Abstract

Gene-culture coevolutionary theorists argue that cultural evolution has harnessed various aspects of our evolved psychology to create a variety of different mechanisms for sustaining social norms, including those related to large-scale cooperation. One of these mechanisms, costly punishment, has emerged in experiments as an effective means to sustain cooperation in some societies. If this view is correct, individuals' willingness to engage in the costly punishment of norm violators should be culturally-transmittable, and applicable to both prosocial and anti-social behaviors (to any social norm). Since much existing work shows that norm-based prosocial behavior in experiments develops substantially during early and middle childhood, we tested 245 3- to 8-year olds in a simplified Third Party Punishment Game to investigate whether children would imitate a model's decision to punish at a personal cost, both unequal and equal offers. Our study showed that children, regardless of their age, imitate the costly punishment of both equal and unequal offers, and the rates of imitation increase (not decrease) with age. However, only older children imitate not-punishing for both equal and unequal offers. These findings highlight the potential role of cultural transmission in the stabilization or de-stabilization of costly punishment in a population.

1. Introduction

Sustaining cooperation in large groups among nonrelatives requires mechanisms to suppress free-riding. Examples of real world cooperative dilemmas that are susceptible to free-riding include warfare, irrigation, voting, cooperative hunting, fishery management, paying taxes, neighborhood watch and recycling. In all these situations individuals have to make choices about how much to contribute, if at all. In the end, individuals' payoffs are affected by both their own decisions and those of others. Figuring out how human societies have surmounted this evolutionary and economic challenge stands as a central problem in the human sciences.

Evolutionary theorists have proposed a number of solutions to the dilemma of large-scale cooperation in humans, based on mechanisms involving indirect reciprocity, signaling and punishment, among others (Boyd et al., 2010; Fu et al., 2008; Hauert et al., 2007; Milinski et al., 2002). In particular, some have proposed that cultural evolution can solve the dilemma of large-scale cooperation by harnessing elements of our evolved social psychology (Chudek et al., 2013; Henrich et al., 2010). These theorists have proposed a number of candidate mechanisms ranging from those that extend indirect reciprocity (Panchanathan & Boyd, 2004), to those that exploit peculiarities of our cultural learning psychology (Henrich, 2009). This approach suggests that humans possess an evolved norm-psychology that enables us to readily acquire and adhere to local norms as well as to respond to the norm violations of others (Chudek & Henrich, 2011; Henrich & Henrich, 2007), while recognizing that norms may be sustained by quite different mechanisms in different societies. Field and experimental evidence has already begun to indicate that different norm-sustaining mechanisms have emerged and operate in different places (Ensminger & Henrich, in press; Herrmann et al., 2008; Wiessner, 2005).

For the largest scale human societies, a subset of these models, those involving costly punishment, are of particular interest for three reasons (1) the threat of counter-punishment, which is greatest in small-scale societies, is reduced in larger, more anonymous societies, (2) indirect reciprocity mechanisms weaken as the size of a population increases making them much less effective in larger societies, and (3) empirically, larger communities are willing to engage in costly punishment much more readily than smaller communities (Henrich et al., 2006, 2010; Marlowe et al., 2008). Notably, in these models, the cultural evolution of costly punishment is silent on the proximate, personal motivations for punishing. That is, in favoring large-scale cooperation, cultural evolution may harness, extend and exploit various aspects of our evolved psychology, including concerns about revenge, signaling, reciprocity and reputation. By tapping these pieces of evolved psychology, cultural evolution can more effectively sustain social norms in a manner analogous to the way that learning to read has evolved culturally to exploit various features of our capacities for spoken language and object recognition[1] (Dehaene, 2009).

Of course, costly punishment creates a second-order free-rider problem which arises from the costs associated with punishment. Cultural evolutionary models involving costly punishments have studied a number of mechanisms that can overcome this problem, under some conditions. These mechanisms include recursive punishing strategies (Axelrod, 1986; Boyd & Richerson, 1992), conformist cultural learning (Boyd & Richerson, 1985), signaling (Boyd et al., 2010; Gintis et al., 2001)

[1] Focusing too narrowly on these more proximate motivations and psychological mechanisms for explaining costly punishment misses how diverse institutions (sets of social norms) have shaped and differentially deployed these (see Henrich et. al. (2012) for an illustrative example focus on the spread of normative monogamous marriage).

and inter-group competition (Boyd et al., 2003, 2011), as well as potent combinations of two or more of these mechanisms (Guzman et al., 2007; Henrich & Boyd, 2001).

Central to this approach to cooperation and social norms more generally, is whether social behaviors are indeed culturally transmitted at all. There is already evidence that cultural learning influences helping and cooperating in both children and adults (Fowler & Christakis, 2010; Henrich & Henrich, 2007; Trommsdorff et al., 2007). Research on children reveals not only sophisticated early cultural learning of social rules (Tomasello et al., 2005), but by age three children seem to understand the normative dimensions of property rights (Rossano et al., 2011) and protest antisocial acts (Vaish et al., 2011) or automatically enforce arbitrary rules that they have just acquired (Rakoczy et al., 2009; Schmidt et al., 2012; Wyman et al., 2009). Consistent with this, different societies reveal distinct developmental trajectories on costly sharing in a study of children aged 3-14 in six diverse societies, including foragers, pastoralists, horticulturalists and American children (House et al., 2013). Notably, these population-level differences in development emerged during middle childhood (age 6-9), just when norm adherence dramatically increases in children (Smith et al., 2013).

In contrast with the development of behaviors on the positive side of prosociality—fairness, sharing and helping (for a recent review, see Tomasello & Vaish, 2013)—there has been relatively little work on the acquisition of costly punishment or other sanctioning behaviors. Recently, by means of a three-way sharing game with puppets, Robbins and Rochat (2011) have shown that as opposed to 3 year olds, 5 year-old American children, but not Samoan children, selectively punish stingy offers. In younger children (19-24 month-olds), Hamlin et al. (2011)

show that children directed positive behavior (giving a treat) towards pro-social puppets and negative behavior (taking a treat from) towards anti-social ones. However, no empirical work thus far has explored the development of children's willingness to engage in costly punishment, of either equitable or inequitable divisions, and how this might be influenced—magnified or suppressed—by social learning opportunities.

To explore this, we developed and deployed a simplified version of the Third Party Punishment Game, which has emerged as the simplest and most widely used experimental task for measuring individuals' willingness to pay a cost to inflict punishment on another individual during allocation tasks. The standard version of the game is played by three players: Player 1, Player 2 and Player 3. Player 1 (the "dictator") is given a certain amount of money and must decide how much to allocate to Player 2 (the "receiver"). Player 3 (the "third-party") also receives an endowment and decides whether to pay a certain amount of his or her allocation to punish Player 1 for his/her offer (Fehr & Fischbacher, 2004a). Experiments played with undergraduates in industrialized societies have confirmed that (1) third parties engage in costly punishment in anonymous situations when the norm of egalitarian distribution is violated, and (2) that the threat of punishment in laboratory games often raises the levels of cooperation and increase long-term payoffs (Fehr & Fischbacher, 2004b; Fehr & Gächter, 2000, 2002; Fischbacher et al., 2001; Gatcher, 2012; Henrich et al., 2006). However, as noted above, individual's willingness to engage in third party punishment varies across societies, from zero in some societies to a strong willingness to punish selfishness deviations from equality in others (Henrich et al., 2006, 2010; Marlowe et al., 2008).

