Evolution and Human Behavior 36 (2015) 86–94

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

Transmission and development of costly punishment in children

Gul Deniz Salali [a,*], Myriam Juda [b], Joseph Henrich [c]

[a] Department of Anthropology, University College London
[b] Department of Psychology, University of British Columbia
[c] University of British Columbia, Department of Psychology, Vancouver School of Economics, Canadian Institute for Advanced Research

ARTICLE INFO

Article history:
Initial receipt 22 November 2013
Final revision received 15 September 2014

Keywords:
Costly punishment
Cultural transmission
Children
Cooperation

ABSTRACT

Evolutionary 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, Gintis, & Bowles, 2010; Fu, Hauert, Nowak, & Wang, 2008; Hauert, Traulsen, Brandt, Nowak, & Sigmund, 2007; Milinski, Semmann, & Krambeck, 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 & Henrich, 2011; Henrich et al., 2010). 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, 2014; Herrmann, Thöni, & Gatcher, 2008; Wiessner, 2005). One mechanism to suppress free-riders is diffuse costly punishment, which some researchers have argued is particularly relevant to understand cooperation in large-scale societies.

Devised to study this particular mechanism, third-party punishment games examine whether a specific behavior is a social norm that is enforced by costly punishment. 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 an amount of her allocation to punish player 1 for her offer (Fehr & Fischbacher, 2004b). 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 increases long-term payoffs (Fehr & Fischbacher, 2004b; Fehr & Gächter, 2000, 2002; Fischbacher, Gachter, & Fehr, 2001; Gatcher, 2012; Henrich et al., 2006). However, individual's willingness to engage in third party punishment varies across societies, from zero in some societies to a strong willingness to punish selfish deviations from equality in others (Henrich et al., 2006; 2010; Marlowe et al., 2008).

Such sanctions and punishment, however, create an evolutionary puzzle, since punishing may incur costs on the individuals engaging in

* Corresponding author. Department of Anthropology, University College London, London WC1H 0BW, United Kingdom.
E-mail address: guldeniz.salali.12@ucl.ac.uk (G.D. Salali).

http://dx.doi.org/10.1016/j.evolhumbehav.2014.09.004

punishment. Punishment gives rise to a second-order free-rider problem when these costs, however small, exceed the benefit it creates for the punisher (Henrich & Boyd, 2001). Accordingly, the second-order free riders, those who do not punish and do not pay the cost of punishment, can eventually invade the population. According to one view, cultural transmission mechanisms can solve the second-order free rider problem by stabilizing costly punishment within a society. With respect to this view, individuals can (1) acquire their taste for costly punishment via cultural learning and (2) learn to punish any behavior (Henrich & Boyd, 2001; for a review see Chudek, Zhao, & Henrich, 2013). On the other hand, some researchers have argued that costly punishment arises from a genetically evolved psychology that was favoured because it reduced or eliminated the fitness advantage enjoyed by free-riders; consequently models of cultural transmission are not necessary for the explanation (Baumard, André, & Sperber, 2013; Delton, Krasnow, Cosmides, & Tooby, 2011; Price, Cosmides, & Tooby, 2002; Price, 2005).

Therefore, an important question is whether cooperation and social norms are 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; Midlarsky & Bryan, 1972; Trommsdorff, Friedlmeier, & Mayer, 2007). Research on children reveals not only sophisticated early learning of social rules (Tomasello, Carpenter, Call, Behne, & Moll, 2005), but by age three, children seem to understand the normative dimensions of property rights (Rossano, Rakoczy, & Tomasello, 2011), protest antisocial acts (Vaish, Missana, & Tomasello, 2011), or automatically enforce arbitrary rules that they have just acquired (Köymen et al., 2014; Rakoczy, Brosche, Warneken, & Tomasello, 2009; Schmidt, Rakoczy, & Tomasello, 2012; Wyman, Rakoczy, & Tomasello, 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 adherence to the norm of equal sharing increases in children (Smith, Blake, & Harris, 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. 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. Recently, Jordan, McAuliffe, and Warneken (2014) showed that costly punishment of selfish behavior increases between ages 6 and 8, while in-group favoritism declines. In younger children (19–24 month-olds), Hamlin, Wynn, Bloom, and Mahajan (2011) showed that children directed positive behavior (giving a treat) towards pro-social puppets and negative behavior (taking a treat) towards anti-social ones. However, no empirical work thus far has explored how social learning opportunities may influence children's willingness to engage in costly punishment, of either equitable or inequitable divisions.

