Title

Ontogenetic and Phylogenetic Variation in Prosocial Behavior: Differences in Prosociality Across Human Development and Primate Species

Author

House, Bailey

Publication Date

2013

UNIVERSITY OF CALIFORNIA

Los Angeles

Ontogenetic and Phylogenetic Variation in Prosocial Behavior:

Differences in Prosociality Across Human Development and Primate Species

A Dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Anthropology

by

Bailey Rogers House

2013

ABSTRACT OF THE DISSERTATION

Ontogenetic and Phylogenetic Variation in Prosocial Behavior:
Differences in Prosociality Across Human Development and Primate Species

By

Bailey Rogers House
Doctor of Philosophy in Anthropology
University of California, Los Angeles, 2013
Professor Joan B. Silk, Chair

Cooperation among genetic kin is a widespread phenomenon in nature, but this can't explain the widespread human motivation towards cooperating with non-relatives. This behavior is likely motivated both by evolved psychological mechanisms based on contingent reciprocity, and also by societally-varying cultural beliefs. In this dissertation I develop methods to explore the role of contingency and culture in human prosocial development, and the nature of prosociality across human and non-human primates. Chapter 1 investigates the development of contingent prosociality in American children, and chapter 2 charts the emergence of variation in prosociality across diverse societies. Chapter 3 develops methods for testing a range of prosocial behaviors in captive chimpanzees. These studies shed light on the motivations behind human cooperation, and the ontogenetic and phylogenetic origins of those motivations.

The Dissertation of Bailey Rogers House is approved.

H. Clark Barrett

Robert Boyd

Scott Johnson

Joan B. Silk, Chair

University of California, Los Angeles

2013

To my parents.

CONTENTS

Acknowledgments

Many people played a role in the creation of this dissertation. I must first acknowledge and thank Joan Silk, my committee chair and mentor. Nearly all of my work over the last six years was influenced by her sharp mind and red pen. I came to UCLA with a project in mind and while Joan readily accepted that proposal, over the years she has convinced me to adjust nearly every part of it, so that this dissertation is both exactly what I set out to do and drastically different from what I had in mind. The work is much better as a result. She has vastly improved my thinking and writing, and has always offered just as much support as was necessary. This meant she pushed hard, but also allowed me to try to solve problems on my own. I wouldn't have asked for more or less, and I owe her much for the time and effort she contributed to me and my work. I hope to eventually pay off that debt somehow, although I'm sure she hasn't been counting.

I also owe a debt of gratitude to the other members of my committee. Clark Barrett is one of the clearest thinkers I know, and for this reason alone his contributions to my work would have been invaluable. But I have also learned much from his tendency to seek common ground between competing ideas. While that is often a somewhat thankless task, I have learned that it is a crucial one, and something that I intend to attempt as well. Rob Boyd's depth and breadth of knowledge can be terrifying, but his genuine willingness to explain always makes it worth facing the fear of asking. I will never know as much as Rob, but his questions have motivated me to always be curious to know more about everything. Scott Johnson immediately welcomed me into his lab at the very beginning of my time at UCLA, and has been tremendously supportive ever since. I deeply appreciate his willingness to open his doors to me, and because of this experience I have learned the great value of engaging honestly with those who have different perspectives.

Both my work and my approach to a career in science are much improved thanks to each of the members of my committee, and I am thankful to them all.

Joe Henrich was integral to the design, analysis, and writing of my published Master's work, as well as Chapters 1 and 2 in the current dissertation. Both the practical and material help that Joe has provided have helped make these projects possible, and I cannot overstate how much I appreciate his willingness to assist with my writing and research ideas. Yet, perhaps even more importantly, both Joan and Joe have always been there to encourage and celebrate when I needed it, and that earns my deepest gratitude.

I also owe a great debt to Richard McElreath for his help with the statistical analyses used in all of my work; his help was particularly essential for Chapter 2. I have learned much from Richard and his thoughtful, patient approach that walked me through analyses I never would have thought I could do. I will never be able to answer an email as quickly or as helpfully as Richard always does, but I will be sure try to do so from now on.