The present study directly tests for the existence of costly third party punishment and the imitative learning of costly punishment in children between 3 and 8 years of age. By using an experimental set-up in which groups of children are given instructions via visual demonstrations of a third party punishment game played with stickers, we were able to measure the rates of imitation in punishment behavior. In particular, we seek answers to the following hypotheses:

1) Sharing and some foundational aspects of fairness appear early in development, but understanding the enforcement of social norms, which emerges later, may shape and modify these prosocial tendencies (Tomasello & Vaish, 2013). Therefore, in the presence of sanctioning norms in a society, the frequency of the costly punishment of unequal offers should increase with age as children acquire and internalize these norms. Older children may be less affected by imitative cues of punishment because they have already internalized the local sanctioning norms.
2) However, sensitive to learning even arbitrary norms in new social contexts, children will imitate both the punishment and the non-punishment of both equal and unequal offers from their cultural models, assuming these models possess the relevant social learning preferences, such as those based on age and sex (Chudek et al., 2013).
3) Nevertheless, since eons of cultural evolution may have made egalitarian norms recurrent features of human societies that subsequently shaped our innate repertoire, there may be an asymmetry between imitating the punishment of equal vs. unequal offers among same-sex peers (more imitation of punishment for equal than unequal offers).

2. Study 1

1.1 Methods

1.1.1 Participants

Two hundred and forty-five children (117 females) between 30 and 107 months of age (ages < 60 months: N = 88, mean±SD = 48.5±7.2; 60 ≤ ages < 84: N = 77, mean±SD = 70.9±6.7; ages ≥ 84: N = 80, mean±SD = 93.3 ± 7.2) were tested at Science World in Vancouver, Canada. During recruiting, the parents were told that the researchers were playing a sticker game and that participation was voluntary. The parents that consented to participate were taken to the *Living Lab* where they were able to watch their children from a screen outside the testing room. Information regarding the birth date, language spoken at home, number and age of siblings and gender of the participants was recorded. Parents were given more information about the study while their children were being tested. The majority of participants were of European or Asian descent.

1.1.2 Procedure

The experiment involved two phases, an *observation phase* and a *test phase*. During the observation phase participants watched a video clip that illustrated how to play the game. During the test phase the participant made his/her decision.

Observation phase: Children watched a video in which an adult administers the experiment (the game) to a child, (hereafter referred to as the “model”). In the video, the model was asked to play a game with stickers and given 6 stickers. Then, the administrator consecutively showed the model two pictures of the game being played by two children (electronic supplementary material, Fig. S1, available on the journal’s website at www.ehbonline). In the first picture, two children, allocated to the role of dictator and receiver, were sitting at a table facing each other. The dictator had 4 stickers laid out on his/her side of the table. In the second picture, the model saw the outcome of the game (i.e. the decision of the dictator on whether to share his/her stickers with the receiver). There were two possible outcomes: equal distribution (2-2) or unequal distribution (4-0). After letting the model examine the number of stickers on each side, the administrator asked him/her how many stickers each player had, and whether s/he thought this was “fair”. The model either replied that it was “fair” or “unfair” according to the condition. The administrator subsequently asked the model if s/he would like to give one of her/his own stickers away to make the dictator lose two stickers (note, the word “punishment” was not used). Depending on the condition, the model either decided to punish or not to punish. If s/he punished, s/he physically gave one of his/her stickers to the administrator and the dictator lost 2 of her/his stickers. In the control condition, the video was stopped just before the model’s punishment decision and participants were told that there was a glitch with the video. The video clip (i.e. the observation phase) lasted about 1 minute and 40 seconds for the equal and unequal conditions, and about 1 minute and 30 seconds for the control conditions.

To focus more precisely on the effects of condition, we used a male model (an average-sized 10 year old, Japanese-European mix) with male participants and a female model (a tall 7 year old, European descent) with female participants. This mitigates the complexities introduced by evidence suggesting that children may tend to preferentially copy models who match their own sex (Henrich & Gil-White, 2001; Shutts et al., 2010). We did not match on ethnicity, but control for this in our analyses.

Test Phase: During the test phase the participant took the role of the model and was then given 6 stickers, like the model. The experimenter, taking the role of the administrator, then showed two pictures of two children (different from the ones shown on the video) playing the same game. Note, the experimenter and the administrator were not the same person. As in the video, the first picture showed the initial condition of the game, and the second revealed the outcome. The outcome of the game and the cost of punishment was the same as that shown in the observation phase. The experimenter instructed the participant to examine the outcome of the game and asked how many stickers each player had. Then the experimenter asked whether the outcome of the game was "fair". It was then proposed that the experimenter would take two stickers from the dictator of the game, shown in the picture, if the participant gave up one of his/her own stickers (see electronic supplementary material for the script of the procedure).

1.1.3 Conditions and treatments

Each child experienced one of six different combinations of conditions and treatments. Conditions were the dictator's distribution (Equal vs. Unequal) and treatments were the model's decision to punish (Control vs. No Punishment vs. Punishment). The offer that each participant could respond to in the test phase was the same as that which they observed in the observation phase. Outcomes of each condition and treatment combination are presented in Table 1. The physical position of the dictator in the pictures was counterbalanced (left vs. right) across participants.

1.1.4 Statistical analysis

Since our response variable was binary (punish/don't punish), we used multiple logistic regression procedures in R 2.15.3. Five independent variables were used to predict our binary punish/don't punish outcome variable: *Age* (in months), *Gender* (female, male), *Treatment* (control, no punishment, punishment), *Order* (position of the dictator on the picture: left, right), and *Ethnicity* (of the participant: European, Asian, Other). We did a separate analysis investigating effects of having siblings (electronic supplementary material, Tables S1 and S2).

We applied logistic regression analyses to each condition separately: (1) only the Equal Condition and (2) only the Unequal Condition. To evaluate how well different variables explained our data, we created models containing all the predictors. We then found the subset of the five predictors along with the significant interaction terms that led to optimal models (i.e. most parsimonious models) for each regression. We removed the non-significant variables based on the likelihood ratio test statistic and its associated *p*-value. We report our results by using the regression coefficients with their standard errors, some of which we have converted to odds ratios (OR) in the main text to make them easier to interpret.

1.2 Results

1.2.1 Does seeing a model engage in either costly punishment or no costly punishment influence the likelihood of punishment?

Yes, as shown in Figure 1, participants' rates of punishment increased when they observed the model punishing (Punishment Treatment (PT)) compared to the Control where the model's decision was unknown. However, our degree of certainty that these are different depends on the age of participants and varies across our conditions. First, we'll discuss the results for the Equal Condition and then the Unequal Condition.

For the Equal Condition, Table 2 shows the results of our regression analysis. Here, we present the initial model (Model 1) with all our predictors and the optimal model (Model 2). Controlling for age, ethnicity and other variables, when participants observe the model punish they are 6.75 times more likely to punish ($CI_{.95}$ = [2.35-21.75]) compared to the Control Treatment (CT). By contrast, when participants in the Equal Condition observe a model who does not punish, they are 7.14 times more likely not to punish (OR = 0.14, $CI_{.95}$ = [0.04-0.45]) relative to the CT. Model 2 reveals similar patterns with the maximum likelihood estimated odds ratios converging at 5.83 ($CI_{.95}$ = 2.14-17.48]) and 5.55 (OR = 0.18, $CI_{.95}$ = 0.05-0.55]), respectively. Interestingly, the best model includes only the punishment treatments and the right-left position of the dictator (which were already counterbalanced), but does not include interactions between *Age* and *Treatment*. Visually, Figure 2A illustrates these relationships.