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 simplified third-party punishment game played with stickers, we were able to measure the rates of imitation in punishment behavior. In particular, we test the following hypotheses:

1) While sharing and other aspects of fairness appear early in life (14–18 months), children only begin to acquire and enforce a conceptual understanding of norms at about 3 years of age (Tomasello & Vaish, 2013). As such, we expect the frequency of the punishment of unequal offers, relative to equal offers, to increase with age.
2) We predict that children will imitate punishment as well as non-punishment of both equal and unequal offers—because either might be the local norm for novel social interaction they engage in, and imitation is a rapid means of adapting to the local rules. However, as children's adherence to the norm of equal sharing increases with age (Blake, McAuliffe, & Warneken, 2014; Smith et al., 2013), we expect older children to be less imitative when the model's action violates the fairness expectations, because older children will be bringing in equality norms acquired outside the experiment and applying them to this novel context.

2. Study 1

2.1. Methods

2.1.1. Participants

Two hundred and forty-five children (117 females) between 30 and 107 months of age (ages < 60 months: $N = 88$, mean $\pm$ SD $= 48.5 \pm 7.2$; $60 \leq$ ages < 84: $N = 77$, mean $\pm$ SD $= 70.9 \pm 6.7$; ages ≥ 84: $N = 80$, mean $\pm$ SD $= 93.3 \pm 7.2$) 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. The languages spoken in bilingual households were Mandarin, Cantonese, Chinese, Japanese, Korean, Punjabi, French, Spanish, Farsi, Latvian, Tagalog, Malayam, Portuguese, Hindi, Hebrew and Cree.

2.1.2. 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.

2.1.2.1. 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 photos 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 photo, 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 photo, 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: an equal distribution (2–2) or an 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, Banaji, & Spelke, 2010). We did not match on ethnicity, but control for this in our analyses.

2.1.2.2. Test phase. During the test phase the participant took the role of the model and was then given 6 stickers, like the model. Before starting the experiment, the experimenter asked to the participant whether s/he liked the stickers. The experimenter, taking the role of the administrator, then showed two photos 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 photo 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 photo, if the participant gave up one of his/her own stickers (see electronic supplementary material for the script, procedure and additional information on the experimental design).

We used photos and videos, and not real-time players for two reasons. First, the use of video helped us guarantee consistency across participants and conditions. By contrast, while adding realism, the use of real actors, particularly child actors, risks inconsistencies and adds noise, or even potentially systematic bias. Second, since we wanted our child participants to react to other children (i.e. "peer punishment") the practicalities of training child actors to run hundreds of participants were prohibitive.

2.1.3. 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. We kept the sample size as equal as possible with regards to gender and age in each of the 6 groups (see Supplementary Table S1). The physical position of the dictator in the photos was counterbalanced (left vs. right) across participants.

2.1.4. Statistical analysis

Since our response variable was binary (punish/don't punish), we used logistic regression procedures in R 3.0.2. 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 photo: left, right), and ethnicity (of the participant: European, Asian, other). We did a separate analysis investigating effects of having siblings (electronic supplementary material, Tables S2 and S3).

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 both the full model and the most parsimonious models 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.

2.2. Results

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

Yes, as shown in Fig. 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. Fig. 2A visually illustrates these relationships.

In the unequal condition, as shown in Fig. 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 Fig. 2B emerges as the interaction of our no-punishment treatment (NPT) and age in Table 3. The *NPT* $\times$ *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 Fig. 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 Fig. 2B show that for the

Table 1
Experimental conditions, treatments and the corresponding outcomes that were shown in the observation phase.

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

"Not known" means this information was not provided to the participant.