I also must thank my many collaborators on the work described in this dissertation. Chapter 1 was a collaborative project with Joan Silk, Joe Henrich, and Barbara Sarnecka. I have known Barbara for many years, and I am very appreciative of her friendship and willingness to allow me to participate in her laboratory at UC Irvine and to conduct the study on which Chapter 1 is based. Chapter 2 was a collaboration between many people: Joan Silk, Joe Henrich, Stephen Laurence, Clark Barrett, Richard McElreath, Brooke Scelza, Adam Boyette, and Barry Hewlett. All were willing to put forth the effort that was needed into the project, and I couldn't have asked for more helpful and supportive coauthors. I must specifically thank Stephen Laurence at Sheffield University for picking up the project as part of a larger cross-cultural research program through the AHRC Culture and the Mind Project. His help was crucial in shepherding the project through data collection and writing. If Steve hadn't believed in the work at the beginning, the project never would have happened. I also want to thank Brooke, Adam, and Barry for their great comments and ideas on the manuscript, and also Stacy Rosenbaum for her help with data collection.

Chapter 3 would have been impossible without the strong support of Susan Lambeth and Steven Schapiro at the MD Anderson Michale E. Keeling Center for Comparative Medicine in Bastrop, TX. Steve allowed me to work with the chimpanzees at the facility, and he provided helpful suggestions and comments on the manuscript. Susan was a pillar of positivity and support through my entire stay, and I will never forget her warmth and openness. I must also thank all of the staff at the facility, and in particular Erica Thiele for her help as I learned how to work with the animals.

It is also important to thank the communities to which I belong, and that have supported my work through both financial support and invaluable feedback. The UCLA Center for Behavior, Evolution, and Culture (BEC) has been my home for the last six years and I cannot overstate how much I appreciate the value of this group. It has been a privilege to be a member and to have access to the eminent scholars affiliated with the Center, and those who have been invited to speak there. Similarly, I must thank the members of the Experimental Biological Anthropology (XBA) lab group in the UCLA Department of Anthropology, which has provided many students with critical opportunities for presenting our work and receiving feedback. I must thank the faculty that have organized BEC and XBA over the last six years: Dan Fessler, Clark Barrett, Joe Manson, Brooke Scelza, Joan Silk, and Rob Boyd. I must also thank my colleagues at UCLA for years of emotional and intellectual support: Cristina Moya, Charles Perreault, Irene Godoy, Karthik Panchanathan, Matt Gervais, Michelle Kline, Ming Xue, Sarah Mathew, Siamak

Naficy, Stacy Rosenbaum, and Willem Frankenhuis. I also wish to thank Hojin Kim, Jim Morgante, Lauren Burakowski, Lauren Krogh, Raphael Castaneda, and Sarah Gaither from the UCLA Baby Lab for their help and friendship. I similarly must thank my colleagues at the UC Irvine Cognitive Development Lab: Emily Slusser, James Negen, and Meghan Goldman. Finally, I want to thank my colleagues during my research in Bastrop, TX: Lara Wood, Lisa O'Bryan, Lydia Hopper, and Sara Price. Their friendship and ideas mattered very much to me.

The staff at the UCLA Department of Anthropology also deserve a special honor for guiding all of the undergraduate and graduate students to graduation. I owe special debts to Ann Walters and Tracy Humbert, without whom I almost certainly would have stumbled. Ann keeps the entire system moving, and Tracy goes well above and beyond what she should have to do to ensure that we stay employed with a roof above our heads. Both Ann and Tracy make students feel as though there is someone out there looking out for us, and that is at least as valuable as everything else they do for us. They provide rays of light when grad school seems dark, and for that I am most grateful.

It is also important to acknowledge the financial support I received. The work for Chapters 1 and 2 were in part supported by a two-year training fellowship from NSF IGERT through the UCLA Interdisciplinary Relationship Science Program. Much of the fieldwork described in Chapter 2 was also funded in part by the Culture and the Mind program, affiliated with the UK Arts and Humanities Research Council. The work in Chapter 3 was funded through a UCLA Graduate Research Mentorship appointment. This dissertation was also written during a Dissertation Year Fellowship awarded by UCLA's Graduate Division. I am grateful to each of these institutions and programs for enabling my work and this dissertation.

Finally, I must thank all of those friends and family who made sacrifices to help me through the last six years. I must give particular thanks to my parents Kathy Adams-House and Jeffrey House, who have been enormously supportive and accepting of the choices I needed to make to be successful in what I have chosen to do. Also deserving a special honor is Amy Malek, my partner of the last four years who has been there to listen to me, to support and critique my work for the better, to make me take breaks, and to encourage me when I needed it most. I really wouldn't be where I am today without these three people, and they should feel pride in being part of this accomplishment.

Chapter 1 has been published:

> House, B.R., Henrich, J., Sarnecka, B., Silk, J.B., 2013. *The development of contingent reciprocity in children*. Evolution and Human Behavior 34, 86–93.