In the Unequal Condition, as shown in Figure 2B, the model's influence on participants' decisions was much more dependent on age. Older children copied the model more when the model was not engaging in costly punishment. The effect shown visually in Figure 2B, emerges as the interaction of our No-Punishment Treatment (NPT) and Age in Table 3. The *NPT x Age* interaction term (Table 3, Models 2 and 3) revealed a roughly 4-fold (OR = 0.24, $CI_{.95}$ = [0.11-0.50]) decrease in the odds of punishment for a 1 year increase in age for the participants in NPT compared to the CT. The green line in Figure 2B illustrates this effect.

The probability of punishing is also increased by opportunities to observe a model punishing, but the ability to detect these effects weakens with age as punishment increases in the control condition with Age. Nevertheless, the confidence intervals in Figure 2B show that for the youngest children the probability of punishing is quite different in the PT vs. the CT.

1.2.2 What is the developmental trajectory for imitation of the Model's decision for punishment and costly punishment?

As just noted, the Equal and Unequal Conditions create distinct age trajectories. For unequal distributions, children's decision was influenced by both their age and an interaction between their age and the availability of cultural transmission. In our control experiment, Figure 2B shows how the likelihood of punishing increases with age. When participants observed a model punishing, the entire curve shifts up, and children at every age punish more, on-average. However, when children observed a model refraining from punishment, older children punish much less. Model 3 in Table 3 shows these effects as they emerge from our multivariate logistic regression models, as the coefficients on *Age* and *Treatment x Age* interaction were significant predictors. Using Model 3, an increase of one year in the CT predicts an increase of 2.56 ($CI_{.95}$ = [1.37–4.78]) in the odds of punishment. Similarly, a 1 year increase in age in the PT predicts a 1.87 ($CI_{.95}$ = [1.00–3.52]) increase in the odds of punishment. By contrast, a one year increase in the NPT predicts a 1.63 (OR = 0.61, $CI_{.95}$ = [0.40–0.94]) decrease in the likelihood of punishment.

For equal distributions, Figure 2A shows that children's age did not strongly affect their decision to punish, except perhaps for small (non-significant) declines under the CT and NPT, and removal of *Age* from the regression model did not diminish the model's predictive power (Table 2). Thus, of our theoretically relevant predictor variables (Age and Treatment), it was the treatment effect—cultural transmission opportunities—that dominated.

Figure 2B shows that younger children did not selectively engage in costly punishment for unequal offers in the CT. To investigate whether there exist age differences in children's understanding of what's fair we analyzed the number and age of the participants who failed to answer the question on whether the outcome of the game was fair ("Do you think this is fair") in a manner consistent with local social norms (in this society, equal allocations are considered "fair" while unequal allocation are "unfair"). Not surprisingly, two-thirds of the 24 children who failed to answer the fairness question in a manner consistent with local norms were younger than 5 years old (electronic supplementary material, Figs. S2 and S3). However, neither controlling for the responses given to the fairness question nor excluding the non-norm compliant responses affected the outcomes of the regression models (electronic supplementary material, Tables S3-6).

We analyzed the age differences in the Equal versus the Unequal CTs to further investigate the development of children's internalized reactions to norm violations and their tendency in engaging in selective costly punishment (i.e. punishing unequal offers) in the absence of a model. As shown in Figure 3A, 3-4 year-olds did not significantly differ in their proportion of punishment for the Equal and Unequal Control Treatments (mean proportion for Equal: 0.40, for Unequal: 0.23; Pearson $\chi^2(1, n = 28) = 0.30$, $P = 0.71$). Although not significant, 5-6 year-olds engaged in 25% more costly punishment in the Unequal Condition than in the Equal one (mean proportion for Equal: 0.50, for Unequal: 0.75; Pearson $\chi^2(1, n = 24) = 0.71$, $P = 0.20$). Finally, 7-8 year-olds engaged in significantly higher levels of costly punishment in the Unequal Control Treatment than in the Equal Control one (mean proportion for Equal: 0.36, for Unequal: 0.83; Pearson $\chi^2(1, n = 26) = 4.21$, $P < 0.05$). These results indicate that selective costly punishment of unequal distributions emerges not earlier than the age of 5 in this society (see electronic supplementary material Fig. S4, Tables S7 and S8 for the analysis of the subset of the data where participants were older than 60 months).

2 Study 2

We designed Study 1 to generate a substantial amount of punishment across all conditions in order to avoid "floor effects" that might prevent us from observing certain treatment effects. For example, if children had been completely unwilling to punish in the Equal-Control, we could not have shown that observing no punishment (NPT) reduces punishing. However, this created an asymmetry between "taking an action" (punishing) vs. "not taking an action" that may have distorted our results in some way. To address this, we modified the Control Treatment of Study 1 by replacing "not punishing" with a hand clap option for the participants, so that they can choose between two actions: punishing or clapping their hands.

2.1 Methods

2.1.1 Participants

Eighty-nine children (45 females) between 31 and 111 months of age (ages < 60 months: $N = 36$, mean±SD = 49.03±6.83; 60 ≤ ages < 84: $N = 36$, mean±SD = 70.47±6.63; ages ≥ 84: $N = 17$, mean±SD = 94.71±9.57) were tested at the Science World in Vancouver, Canada. Information regarding the birth date, language spoken at home, number and age of siblings and gender of the participants was recorded. Parents were given more information about the study while their children were being tested. The majority of participants were of European or Asian descent.

2.1.2 Procedure

As in Study 1, the experiment involved an *observation* and a *test* phase. During the observation phase the participant watched a video clip showing an illustration of the game as instructions, and played the actual game in the test phase. Only the Control Treatment was applied in Study 2.

The only change from Study 1 was made in the test phase. In Study 1, the participants were given an option to punish with the following question: "If you give me one of your stickers, I will take away 2 stickers from Jane. Would you like to give me one of your stickers so that Jane loses two stickers or would you like to keep all your stickers?” In this study, we introduced a second, neutral option (clapping hands), so that participants could choose between giving away one of their stickers or clapping their hands. The prompt for punishment was changed as follows: “Now to end this game you can either clap your hands, or give me one of your stickers so that I will take away 2 stickers from Jane. Would you like to give away one of your stickers so that Jane loses 2 stickers, or would you like to clap your hands?” The rest of the procedure during the test phase was identical to Study 1.

2.1.3 Conditions:

Equal and Unequal Control Treatments were used for a comparison of the baseline levels of punishment between Study 1 and Study 2. The position of the dictator in the pictures (whether the dictator was sitting on the left or on the right side of the table) was again counter-balanced across participants.

2.1.4 Statistical analysis:

We used proportion test (a chi-square test with continuity correction) in R 2.15.3 to investigate whether the levels of punishment in the Control Conditions differed from those in Study 1. To investigate the effects of age on the odds of punishment, we used logistic regression models where the age was our primary predictor; and the number of siblings, gender and the position of the dictator on the pictures were our control variables.