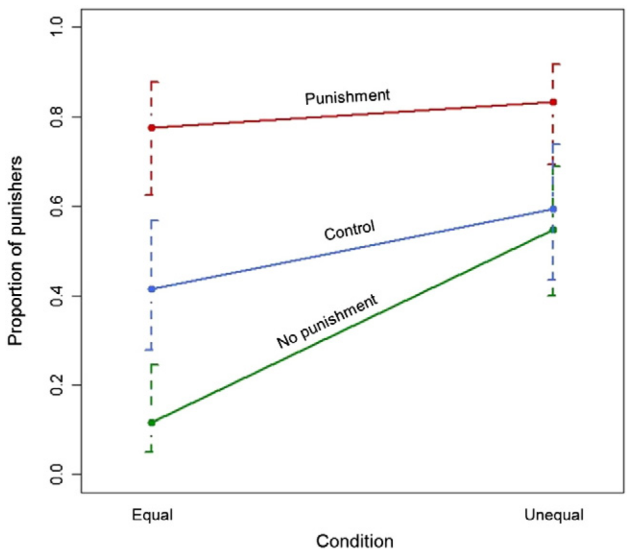

Fig. 1. Proportion of punishers in control (CT), punishment (PT) and no-punishment (NPT) treatments for equal and unequal conditions. The blue line corresponds to the CT, the green line to the NPT, and the red line corresponds to the PT. Error bars indicate 95% confidence intervals.

Table 2
Logistic regression models for probability of punishment for the equal condition.

Predictors	Model 1		Model 2	
	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	0.08	0.36
Pseudo-R^2 (Hosmer–Lemeshow)	0.30		0.27	
-2 log likelihood	117.4		122.1	
N	122		122	

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. Model 1 is the initial model containing all the predictors. Model 2 is the optimal model based on the likelihood ratio test statistic and its associated *p*-value. Omitting non-significant predictors did not affect the model fit (for model 1 and model 2: $P[\chi^2(4) > 4.68] = 0.32$).
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

youngest children the probability of punishing is quite different in the PT vs. the CT.

2.2.2. 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 show 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, Fig. 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 NPT × 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, Fig. 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.

Fig. 2 shows that younger children did not selectively engage in costly punishment for unequal offers in the CT. Moreover, based on our best-fit age curves, we find that selective punishment of unequal offer begins to emerge at age 5 (60 months) and achieves conventional level of significance at around 68 months (electronic supplementary material, Fig. S2). One important concern suggested by this is that the youngest children may not fully comprehend the experiment. To further check the comprehension of younger participants and test the robustness of our results we performed additional analyses (electronic supplementary material, Section 3 Tables S4-5). First, we controlled for the correctness of the answers given to the question of number of stickers each player had on the photo. Adding a control variable for sticker counts did not improve our optimal model's predictive power. Second, we excluded the participants who did not accurately report the number of stickers. In the optimal model from the resulting regression analysis, the significant predictors remained the same as in our main optimal model (Table S4, model 3.2). Third, we controlled for the answers given to the question on whether the outcome of the game was fair ("Do you think this is fair"). Adding this control variable did not improve our optimal model (Table S4, model 4 and Table S5, model 2). Fourth, we excluded the participants who failed to answer the fairness question in a manner consistent with local social norms. For both equal and unequal conditions, results of the regression analyses remained the same. To summarize, neither controlling for our comprehension checks nor dropping participants who failed the checks altered our main results.

Finally, we further investigated the robustness of our results by excluding (1) only the 3 year-olds, (2) both the 3 and 4 year-olds from our analyses. Electronic supplementary material, Sections 3.3.3–4 detail our results from these subsets of data. For the equal condition, our optimal model, and the significant predictors remained the same when 3 year-olds were excluded. When we excluded both the 3- and 4 year-olds, the optimal model included our theoretically most important predictors (*NPT* and *PT*), but not the order variable, unlike our main model. For the unequal condition, the optimal model (including treatment, age and treatment × age interactions) remained the same when we excluded 3 and 3–4 year-olds, however because of the decline in statistical power (due to smaller sample size) and the compression of the age range some predictors lose statistical significance. Tables S4 and S5 compare the effect sizes of our key and control predictors for our main model and other models.

In summary, study 1 showed that children imitated the model's decision on punishing equal distributions regardless of their age. In the absence of the model's influence, the rates of punishment of unequal distributions increased by age. Older children were more inclined to imitate when the model did not punish the unequal offer.