Chapter 2 is in press:

> House, B.R., Silk, J.B., Henrich, J., Barrett, H.C., Scelza, B.A., Boyette, A.H., Hewlett, B., McElreath, R., Laurence, S., in press. *The Ontogeny of Prosoical Behavior across Diverse Societies*. Proceedings of the National Academy of Sciences of the United States.

Chapter 3 is under review:

> House, B.R., Silk, J.B., Lambeth, S., Schapiro S.J., under review. *Task Design Influences Prosociality in Captive Chimpanzees (Pan troglodytes)*

VITA

March 24, 1980	Born, Charleston, SC
2003	BA, Cognitive Science Hampshire College
2009	MA, Biological Anthropology University of California, Los Angeles
2008 - 2010	Training Fellow NSF IGERT Interdisciplinary Relationship Science Program University of California, Los Angeles
2010	Teaching Assistant University of California, Los Angeles
2012-2013	Dissertation Year Fellow University of California, Los Angeles

PUBLICATIONS

Silk, J.B, House, B.R. (2011) From Social Bonds to Social Preferences: the Foundations of Human Moral Sentiments. *Proceedings of the National Academy of Sciences, 108*, 10910–10917.

Silk, J.B, House, B.R. (2012) The Phylogeny and Ontogeny of Social Preferences. In Vonk, J. and Shackelford, T. (Eds.), *The Oxford Handbook of Comparative Evolutionary Psychology*. (pp. 381–397).

House, B.R., Silk, J.B., Henrich, J., Brosnan, S.F. (2012) The Ontogeny of Human Prosociality: behavioral experiments with children aged 3 to 8. *Evolution and Human Behavior, 33*(4), 291-308.

Frankenhuis, W.E, House, B.R., Barrett, H.C., Johnson, S.P. (2013) Infants' perception of chasing. *Cognition*, *126*(2), 224-233.

House, B.R., Henrich, J., Sarnecka, B.W., Silk, J.B. (2013) The Ontogeny of Contingent Reciprocity in Children. *Evolution and Human Behavior*, *34*(2), 86-93.

House, B.R., Silk, J.B., Henrich, J., Barrett, H.C., Scelza, B., Boyette, A., Hewlett, B.S., Lawrence, S. (in press) The Ontogeny of Prosocial Behavior Across Diverse Societies. *Proceedings of the National Academy of Sciences.*

1

THE DEVELOPMENT OF CONTINGENT RECIPROCITY IN CHILDREN[1]

Introduction

Cooperation among kin is widespread in nature, but humans differ from most other species because we regularly cooperate with both relatives and non-relatives (Henrich and Henrich, 2007). Kin selection can lead to the evolution of prosocial behaviors that confer benefits on others that are related through descent from a common ancestor (Hamilton, 1964), but cannot account for cooperation between nonkin. Reciprocal altruism provides a mechanism for cooperation to evolve among pairs of nonrelatives (Axelrod and Hamilton, 1981; Trivers, 1971). Selection is expected to favor mechanisms that lead individuals to conditionally help others as long as the costs of helping are outweighed by the future benefits scaled by the likelihood of future interactions. For example, cooperation will be sustained if the benefits of cooperating are at least twice the costs, and if there is more than a 50% chance that interactions will be repeated. Reciprocal altruism requires individuals to keep track of past interactions in some way, assess the likelihood of future interactions, and condition their own behavior on the previous behavior of their partners (Axelrod and Hamilton, 1981; Trivers, 1971). Humans engage in contingent

---

[1] Originally published as:
House, B., Henrich, J., Sarnecka, B., Silk, J.B., 2013. The development of contingent reciprocity in children. Evolution and Human Behavior 34, 86–93.

cooperation in at least some settings (Gurven, 2006), but we know very little about how the capacity for contingent reciprocity develops as children mature. The goal of this paper is to fill this gap by exploring the development of contingent prosocial behavior in children using an experimental task similar to one previously used with captive chimpanzees. This affords insight both into the developmental trajectory of human reciprocity, and also the phylogeny of this behavior.

There is considerable evidence that humans condition their own cooperation on the cooperation of others. In some small-scale societies, individuals and family units transfer greater quantities of goods to those that previously transferred greater quantities to them (Bliege Bird et al., 2002; Gurven, 2006, 2004; Gurven et al., 2002, 2000). There is also evidence that these transfers are contingent on past behavior. Among the Ache, the quantity of food received by one family from another in one time period was positively related to the quantity of food given to the same family in a subsequent time period (Gurven, 2006).