2.2 Results

There was no significant change in the levels of punishment across Control Treatments of Study 1 and 2 (Fig. 3). Although the level of punishment in the Equal

Condition of Study 2 was higher than that of Study 1, the difference was not significant at conventional levels (mean proportion of punishment in Study 1: 0.41, Study 2: 0.55; Pearson $\chi^2(1, n = 88) = 1.174$, $P = 0.279$). The results were the same for the Unequal Conditions (mean proportion of punishment in Study 1: 0.59, Study 2: 0.76; Pearson $\chi^2(1, n = 79) = 1.831$, $P = 0.176$). If anything, the clapping option seems to increase punishing relative to doing nothing.

As in Study 1, Figure 3 and Table 4 shows that participants' age did not have an effect on their odds of punishment in the Equal Condition ($P = 0.46$). However, as shown in Figure 3 and Table 5, older children in the Unequal Condition punished slightly more than the younger ones ($P = 0.06$). Accordingly, a 12-month increase in age increased the odds of punishment by 2.1 times ($CI_{.95} = [0.97 – 12.43]$).

3 Discussion

Our results indicate that costly punishment can be acquired via social learning. First, we found that selective costly punishment of unequal offers does not emerge until 5 years of age. Second, children, regardless of their age, readily imitated costly punishment in either the Equal or Unequal conditions, though these effects are not symmetric. Third, children get more imitative as they get older, especially when the model does not punish. These findings suggest that sanctioning norms can be transmitted via cultural learning, which can lead to the acquisition and maintenance of norms for both prosocial and anti-social behaviors. Below, we will discuss each of these findings in light of recent relevant research investigating children's prosocial and imitative behaviors. Then, we will address potential concerns with our experimental design.

We will start by discussing our investigation of the developmental trajectory of costly punishment by comparing the Equal and Unequal Control Treatments, where there were no opportunities for social learning. We examined the emergence of children's willingness to engage in the selective costly punishment of unequal offers. We predicted that the frequency of costly punishment of unequal offers should increase with age as children acquire and internalize the local sanctioning norms in their society. Consistent with this prediction, our results suggest that greater (costly) punishment of unequal offers starts developing at 5 years of age, and by the age of 7, children selectively punish unequal offers. Our results are consistent with Robbins and Rochat's (2011) result that 5 year-olds but not 3-year olds in America selectively punish unequal distributions. Further work on the development of fairness in children indicates that whereas 3-year-olds tend to be selfish in their distributions, children begin showing tendencies towards equitable distribution at 5 to 7 years of age (Benenson et al., 2007; Blake & Rand, 2010; Fehr et al., 2008; Gummerum et al., 2010; Rochat et al., 2009).

Recent findings, however, suggest an early awareness of equal vs. unequal distributions. For example, 15-month old infants have been shown to have expectations regarding the equal distribution of resources and 16 month olds expect a recipient to approach an equal rather than unequal distributor (Geraci & Surian, 2011; Schmidt & Sommerville, 2011). Moreover, Hamlin et al. (2011) found that 8-month-old infants selectively prefer characters (stuffed animals) who act positively toward prosocial individuals and characters who act negatively towards antisocial ones. Their results also indicate that young toddlers (19-24 months) direct positive behaviors (giving treats) toward prosocial others and negative behaviors (taking treats) toward antisocial others.

Studies on 8 to 16 month-old infants show us that infants rapidly develop certain intuitions or expectations that no doubt influence their subsequent learning and development. However, it would be a mistake to immediately project these findings, via a straight line, onto adult behavior or even the behavior of older children. Recent work on costly sharing with children aged 3 and up across diverse societies shows that, first, young children become *less prosocial* (from age 3 to middle childhood), and then at some point in middle childhood, start becoming more prosocial (egalitarian) and diverge towards the behaviors of adults in their own societies (House et al., 2013; also see Trommsdorff et al., 2007).

Our findings here support the view that cultural learning builds on existing aspects of an evolved social psychology, as children can readily acquire social norms against, and tastes for punishing, equal distributions via cultural transmission (Chudek & Henrich, 2011; Andres et al., 2007; Henrich & Boyd, 2001). Not only does this work show that costly punishment is culturally transmittable, but the finding supports the implication that both prosocial and non-prosocial behaviors can be readily acquired via cultural transmission. Moreover, our results are relevant to a recent finding that even when children can verbally endorse social rule, they only begin to comply with these rules in middle childhood (Smith et al., 2013). This was evident in the finding that although 60 to 80 percent of 3 to 5 year-old children answered the fairness question (regarding the distribution of stickers) in a manner consistent with local norms, they did not punish unequal offers as often as older children did in the Control Treatment (electronic supplementary material, Fig. S3B).

One reason why we observed cultural learning for both equal and unequal offers could be that children often engage in surprisingly unselective blanket copying or “over-imitation” (for a review, see Whiten et al., 2009). Horner & Whiten (2005) observed that unlike chimpanzees, 3 to 4 years olds copied model’s actions even when they appeared causally irrelevant. Further studies have replicated this phenomenon of over-imitation in children (Lyons et al., 2007, 2011).

In our study, older children were more affected by imitative cues of punishment even when the model was violating the fairness norms: they decreased punishing unequal offers when the model was not punishing and they engaged in more punishment when the model was punishing the equal offers (see Fig. 2). Interestingly, when McGuigan et al. (2007) extended the above mentioned imitation studies to include 5-year olds, predicting that over-imitation would decline in cognitively more mature children, he observed an opposite effect: levels of imitation increased from 3 to 5 years of age. Later, it has been shown that the observed increase in over-imitation with age extends to adults with adults imitating even more than children (McGuigan et al., 2011). It has been suggested that adult humans continue to rely on “automatic coding” processes as they age, possibly more often when they perform a novel task and particularly in the presence of “expert” models (McGuigan et al. 2011). It is possible that our experiment presented a novel task (e.g. deciding whether to give up one sticker to punish a distributer) which induced older children to imitate the model even when the model’s action was violating the fairness expectation. The question of whether the same trend in the levels of imitation in costly punishment will extend to adults remains to be answered.

One other possible explanation why we observed higher levels of imitation in older children may be that the ages of the models (7 for the female model, and 10 for the male model) were closer to the older participants' age. Some evidence indicates that learners use cues of health, prestige, ethnic markers, sex, and age in figuring out who to learn from or imitate (Chudek et al., 2012, 2013; Efferson et al., 2008; Ryalls & Gul, 2000; Shutts et al., 2010). Therefore, it may be that the younger children in our study were less attentive to the model's decision than the older ones.

Our results show that punishment behavior can be culturally transmitted, which may lead to the stabilization of initially costly behaviors in a population. Prior work has already shown that cooperative behavior can be transmitted. Transmission in structured or structuring populations can lead to the clustering of cooperators and hence "social viscosity" (i.e. positive assortment of individuals who adhere to similar norms; Eshel & Cavalli-Sforza, 1982; Fowler & Christakis, 2010; Nowak & May, 1992; Ohtsuki et al., 2006). The existence of social viscosity is vitally important in maintaining large-scale cooperation (Fehr & Fischbacher, 2004b; Fischbacher et al., 2001; Fletcher & Doebeli, 2009; Keser & Winden, 2000). The transmission of third party punishing norms can potentially increase the rates of costly norm-support (so called 'altruistic') punishment by providing a proximate mechanism to mitigate the second order free rider problem. These costs can be psychological or social such as reduced status, ego depletion, negative reputations, being less trusted, susceptibility to retaliation (Adams & Mullen, 2012), damage to relationships, escalation of disputes into violence, time and energy costs (Wiessner, 2005) or emotional tensions

(Fehr & Fischbacher, 2004b). A recent cultural evolutionary model has shown that the total cost of punishment declines as the number of punishers increase in a society (Boyd et al., 2010). Accordingly, modest amounts of positive assortment of punishers in the formation of groups allow for the evolution of third-party punishing norms. Here, we argue that, under some social-ecological conditions (e.g., large groups), cultural transmission mechanisms such as imitative learning can help spread and stabilize punishment. This is achieved by diffusing the costs associated with individual punishment over the group of positively assorted punishers, and by the negative evaluations of norm-violators who fail to punish.