3. Study 2

In the absence of cultural transmission opportunities, study 1 generated a substantial amount of punishment across all conditions, perhaps due to the demands of the task created by asking the children to decide

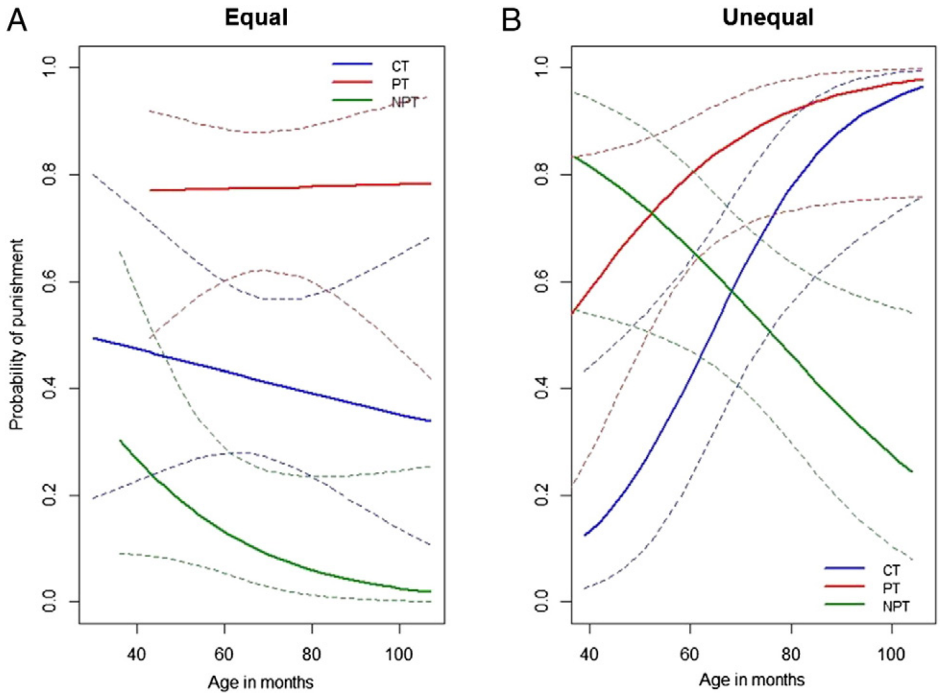

Fig. 2. Probability of punishment by age. Solid curves are drawn based on the predictions from logistic regression models for each condition. Dashed curves represent 95% confidence intervals. (A) Probability of engaging in costly punishment when the distribution was equal. Age did not have a significant 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.

between punishing or doing nothing. We had figured that this action asymmetry might elevate punishment levels, which was a desirable design feature, since it enabled us to avoid "floor effects" that might have prevented 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) actually reduces punishing. However, to explore such task demand effects, 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.

3.1. Methods

3.1.1. Participants

Eighty-nine children (45 females) between 31 and 111 months of age 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.

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 × Age			−0.12***	0.03	−0.12***	0.03
PT × 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	

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. Model 1 is the initial model containing all the predictors. Model 3 is the optimal model based on the likelihood ratio test statistic and its associated *p*-value. Adding the *Treatment × Age* interaction term significantly increased the model's predictive power (for model 1 and model 2: $P[\chi^2(2) > 20.10] < 0.001$), while omitting non-significant predictor ethnicity did not affect the model fit (for model 2 and model 3: $P[\chi^2(2) > 0.03] = 0.98$).

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

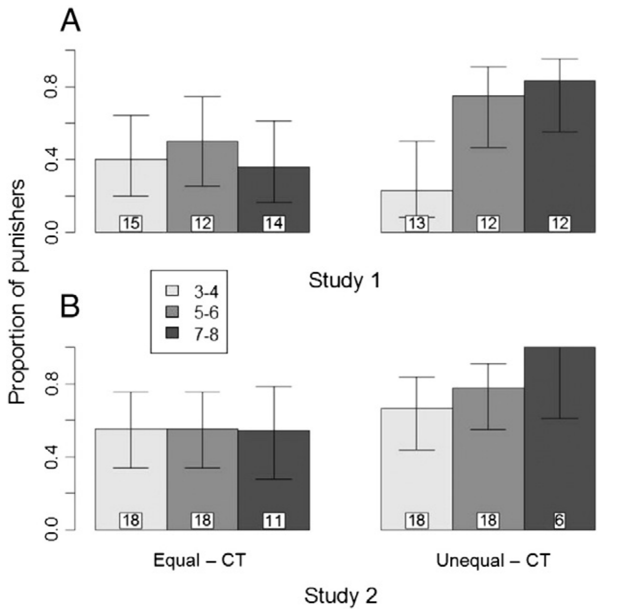

Fig. 3. Proportion of punishers by age in CT for equal and unequal conditions in (A) study 1 and (B) study 2. Different age groups (3–4, 5–6, and 7–8 years) are represented by light to dark gray bars. Boxes inside the bars represent the sample size for the corresponding condition. Error bars indicate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

3.1.2. 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.