Several studies have explored the development of reciprocal behavior in children (see Appendix 1). Fujisawa et al. (2008) studied naturally occurring interactions among children in 3-4 year-old Japanese children, and found that children's tendency to provide help and give objects (e.g. toys) to peers correlated with the peers' tendency to act prosocial towards them. Children were not given explicit instructions about how they should behave during these observations, so this study provides a good source of naturalistic data on reciprocity in children, but correlational data do not provide clear evidence of contingency in behavior.

Experimental studies allow a more explicit analysis of contingency. Testing pairs of American fourth graders Staub and Sherk (1970) allocated a number of candies to one child in each pair, and allowed them to transfer some to the other child or keep them all. Later, the children were allowed to draw pictures, but only one crayon was provided, and it was given to the child who was non-endowed previously. Children shared crayons more with children who had shared the most candy with them. Levitt et al. (1985) placed a barrier in the middle of a playroom to separate a pair of children aged 2.5-3 years, one of who was provided with a toy and instructed by their parent to pass the toy to the child on the other side of the barrier. Later in the session, the second child was provided with a toy, and in 9 out of 10 dyads this child only shared if the first child had shared before. These data suggest a contingency in children's willingness to share, but it is possible that children were responding to the adult's instructions to share, not the behavior of their partners. Fishbein and Kaminski (1985) had pairs of 6-11 year old American children play a game in which each player had the opportunity to help the other advance toward a goal. Children helped their partner (actually a stooge who had been trained to always help) about 68% of the time after their partner had helped them. However, subjects were less likely to reciprocate help if their partners had been instructed to help by the experimenter, than if their partners helped them without explicit instructions. This suggests that children condition their prosocial behavior on the perceived intentions of their partners, and on the actions and desires of adults, and raises concerns about the interpretation of results from studies in which children are instructed to share by their parents or other adults.

Birch and Billman (1986) endowed pairs of 3-5 year-old children (from the same school) with asymmetrical quantities of food (10 pieces vs. 1 piece). They then observed whether the 'rich' child shared with the 'poor' child. Of 14 children who received food when they were 'poor', 13 subsequently shared when they were 'rich'. However, of 13 children who had not received food when they were 'poor', only 7 shared later when they were 'rich'. This finding suggests a contingency between sharing and being shared with, but because children are not re-paired with the same child who shared with them before, their responses may be evidence either for generalized reciprocity (Barta et al., 2011) or for a norm psychology that is trying to learn relevant rules about sharing (Chudek and Henrich, 2011).

Dahlman et al. (2007) conducted a study in which children were paired with anonymous recipients, and played a series of three ‘games’. In each game, one child (the actor) was allowed to choose between two outcomes that had different payoffs for themselves and another child. Then, the recipients were informed of their decisions and were allowed to choose from the same set of options. Three to five year old children’s choices were not affected by the choices that their partners had made, but 6-8 year old children tended to match the previous behavior of their partners. However, the difference in the extent of reciprocity among the younger and older children was only significant in one of the three games, which has come to be known as the Prosocial Game (Fehr et al., 2008; House et al., 2012; Thompson et al., 1997). In this game, actors chose between one option that provided a reward to themselves and a reward to the other child, and a second option that provided a reward to the actor, but nothing to the other child.

These studies do not provide a clear picture of the development of contingent reciprocity as children mature. Observational evidence suggests that 3-4 year old children are most helpful to those that are most helpful to them, but correlational data do not provide evidence that children are using contingent behavioral strategies. Similarly, evidence that toddlers shared more with those who have previously shared with them is confounded by the fact that the children had been instructed to share. Fishbein and Kaminski (1985) found no effects of age on the reciprocal behavior of the 6-11 year old children that they tested, but it is not clear when contingent strategies first emerge. Moreover, most experimental studies have been limited to a single round of exchanges, and do not tell us whether children's behavior changes as they gain experience with the task and the behavior of their partners.