Finally, we address three concerns with our study. First, children may have been influenced by the demands of the task (i.e. “experimenter demand effect”) by being asked to decide whether or not to give away one sticker to punish the dictator (Zizzo, 2009). In our Equal Condition, especially, there may be a substantial experimenter demand effect, which might have encouraged the children to punish even though they, especially the older children, were not inclined to punish. However, the existence of the demand effect is a design feature of our experiment, since it allowed us to study the effect of the model who did not engage in the punishment on both equal and unequal offers. In the absence of the demand effect, we might not have been able to observe that a non-punishing model reduces punishing. In the Equal Control the proportion of punishers was 41% (aged and gender aggregated), and this proportion declined to 12%--but didn’t hit the floor—in the Equal NPT, and increased to 74% in Equal PT. These results clearly demonstrate high levels of cultural learning for costly punishment.

A second concern is the asymmetry between “taking an action” (punishing) vs. “not taking an action”. For this, the control treatment of Study 1 was modified by replacing “not punishing” with a hand clap option for the participants, so they could choose between two actions: punishing or clapping their hands. We predicted that the introduction of the second option would reduce the experimenter demand effect. Contrary to our expectations, participants’ level of punishment did not decline in Study 2. Although it is unclear, it might have been that some children were reluctant to clap their hands in the absence of an obvious cause, or thought they were applauding the dictators’ behavior.

A third concern is that our age trajectory may, in part, capture an increase in comprehension of the situation rather than anything about social motives. Several features of the data mitigate this concern. First, we reran our analyses in various ways, aiming to address this issue: (a) we ran our regression analyses by including children’s answers to our explicit “fair” or “unfair” question as a control variable (electronic supplementary material, Tables S3 and S4); this variable should systematically account for many of those who did not understand the game, at an explicit level; (b) we also reran by excluding all the children who answered our question in a way that was inconsistent with adult norms (electronic supplementary material, Tables S5 and S6); and (c) we examined only children age 5 and up (electronic supplementary material, Fig. S4, Tables S7 and S8). In all cases, our qualitative results held. Second, it’s worth noting that while some of the youngest children likely did not completely comprehend the situation (and answered randomly), it is still the case that systematic differences between the treatments and conditions emerge, even among the youngest participants. The key features of the experiment are influencing their choices.

Acknowledgments

We would like to thank Andy Baron and the Living Lab at the Science World, Vancouver, for providing a research site for these experiments; and Xuanni Yang, Roy Zhou, Randip Gill, Sara Coyle for their assistance in data collection and entry. GDS was supported by the Erasmus Mundus Masters Scholarship from the European Union. MJ was supported by the National Research Fund, Luxembourg, and co-funded under the Marie Curie Actions of the European Commission (FP7-COFUND).

References

Adams, G. S., & Mullen, E. (2012). The social and psychological costs of punishing. *Behavioral and brain sciences*, *35*, 15–16.

Axelrod, R. (1986). An evolutionary approach to norms. *American political science review*, *80*(4), 1095–1111.

Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children’s altruistic behavior in the dictator game. *Evolution and Human Behavior*, *28*(3), 168–175. doi:10.1016/j.evolhumbehav.2006.10.003

Blake, P. R., & Rand, D. G. (2010). Currency value moderates equity preference among young children. *Evolution and Human Behavior*, *31*(3), 210–218. doi:10.1016/j.evolhumbehav.2009.06.012

Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science (New York, N.Y.)*, *328*(5978), 617–20. doi:10.1126/science.1183665

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(6), 3531–5. doi:10.1073/pnas.0630443100

Boyd, R., & Richerson, P. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and sociobiology*, *195*, 171–195.

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral ecology and sociobiology*, *65*(3), 431–444. doi:10.1007/s00265-010-1100-3

Chudek, M., Brosseau-Liard, P., Birch, S., & Henrich, J. (2013). Culture-gene coevolutionary theory and children's selective social learning. In M. Banaji & S. Gelman (Eds.), *Navigating the Social World: The Early Years*. Oxford University Press.

Chudek, M., Heller, S., Birch, S., & Henrich, J. (2012). Prestige-biased cultural learning: bystander’s differential attention to potential models influences children's learning. *Evolution and Human Behavior*, *33*(1), 46–56. doi:10.1016/j.evolhumbehav.2011.05.005

Chudek, M., & Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in cognitive sciences*, *15*(5). doi:10.1016/j.tics.2011.03.003

Chudek, M., Zhao, W., & Henrich, J. (2013). Culture-Gene Coevolution, Large-Scale Cooperation and the Shaping of Human Social Psychology. In K. Sterelny, R. Joyce, B. Calcott, & B. Fraser (Eds.), *Cooperation and Its Evolution*. MIT Press.

Dehaene, S. (2009). *Reading in the Brain: The Science and Evolution of a Human Invention*. Viking Books.

Efferson, C., Lalive, R., & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science (New York, N.Y.)*, *321*(5897), 1844–9. doi:10.1126/science.1155805

Ensminger, J., & Henrich, J. (Eds.). (n.d.). *Experimenting with Social Norms: Fairness and Punishment in Cross Cultural Perspective*. New York: Russell Sage Press.

Eshel, I., & Cavalli-Sforza, L. (1982). Assortment of encounters and evolution of cooperativeness. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 1331–1335.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*(7208), 1079–83. doi:10.1038/nature07155

Fehr, E., & Fischbacher, U. (2004a). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. doi:10.1016/S1090-5138(04)00005-4

Fehr, E., & Fischbacher, U. (2004b). Social norms and human cooperation. *Trends in cognitive sciences*, *8*(4), 185–90. doi:10.1016/j.tics.2004.02.007

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, *90*(4), 980–994.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–40. doi:10.1038/415137a

Fischbacher, U., Gachter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404.

Fletcher, J. A., & Doebeli, M. (2009). A simple and general explanation for the evolution of altruism. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1654), 13–9. doi:10.1098/rspb.2008.0829

Fowler, J. J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National …*, *107*(12), 5334–5338. doi:10.1073/pnas.0913149107

Fu, F., Hauert, C., Nowak, M., & Wang, L. (2008). Reputation-based partner choice promotes cooperation in social networks. *Physical Review E*, *78*(2), 026117. doi:10.1103/PhysRevE.78.026117

Gatcher, S. (2012). In the lab and the field: punishment is rare in equilibrium. *Behavioral and Brain Sciences*, *35*(1), 26–28.