3.1.3. 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 photos (whether the dictator was sitting on the left or on the right side of the table) was again counter-balanced across participants.

3.1.4. Statistical analysis

We used proportion test (a chi-square test with continuity correction) in R3.0.2. 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 photos were our control variables.

Table 4
Regression models for equal control treatment—study 2.

Predictors	Model 1		Model 2	
	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: $P[\chi2(3) > 1.99] = 0.57$). 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).

3.2. 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 $\chi^2(1, n = 88) = 1.174, p = 0.28$). The results were the same for the unequal conditions (mean proportion of punishment in study 1: 0.59, study 2: 0.76; Pearson $\chi^2(1, n = 79) = 1.83, p = 0.18$). If anything, the clapping option seems to increase punishing relative to doing nothing.

As in study 1, 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 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 ($CI_{.95} = [0.97–12.43]$).

4. Discussion

Our results suggest that costly punishment can be transmitted by cultural learning in children. First, children, regardless of their age, readily imitated costly punishment in either the equal or unequal conditions, though these effects are not symmetric. Second, children get more imitative as they get older, especially when the model does not punish. Third, based on our best-fit age curves, selective punishment of unequal offers begins to emerge at 5 years of age, but does not achieve conventional level of significance until 68 months. These findings support the notion that the sanctioning of norm violators can be influenced via cultural learning, leading to the acquisition and maintenance of norms for both prosocial and anti-social behaviors. Below, we discuss each of these findings in light of recent relevant research investigating children's prosocial and imitative abilities and tendencies. Then, we will address potential concerns with our experimental design.

Our findings 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 either equal or unequal distributions via cultural transmission (Chudek & Henrich, 2011; Guzman, Rodriguez-Sickert, & Rowthorn, 2007; Henrich & Boyd, 2001). Not only does this work show that costly punishment is culturally transmittable, but supports the implication that both the sanctioning of prosocial and non-prosocial behaviors can be readily acquired via cultural transmission. The finding that children imitate punishing for equal offers suggests that arbitrary or even maladaptive norms can be acquired and spread in a population as well as potentially evolved predispositions (such as young babies' capacity for moral judgment, see Bloom, 2013). Indeed, many societies possess maladaptive social traits such as disease-spreading cannibalism of dead relatives (Lindenbaum,

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	

Omitting non-significant predictors did not affect the model fit (for model 1 and model 2: $P[\chi2(3) > 1.51] = 0.68$). 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).
* $p = 0.06$.

2008) or fertility decline due to reproductive decisions in modern populations (Colleran, Jasienska, Nenko, Galbarczyk, & Mace, 2014). Future empirical studies will shed more light into the extent of influence of cultural transmission on our evolved predispositions.

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, McGuigan, Marshall-Pescini, & Hopper, 2009). Horner and Whiten (2005) observed that unlike chimpanzees, 3 to 4 year olds copied model's actions even when they appeared causally irrelevant. Further studies have replicated this phenomenon of over-imitation in children (Lyons, Young, & Keil, 2007; 2011). Future investigations can address the extent of imitation in other domains in children.

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, Whiten, Flynn, and Horner (2007) extended the above mentioned imitation studies to include 5-year olds, predicting that over-imitation would decline in cognitively more mature children, they observed an opposite effect: levels of imitation increased from 3 to 5 years of age. Later worked showed that the observed increase in over-imitation with age extends to adults with adults imitating even more than children (McGuigan, Makinson, & Whiten, 2011). This team has 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 violated the fairness expectations the children imported into the experiment. 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, Heller, Birch, & Henrich, 2012; Chudek, Brosseau-Liard, Birch, & Henrich, 2013; Efferson, Lalive, & Fehr, 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, Hauert, Lieberman, & Nowak, 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, 2004a). 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.

