmental transition in more detail. One possibility is that there is a general developmental effect, with children gradually becoming more likely to engage in costly punishment with age. In line with the possibility, various studies on children's sharing behavior show steady increase in children's tendency to divide resources according to an equality norm (Benenson, Pascoe, & Radmore, 2007; Blake & Rand, 2009; Gummerum, Hanoch, Keller, Parsons, & Hummel, 2010; Smith, Blake, & Harris, 2013), suggesting that by six years of age, children adhere to this norm so strongly that they enforce it on others as well. An alternative possibility is that there are specific psychological and social factors that account for its ontogenetic emergence. For example, it is possible that as children spend more time in peer groups, they have had more opportunity to negotiate norms and intervene against norm-violations (Killen & Smetana, 2006; Piaget, 1965).

In sum, our study suggests that from the age of six, children have an understanding of social norms against unfair behavior and that they are willing to pay personal costs to enforce fairness in others. This finding has important implications for our understanding of the role of costly third-party punishment in humans. Given that costly third-party punishment emerges relatively early in development, it is likely that third party punishment serves a role in promoting fair behavior in peer interactions. This suggests that costly third-party punishment is an integral part of children's nascent sense of fairness.

Acknowledgements

We thank the research assistants who contributed to this project and Scott Mackie for his help with the apparatus design. We thank Josh Buckholtz, Josh Greene, Nichola Raihani and three anonymous reviewers for helpful comments on an earlier version of this manuscript. We are also grateful for financial support from the Harvard University Mind/Brain/Behavior Interfaculty Initiative, the Harvard University BLISS Fellowship Program, the Herchel Smith Harvard Undergraduate Science Research Program and the Science of Generosity Initiative of the John Templeton Foundation.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2014.08.013.

References

Agresti, A., & Coull, B. A. (2010). Approximate is better than "exact" for interval estimation of binomial proportions. *American Statistician, 52*(2), 119–126.

Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin, 137*(6), 881–909. http://dx.doi.org/10.1037/a0025354.