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: infants' reactions to equal and unequal distributions of resources. *Developmental science*, *14*(5), 1012–20. doi:10.1111/j.1467-7687.2011.01048.x

Gintis, H., Smith, E. a, & Bowles, S. (2001). Costly signaling and cooperation. *Journal of theoretical biology*, *213*(1), 103–19. doi:10.1006/jtbi.2001.2406

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*, 25–34. doi:10.1016/j.joep.2009.09.002

Guzman, A. R., Rodriguez-Sickert, C., & Rowthorn, R. (2007). When in Rome, do as the Romans do: the coevolution of altruistic punishment, conformist learning, and cooperation. *Evolution and Human Behavior*, *28*(2), 112–117. doi:10.1016/j.evolhumbehav.2006.08.002

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 19931–6. doi:10.1073/pnas.1110306108

Hauert, C., Traulsen, A., Brandt, H., Nowak, M. a, & Sigmund, K. (2007). Via freedom to coercion: the emergence of costly punishment. *Science (New York, N.Y.)*, *316*(5833), 1905–7. doi:10.1126/science.1141588

Henrich, J. (2009). The evolution of costly displays, cooperation and religion. *Evolution and Human Behavior*, *30*(4), 244–260. doi:10.1016/j.evolhumbehav.2009.03.005

Henrich, J., & Boyd, R. (2001). Why people punish defectors. Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of theoretical biology*, *208*(1), 79–89. doi:10.1006/jtbi.2000.2202

Henrich, J., Boyd, R., & Richerson, P. J. (2012). The puzzle of monogamous marriage. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *367*(1589), 657–69. doi:10.1098/rstb.2011.0290

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., … Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science (New York, N.Y.)*, *327*(5972), 1480–4. doi:10.1126/science.1182238

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., … Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767. doi:10.1126/science.1127333

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission.

Evolution and human behavior : official journal of the Human Behavior and Evolution Society, *22*(3), 165–196.

Henrich, N., & Henrich, J. (2007). *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. Oxford University Press.

Herrmann, B., Thöni, C., & Gatcher, S. (2008). Antisocial Punishment Across Societies. *Science*, *319*, 1362–1367.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal cognition*, *8*(3), 164–81. doi:10.1007/s10071-004-0239-6

House, B. R., Silk, J., Henrich, J., Barrett, C., Scelza, B., Boyette, A., … Laurence, S. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1221217110

Keser, C., & Winden, F. Van. (2000). Conditional Cooperation and Voluntary Contributions to Public Goods *. *Scand. J. of Economics*, *102*(1), 23–39.

Lyons, D. E., Damrosch, D. H., Lin, J. K., Macris, D. M., & Keil, F. C. (2011). The scope and limits of overimitation in the transmission of artefact culture. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *366*(1567), 1158–67. doi:10.1098/rstb.2010.0335

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19751–6. doi:10.1073/pnas.0704452104

Marlowe, F., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., … Tracer, D. (2008). More “altruistic”punishment in larger societies. *Proc. R. Soc. B*, *275*, 587–590. doi:10.1098/rspb.2007.1517

McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British journal of psychology (London, England : 1953)*, *102*(1), 1–18. doi:10.1348/000712610X493115

McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3- and 5-year-old children. *Cognitive Development*, *22*(3), 353–364. doi:10.1016/j.cogdev.2007.01.001

Milinski, M., Semmann, D., & Krambeck, H.-J. (2002). Reputation helps solve the “tragedy of the commons”. *Nature*, *415*(6870), 424–6. doi:10.1038/415424a

Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, *359*(6398), 826–829.

Ohtsuki, H., Hauert, C., Lieberman, E., & Nowak, M. a. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, *441*(7092), 502–5. doi:10.1038/nature04605

Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, *432*(7016), 499–502. doi:10.1038/nature02978

Rakoczy, H., Brosche, N., Warneken, F., & Tomasello, M. (2009). Young children's understanding of the context-relativity of normative rules in conventional games. *British Journal of Developmental Psychology*, *27*(2), 445–456. doi:10.1348/026151008X337752

Robbins, E., & Rochat, P. (2011). Emerging signs of strong reciprocity in human ontogeny. *Frontiers in psychology*, *2*(December), 353. doi:10.3389/fpsyg.2011.00353

Rochat, P., Dias, M. D. G., Broesch, T., Passos-Ferreira, C., Winning, a., & Berg, B. (2009). Fairness in Distributive Justice by 3- and 5-Year-Olds Across Seven Cultures. *Journal of Cross-Cultural Psychology*, *40*(3), 416–442. doi:10.1177/0022022109332844

Rossano, F., Rakoczy, H., & Tomasello, M. (2011). Young children's understanding of violations of property rights. *Cognition*, *121*(2), 219–27. doi:10.1016/j.cognition.2011.06.007

Ryalls, B., & Gul, R. (2000). Infant imitation of peer and adult models: Evidence for a peer model advantage. *Merrill-Palmer Quarterly*, *46*(1), 188–202.

Schmidt, M., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition*, *124*(3), 325–33. doi:10.1016/j.cognition.2012.06.004

Schmidt, M., & Sommerville, J. a. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. (M. Perc, Ed.)*PloS one*, *6*(10), e23223. doi:10.1371/journal.pone.0023223

Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental Science*, *13*(4), 599–610. doi:10.1111/j.1467-7687.2009.00913.x.Social

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. doi:10.1371/journal.pone.0059510

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and brain sciences*, *28*(5), 675–91; discussion 691–735. doi:10.1017/S0140525X05000129

Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual review of psychology*, *64*, 231–255. doi:10.1146/annurev-psych-113011-143812

Trommsdorff, G., Friedlmeier, W., & Mayer, B. (2007). Sympathy, distress, and prosocial behavior of preschool children in four cultures. *International Journal of Behavioral Development*, *31*(2007), 284–293.

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *The British journal of developmental psychology*, *29*(Pt 1), 124–30. doi:10.1348/026151010X532888

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee.

Philosophical transactions of the Royal Society of London. Series B, Biological sciences, *364*(1528), 2417–28. doi:10.1098/rstb.2009.0069

Wiessner, P. (2005). Norm Enforcement among the Ju /' hoansi Bushmen. *Human Nature*, *16*(2), 115–145.

Wyman, E., Rakoczy, H., & Tomasello, M. (2009). Normativity and context in young children's pretend play. *Cognitive Development*, *24*(2), 146–155. doi:10.1016/j.cogdev.2009.01.003

Zizzo, D. J. (2009). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. doi:10.1007/s10683-009-9230-z

*Title Page

Transmission and Development of Costly Punishment in Children

GUL DENIZ SALALI*

Department of Anthropology,

University College London

MYRIAM JUDA

Department of Psychology,

University of British Columbia

JOSEPH HENRICH

University of British Columbia,

Department of Psychology;

Vancouver School of Economics;

Canadian Institute for Advanced Research

* Corresponding author. Department of Anthropology, University College London, London WC1H 0BW, United Kingdom

E-mail address: guldenizsalali@gmail.com

Word count: 7,978

Figure legends

Fig. 1. Proportion of punishers in Control (CT), Punishment (PT) and No-Punishment (NPT) Treatments for Equal and Unequal Conditions. The blue line corresponds to the CT, the green line to the NPT, and the red line corresponds to the PT. Error bars indicate 95% confidence intervals.

Fig. 2. Probability of punishment by age. Solid curves are drawn based on the predictions from logistic regression models for each condition. Dashed curves represent 95% confidence intervals. (A) Probability of engaging in costly punishment when the distribution was equal. Age did not have a significant effect in any of the treatments for the Equal-CT, $P = 0.62$; for Equal-PT, $P = 0.95$; for Equal-NPT, $P = 0.11$). (B) Probability of engaging in costly punishment at each age when the distribution was unequal. For details, see the main text.