There are some limitations to 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 the 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 reason, or thought they were applauding the dictators' behavior. One possible improvement to the design of study 2 could involve asking children to choose between taking 5 or 6 stickers rather than asking them to keep all or lose one. In this case, the 6 stickers would be laid on the table at the beginning of the experiment, and after the presentation of the dictator's offer, the participant is asked whether they would like to take all the 6 stickers, or leave one to punish the dictator.

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. However, several features of the data mitigate this concern. First, we reran our analyses in various ways, aiming to address this issue: (a) by including a control variable which accounted for the children's ability to accurately count the number of stickers; b) we excluded children who were not able to count correctly; (c) we included children's answers to our explicit "fair" or "unfair" question as a control variable; this variable should systematically account for many of those who did not understand the game, at an explicit level; (d) we reran by excluding all the children who answered our question in a way that was inconsistent with adult norms; and (e) we reanalyzed, excluding the children who failed to count the stickers correctly and those who did not answer the fairness question correctly. In all cases, our optimal models and significant predictors remained the same as our main optimal model. Even when we removed both those who failed the sticker-count check and the fairness question, our main results stand (see electronic supplementary material, Tables S4 and 5). Next, we examined (a) only children age 4 and older; (b) only children age 5 and older. For the equal condition, our optimal model and significant key predictors (NPT and PT) remained the same, even after excluding both 3 and 4 year-olds. For the unequal condition, our optimal models remained the same. However, exclusion of 3–4 year-olds altered our key coefficients and degree of confidence in those estimates, due to the compression in age range and subsequent reduction of statistical power. Nevertheless, it is impressive that we retained our qualitative findings.

A fourth concern relates to the use of photos of a dictator game that was played between two children, and whether children truly comprehended the task in the absence of real actors. We used photos mainly to maintain consistency across participants and for practical difficulties of using child actors. Nevertheless, our design could be improved by the use of a pre-recorded video of children playing the dictator game. The participant could be told that the children on the screen are playing the game in a different room at the time of the experiment. Moreover, existing studies often relied on puppet actors. However, it is unclear how children perceive puppets; children aged 3 and older may know that puppets are not real actors. So, our work provides a complement to existing work using puppets.

It should be noted that during recruitment and the experiment, we told participants that they were going to play a sticker game. The word "game" was used to engage children in the experiment. Although the game itself is not real, the way the children played the game in the photo represented a situation that may have invoked participants' moral judgments. Nevertheless, one possible improvement might be to inform the participants that they will see photos of an event that took place between two children and be asked about the outcomes of that event, rather than using the word "game".

Finally, problems associated with the use of deception are not relevant in the context of this study. Deception poses a risk of influencing participant behavior in the case where future participants have already learnt about such deception in class, in prior laboratory experiences, or from social contacts at the university. However, our use of children at the Vancouver's Science World does not pose this risk since (a) we did not need to de-brief the children on the fact that the other children are not actually punished (they never know) and (b) children, unlike undergraduates, do not learn about economic or psychology experiments in school prior to entering the lab.

Here we found that sanctioning behaviors can be transmitted through imitation in children, with increasing imitation as children get older. These findings emphasize the likely role of cultural transmission in the spread of sanctioning norms in a population.

Supplementary Materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.evolhumbehav.2014.09.004.

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; Xuanni Yang, Roy Zhou, Randip Gill, Sara Coyle for their assistance in data collection and entry; and Matthew Gwynfryn Thomas for proofreading the manuscript. 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). JH would like to thank the Canadian Institute for Advanced Research (CIFAR) and SSHRC.

References

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

Baumard, N., André, J. -B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *The Behavioral and Brain Sciences*, *36*(1), 59–78, http://dx.doi.org/10.1017/S0140525X11002202.

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge–behavior gap. *Trends in Cognitive Sciences*, 1–3, http://dx.doi.org/10.1016/j.tics.2014.08.003.

Bloom, P. (2013). *Just Babies.* New York: Crown.

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), 617–620, http://dx.doi.org/10.1126/science.1183665.

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, http://dx.doi.org/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), http://dx.doi.org/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.

Colleran, H., Jasienska, G., Nenko, I., Galbarczyk, A., & Mace, R. (2014). Community-level education accelerates the cultural evolution of fertility decline. *Proceedings of the Royal Society B*, *281*(1779), 20132732.

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences*, *108*(32), 13335–13340, http://dx.doi.org/10.1073/pnas.1102131108.