Bates, D., Maechler, M., Bolker, B., 2012. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0, http://CRAN.R-project.org/package=lme4.
Baumard, N., Andre, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences, 36*, 59–122.
Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and Human Behavior, 28*(3), 168–175. http://dx.doi.org/10.1016/j.evolhumbehav.2006.10.003.
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. http://dx.doi.org/10.1016/j.cognition.2011.04.006.
Blake, P. R., & Rand, D. G. (2009). Currency value moderates equity preference among young children. *Evolution and Human Behavior, 31*(3), 210–218. http://dx.doi.org/10.1016/j.evolhumbehav.2009.06.012.
Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., et al. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution, 24*(3), 127–135. http://dx.doi.org/10.1016/j.tree.2008.10.008.
Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., et al. (2008). The neural correlates of third-party punishment. *Neuron, 60*(5), 930–940. http://dx.doi.org/10.1016/j.neuron.2008.10.016.
Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience, 15*(5), 655–661. http://dx.doi.org/10.1038/nn.3087.
Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *PLoS ONE, 4*(8), e6699.
Damon, W. (1977). *The social world of the child*. London, England: Jossey-Bass.
Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature, 446*(7137), 794–796. http://dx.doi.org/10.1038/nature05651.
Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry, 41*(1), 20–26.
Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior, 62*(1), 287–303.
Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature, 454*(7208), 1079–1083. http://dx.doi.org/10.1038/nature07155.
Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63–87. http://dx.doi.org/10.1016/S1090-5138(04)00005-4.
Fehr, E., Fischbacher, U., & Gächter, S. (2003). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature, 13*(1), 1–25.
Gardner, A., & West, S. A. (2004). Cooperation and punishment, especially in humans. *The American Naturalist, 164*(6), 753–764. http://dx.doi.org/10.1086/425623.
Gummerum, M., Hanoch, Y., Keller, M., Parsons, K., & Hummel, A. (2010). Preschoolers' allocations in the dictator game: The role of moral emotions. *Journal of Economic Psychology, 31*(1), 25–34. http://dx.doi.org/10.1016/j.joep.2009.09.002.
Hamann, K., Warneken, F., Greenberg, J. R., & Tomasello, M. (2011). Collaboration encourages equal sharing in children but not in chimpanzees. *Nature, 476*(7360), 328–331. http://dx.doi.org/10.1038/nature10278.
Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature, 228*(5277), 1218–1220.
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. http://dx.doi.org/10.1073/pnas.1110306108.
Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., et al. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science, 327*(5972), 1480–1484. http://dx.doi.org/10.1126/science.1182238.
Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., et al. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767–1770. http://dx.doi.org/10.1126/science.1127333.
Karniol, R. (1978). Children's use of intention cues in evaluating behavior. *Psychological Bulletin, 85*(1), 76–85.
Kenward, B., & Östh, T. (2012). Enactment of third-party punishment by 4-year-olds. *Frontiers in Psychology, 3*, 1–9. http://dx.doi.org/10.3389/fpsyg.2012.00373.
Killen, M., Rutland, A., Abrams, D., Mulvey, K. L., & Hitti, A. (2013). Development of intra- and intergroup judgments in the context of moral and social-conventional norms. *Child Development, 84*(3), 1063–1080.
Killen, M., & Smetana, J. G. (Eds.). (2006). *Handbook of moral development*. Mahwah, NJ: Lawrence Erlbaum Associates.
Krasnow, M. M., Cosmides, L., Pedersen, E. J., & Tooby, J. (2012). What are punishment and reputation for? *PLoS ONE, 7*(9), e45662.
Lehmann, L., & Keller, L. (2006). The evolution of cooperation and altruism – A general framework and a classification of models. *Journal of Evolutionary Biology, 19*(5), 1365–1376.
Leman, P. J., & Björnberg, M. (2010). Conversation, development, and gender: A study of changes in children's concepts of punishment. *Child Development, 81*(3), 958–971.
LoBue, V., Nishida, T., Chiong, C., DeLoache, J. S., & Haidt, J. (2010). When getting something good is bad: Even three-year-olds react to inequality. *Social Development, 20*(1), 154–170. http://dx.doi.org/10.1111/j.1467-9507.2009.00560.x.
Masters, J. C., & Furman, W. (1981). Popularity, individual friendship selection, and specific peer interaction among children. *Developmental Psychology, 17*(3), 344–350.
Nelissen, R. M., & Zeelenberg, M. (2009). Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions. *Judgment and Decision Making, 4*(7), 543–553.
Nelson, W. R. (2002). Equity or intention: It is the thought that counts. *Journal of Economic Behavior and Organization, 48*(4), 423–430.
Piaget, J. (1965). *The moral judgment of the child*. New York: Free Press.
R Development Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Raihani, N. J., & McAuliffe, K. (2012). Human punishment is motivated by inequity aversion, not a desire for reciprocity. *Biology Letters, 8*(5), 802–804. http://dx.doi.org/10.1038/nature06723.
Raihani, N. J., Thornton, A., & Bshary, R. (2012). Punishment and cooperation in nature. *Trends in Ecology and Evolution, 27*(5), 288–295. http://dx.doi.org/10.1016/j.tree.2011.12.004.
Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology, 44*(3), 875–881.
Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees. *Proceedings of the National Academy of Sciences, 109*(37), 14824–14829. http://dx.doi.org/10.1073/pnas.1203179109/-/DCSupplemental.
Robbins, E., & Rochat, P. (2011). Emerging signs of strong reciprocity in human ontogeny. *Frontiers in Psychology, 2*. http://dx.doi.org/10.3389/fpsyg.2011.00353/abstract.
Rogoff, B. (2003). *The cultural nature of human development*. New York: Oxford University Press.
Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 300*(5626), 1755–1758. http://dx.doi.org/10.1126/science.1082976.
Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition, 124*(3), 325–333. http://dx.doi.org/10.1016/j.cognition.2012.06.004.
Schmidt, M. F. H., & Sommerville, J. A. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS ONE, 6*(10), 1–7. http://dx.doi.org/10.1371/journal.pone.0023223.
Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General, 141*(2), 382–395. http://dx.doi.org/10.1037/a0025907.
Smetana, J. G. (1983). Social-cognitive development: Domain distinctions and coordinations. *Developmental Review, 3*(2), 131–147.
Smetana, J. G. (1984). Toddlers' social interactions regarding moral and conventional transgressions. *Child Development*, 1767–1776.
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. http://dx.doi.org/10.1371/journal.pone.0059510.
Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Reviews of Psychology, 64*, 231–255.
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. http://dx.doi.org/10.1348/026151010X532888.
Warneken, F., Lohse, K., Melis, A. P., & Tomasello, M. (2011). Young children share the spoils after collaboration. *Psychological Science, 22*(2), 267–273.
Wyman, E., Rakoczy, H., & Tomasello, M. (2009). Normativity and context in young children's pretend play. *Cognitive Development, 24*(2), 146–155.