Fig. 3. Proportion of punishers by age in CT for Equal and Unequal Conditions in (A) Study 1 and (B) Study 2. Different age groups (3-4, 5-6, and 7-8 years) are represented by light to dark gray bars. Boxes inside the bars represent the sample size for the corresponding condition. Error bars indicate 95% binomial confidence intervals.

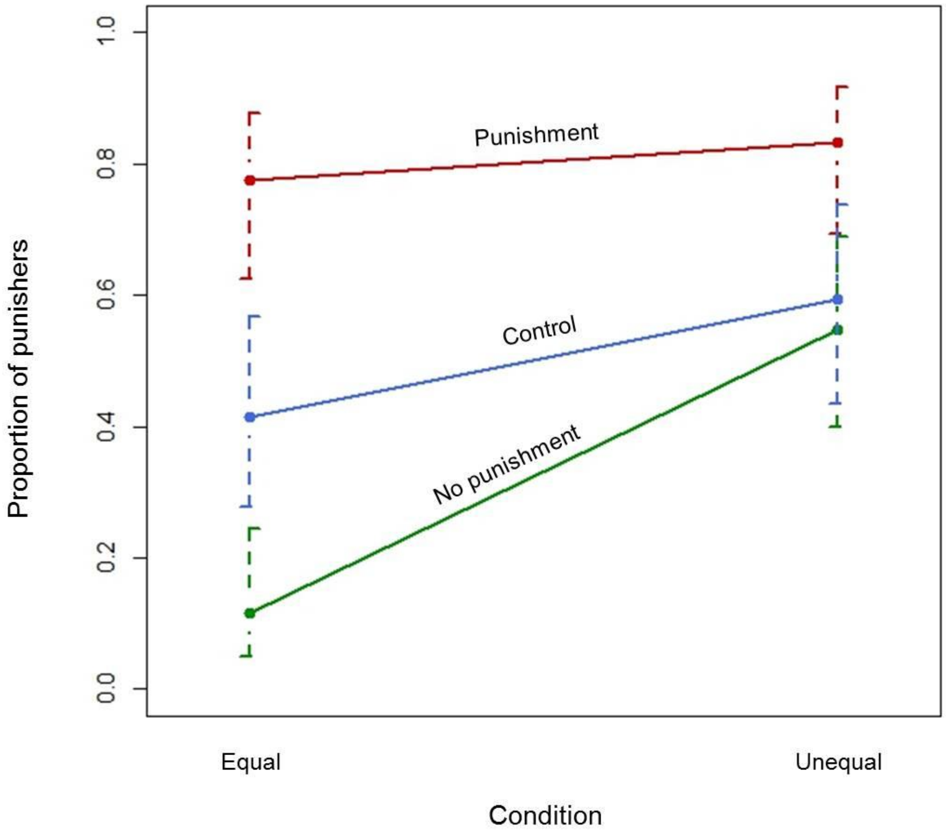
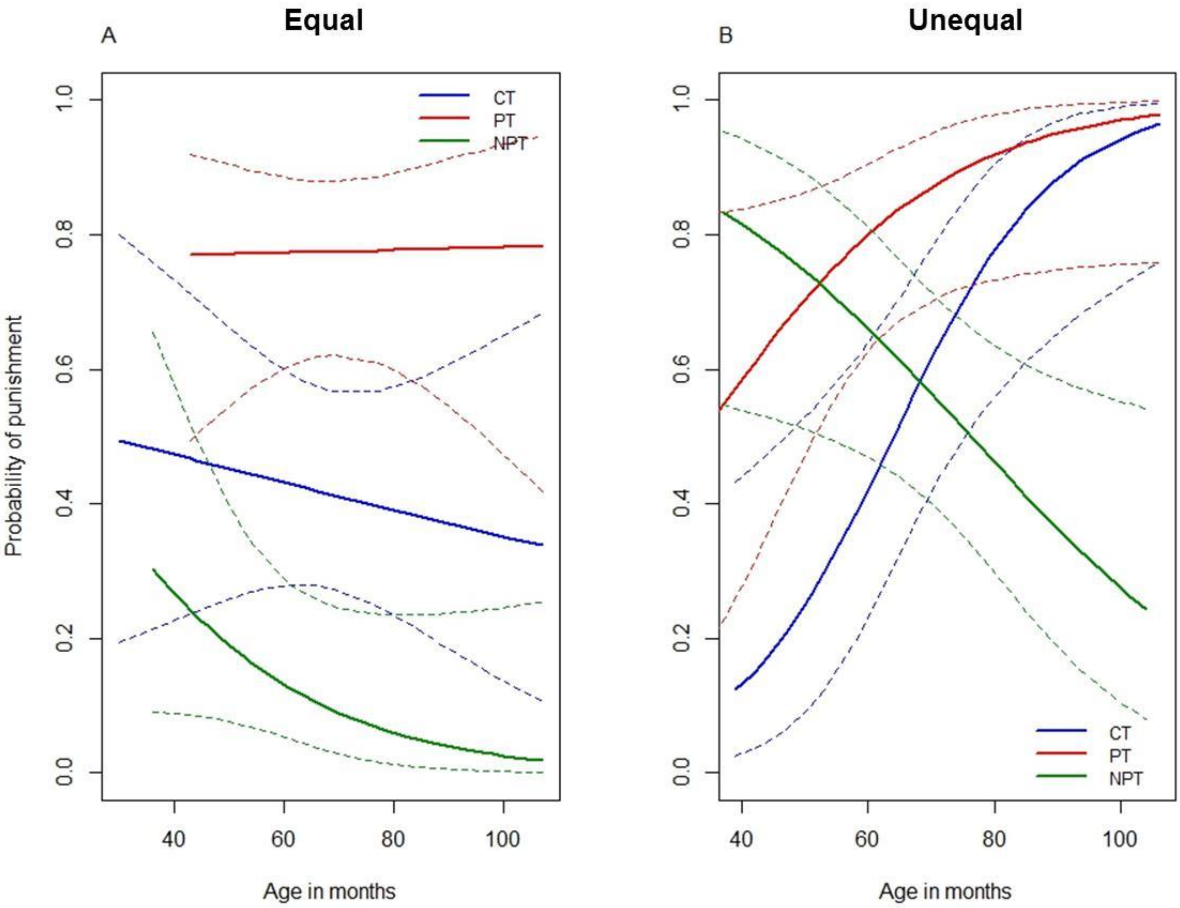

1.0
0.8
0.6
0.4
0.2
0.0
Proportion of punishers
Punishment
Control
No punishment
Equal
Unequal
Condition

Fig. 1.



Equal
A
CT
PT
NPT
Probability of punishment
0.0
0.2
0.4
0.6
0.8
1.0
40
60
80
100
Age in months
Unequal
B
CT
PT
NPT
0.0
0.2
0.4
0.6
0.8
1.0
40
60
80
100
Age in months

Fig. 2.