Efferson, C., Lalive, R., & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science (New York, N.Y.)*, *321*(5897), 1844–1849, http://dx.doi.org/10.1126/science.1155805.

Ensminger, J., & Henrich, J. (Eds.). (2014). *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., & Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190, http://dx.doi.org/10.1016/j.tics.2004.02.007.

Fehr, E., & Fischbacher, U. (2004b). 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., & 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), 137–140, http://dx.doi.org/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), 13–19, http://dx.doi.org/10.1098/rspb.2008.0829.

Fowler, J. J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(12), 5334–5338, http://dx.doi.org/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, http://dx.doi.org/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.

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, http://dx.doi.org/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), 19931–19936, http://dx.doi.org/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), 1905–1907, http://dx.doi.org/10.1126/science.1141588.

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, http://dx.doi.org/10.1006/jtbi.2000.2202.

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 (New York, N.Y.), *327*(5972), 1480–1484, http://dx.doi.org/10.1126/science.1182238.
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.
Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., et al. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767, http://dx.doi.org/10.1126/science.1127333.
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), 164–181, http://dx.doi.org/10.1007/s10071-004-0239-6.
House, B. R., Silk, J., Henrich, J., Barrett, C., Scelza, B., Boyette, A., et al. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences*, http://dx.doi.org/10.1073/pnas.1221217110.
Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences of the United States of America*, http://dx.doi.org/10.1073/pnas.1402280111.
Keser, C., & Winden, F. Van (2000). Conditional cooperation and voluntary contributions to public goods *. *The Scandinavian Journal of Economics*, *102*(1), 23–39.
Köymen, B., Lieven, E., Engemann, D. a, Rakoczy, H., Warneken, F., & Tomasello, M. (2014). Children's norm enforcement in their interactions with peers. *Child Development*, *85*(3), 1108–1122, http://dx.doi.org/10.1111/cdev.12178.
Lindenbaum, S. (2008). Understanding kuru: The contribution of anthropology and medicine. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1510), 3715–3720, http://dx.doi.org/10.1098/rstb.2008.0072.
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), 1158–1167, http://dx.doi.org/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), 19751–19756, http://dx.doi.org/10.1073/pnas.0704452104.
Marlowe, F., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., et al. (2008). More "altruistic"punishment in larger societies. *Proceedings of the Royal Society B*, *275*, 587–590, http://dx.doi.org/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, http://dx.doi.org/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, http://dx.doi.org/10.1016/j.cogdev.2007.01.001.
Midlarsky, E., & Bryan, J. H. (1972). Affect expressions and children's imitative altruism. *Journal Of Experimental Research in Personality*, 6(2–3), 195–203.
Milinski, M., Semmann, D., & Krambeck, H. -J. (2002). Reputation helps solve the "tragedy of the commons". *Nature*, *415*(6870), 424–426, http://dx.doi.org/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), 502–505, http://dx.doi.org/10.1038/nature04605.
Price, M. E. (2005). Punitive sentiment among the Shuar and in industrialized societies: Cross-cultural similarities. *Evolution and Human Behavior*, *26*(3), 279–287, http://dx.doi.org/10.1016/j.evolhumbehav.2004.08.009.
Price, M. E., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior*, *23*(3), 203–231, http://dx.doi.org/10.1016/S1090-5138(01)00093-9.
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, http://dx.doi.org/10.1348/026151008X337752.
Robbins, E., & Rochat, P. (2011). Emerging signs of strong reciprocity in human ontogeny. *Frontiers in Psychology*, *2*, 353, http://dx.doi.org/10.3389/fpsyg.2011.00353.
Rossano, F., Rakoczy, H., & Tomasello, M. (2011). Young children's understanding of violations of property rights. *Cognition*, *121*(2), 219–227, http://dx.doi.org/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), 325–333, http://dx.doi.org/10.1016/j.cognition.2012.06.004.
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, http://dx.doi.org/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, http://dx.doi.org/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), 675–691, http://dx.doi.org/10.1017/S0140525X05000129 (discussion 691–735).
Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology*, *64*, 231–255, http://dx.doi.org/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), 124–130, http://dx.doi.org/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), 2417–2428, http://dx.doi.org/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, http://dx.doi.org/10.1016/j.cogdev.2009.01.003.
Zizzo, D. J. (2009). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98, http://dx.doi.org/10.1007/s10683-009-9230-z.