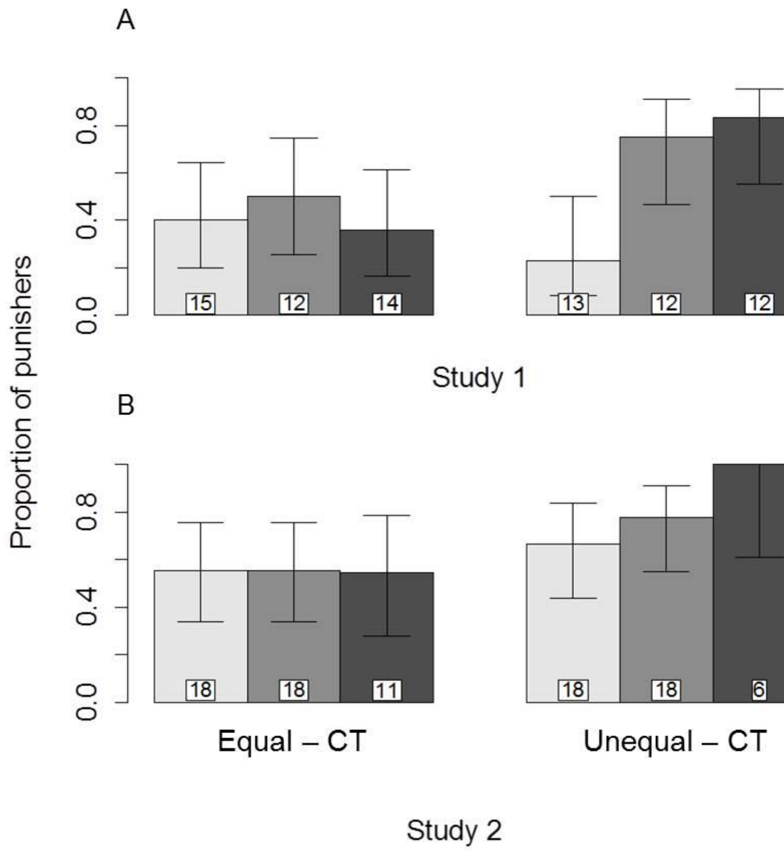

A
Proportion of punishers
0.0
0.4
0.8
15
12
14
13
12
12
Study 1
B
0.0
0.4
0.8
18
18
11
18
18
6
Equal – CT
Unequal – CT
Study 2

Fig. 3.

Table 1

Experimental conditions, treatments and the corresponding outcomes that were shown in the observation phase. “Not known” means this information was not provided to the participant.

Condition and treatment	Dictator’s distribution (self vs. receiver)	Dictator’s outcome	Receiver’s outcome	Model’s outcome
Equal – Control	2-2	Not known	Not known	Not known
Equal – No Punishment	2-2	2	2	6
Equal – Punishment	2-2	0	2	5
Unequal – Control	4-0	Not known	Not known	Not known
Unequal – No Punishment	4-0	4	0	6
Unequal – Punishment	4-0	2	0	5

Table 2

Logistic regression models for probability of punishment for the Equal Condition.

	Model 1		Model 2	
Predictors	Coeff.	SE	Coeff.	SE
Treatment- No Punishment (NPT)	-1.95**	0.62	-1.70**	0.59
Treatment- Punishment (PT)	1.91***	0.56	1.76***	0.53
Age (30-107 months)	-0.01	0.01		
Gender- male	0.25	0.46		
Ethnicity- Asian	-1.15	0.66		
Ethnicity- other	-0.60	0.61		
Order- right	-1.32*	0.51	-1.11*	0.48
(Intercept)	0.75	0.90		
Pseudo-R^2 (Hosmer-Lemeshow)	0.30		0.27	
-2 log likelihood	117.4		122.1	
N	122		122	

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

Logistic regression coefficients and their standard errors. Response variable: punishment. *Treatment* encodes which treatment the participant had (CT, NPT or PT), *Age* is the participant's age in months, *Gender* is the gender of the participant, *Order* encodes on which side the dictator appeared on the picture and *Ethnicity* is the ethnicity of the participant (European, Asian or other). *Ethnicity* is determined by the language(s) of the participants spoken at home. *N* is the number of subjects. Omitting non-significant predictors did not affect the model fit (for Model 1 and Model 2: $P[\chi^2(4) > 4.68] = 0.32$).

Table 3

Logistic regression models for probability of punishment for the Unequal Condition.

	Model 1		Model 2		Model 3	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Treatment- No Punishment (NPT)	-0.26	0.48	8.07***	2.29	8.14***	2.23
Treatment- Punishment (PT)	1.24*	0.55	3.15	2.50	3.27	2.40
Age (36-106 months)	0.02	0.01	0.08**	0.03	0.08**	0.03
Gender- male	-0.15	0.42				
Ethnicity- Asian	-0.29	0.62	-0.05	0.67		
Ethnicity- other	0.32	0.67	0.10	0.71		
Order- right	-0.29	0.42				
NPT x Age			-0.12***	0.03	-0.12***	0.03
PT x Age			-0.03	0.04	-0.03	0.04
(Intercept)	-0.40	0.90	-4.96**	1.85	-5.02**	1.79
Pseudo-R^2 (Hosmer-Lemeshow)	0.08		0.21		0.21	
-2 log likelihood	142.5		122.4		122.4	
N	121		121		121	

* *p*< 0.05; ** *p*< 0.01; *** p< 0.001

Logistic regression coefficients and their standard errors. Response variable: punishment. *Treatment* encodes which treatment the participant had (CT, NPT, PT), *Age* is the participant's age in months, *Gender* is the gender of the participant, *Order* encodes on which side the dictator appeared on the picture and *Ethnicity* is the ethnicity of the participant (European, Asian or other). *N* is the number of subjects. We used quasi-binomial logistic regression in cases where the residual deviance exceeded the residual degrees of freedom. Adding the *Treatment x Age* interaction term significantly increased the model's predictive power ($P[\chi^2(2) > 20.102] < 0.001$), while omitting non-significant predictor *Ethnicity* did not affect the model fit (for Model 2 and Model 3: $P[\chi^2(2) > 0.028] = 0.98$).

Table 4

Regression models for Equal Control Treatment - Study 2

	Model 1		Model 2	
Predictors	Coeff.	SE	Coeff.	SE
Age (31-111 months)	-0.01	0.02		
Gender- male	-1.06	0.66	-0.96	0.60
Order- right	-0.40	0.66		
Siblings- yes	-0.57	0.78		
(Intercept)	2.28	1.45	0.69	0.43
Pseudo-R^2 (Hosmer-Lemeshow)	0.07		0.04	
-2 log likelihood	60.05		62.05	
N	47		47.00	

Omitting non-significant predictors did not affect the model fit (for Model 1 and Model 2: $P[\chi 2(3) > 1.99] = 0.57$). Note that we did not include the Ethnicity variable because of the small sample size of the levels (N = 2 for Asians, and N = 9 for other).

Table 5

Regression model for Unequal Control Treatment - Study 2

	Model 1		Model 2	
	Coeff.	SE	Coeff.	SE
Age (38-108 months)	0.06^	0.03	0.06	0.03
Gender- male	0.41	0.80		
Order- right	-0.85	0.82		
Siblings- yes	-0.23	1.04		
(Intercept)	-2.17	1.96	-2.27	1.81
Pseudo-R^2 (Hosmer-Lemeshow)	0.13		0.10	
-2 log likelihood	40.03		41.54	
N	42		42	

^ p= 0.06

Omitting non-significant predictors did not affect the model fit (for Model 1 and Model 2: P[$\chi 2(3) > 1.51$] = 0.68). Note that we did not include the Ethnicity variable because of the small sample size of the levels (N = 9 for Asians, and N = 5 for other).