The current study is designed to examine the development of contingent reciprocity as children mature, but also to provide a direct comparison between the behavior of human children and that of non-human primates. Reciprocity is a plausible foundation for cooperation in non-human primates, raising additional questions about the phylogeny of the human reciprocity that we are exploring in the present study. Questions about phylogeny are best answered by comparing experimental data across humans and closely related primates. Surprisingly, experimental evidence for contingent reciprocity among our closest primate relatives, chimpanzees, is limited. de Waal (1997) found that chimpanzees were 6% more likely to share food with individuals that had groomed them within the past two hours than with individuals who had not groomed them within this period. Melis et al. (2008) found a weak tendency towards reciprocity in a task in which chimpanzees could help a familiar group member gain access to a food reward by unlocking a door. However, in a task in which chimpanzees could insert tokens into a vending machine that delivered a food reward to a conspecific in an adjacent enclosure, Yamamoto and Tanaka (2009a) found that individuals given free access to the apparatus didn't deliver many rewards to their partners or develop a contingent strategy. Similarly, Brosnan et al. (2009) presented pairs of familiar chimpanzees with a variant of the Prosocial Game in which one animal, the actor, could choose between two options: Option 1 delivered a food payoff to the actor and its partners, while Option 2 delivered a payoff only to the actor. Thus, Option 1 was prosocial (and equitable) and Option 2 was selfish (and inequitable). Prosocial choices were not costly to actors because they could not obtain higher payoffs by choosing the alternative outcome. Subjects alternated between playing the role of actor and recipient across trials. Actors' choices were not consistently affected by the choices of their partners in previous trials. Similar results were obtained in a subsequent study of chimpanzees using the same payoff distributions (Yamamoto and Tanaka, 2010). These methods can be easily adapted for use with children.

Following the procedures of Brosnan et al. (2009), in the current study we paired familiar children aged 3-7.5 years in face-to-face interactions and allowed them to interact repeatedly across multiple rounds in the Prosocial Game. Our results suggest that the propensity to respond in a contingent manner does not develop until about 5.5 years of age, but by this age the performance of children clearly differs from the performance of captive adult chimpanzees in a similar experimental setting.

Methods

Participants. Children were recruited at preschools near the University of California, Irvine. Children received a toy when parents signed the consent form, but at the time of testing children did not receive compensation for their participation beyond the payoffs obtained during the experiment. N=80 children (43 female) between the ages of 3 and 7.5 years (age 3-4: N=33, mean age=4.17, SD=.58; age 5-7.5: N=47, mean age=6.12, SD=.60). Pairs of children were about the same age, and usually drawn from the same class to emulate the methods of chimpanzee studies in which subjects are drawn from the same social groups. Pairs could be either same-sex or mixed-sex pairs, but were never composed of kin. Two participants were excluded from the analyses due to inattention or unwillingness to complete the experiment.

The Experimental Task. Children were seated across from one another on the floor, with the experimenter seated on one side. Two 8.5" x 14" cards were placed on the floor between the children (see Figure 1), and each card had one red circle and one blue circle printed on it. The experimental materials were based on Fehr et al. (2008). For each trial, payoffs were placed in the circles and one of the two children was permitted to choose one of the two cards (binary, forced choice). Payoffs were metal washers (described as "coins"), and children were told that one washer would be exchanged for one sticker at the end of the experiment. Children were only allowed to take the payoffs from the circle that was closest to them on the selected card.

Figure 1: Experimental Setup. Payoffs are individual washers (visible inside each circle below), each of which was exchanged for one sticker after the experiment was completed. In the example trial below the child on the left is the actor, and the child on the right is the recipient.

On each trial, one child was the actor and one child was the recipient. Actors were presented a choice between two options: (1) one washer for the actor and one for the recipient (the 1/1 option), or (2) one for the actor and nothing for the recipient (the 1/0 option). Actors and recipients alternated roles on successive trials. The children stored their payoffs in opaque paper bags that were provided to them, and later exchanged their payoffs for stickers.

Procedure. Experimenters first familiarized themselves with the children at the preschools by spending several hours at the school across multiple days. Children were approached and asked if they would like to play a game with the selected partner. Pairs were led to a quieter part of the school and seated across from each other. The full experimental session presented each child with two training trials and five test trials, for a total of 14 trials. Children alternated as actor and recipient during both training and test trials, and participants were told that they would alternate roles and have several turns in each role (see Appendix 2 for verbal instructions given to children). Children were not informed in advance of the exact length of the experiment, though a few inattentive pairs were informed when it was the last trial. After all testing in a particular classroom was completed, teachers were asked to complete a survey that rated the relationship quality of the pairs of children.

Training. Before each training trial, each child was given the full set of instructions, so each child heard the instructions four times. The first training trial presented the actor with a 1/1 vs. 2/2 choice, meaning that one card delivered only one payoff to each participant, while the second card delivered two payoffs to each participant. The second training trial presented actors with a 1/0 vs. 2/0 choice. These two trials were meant to introduce children to two facts about this game: payoffs obtained were influenced by the choices actors made, and recipients did not necessarily obtain payoffs. These two training trials were always presented in the same order, but the side of presentation for each payoff was counterbalanced across subjects.

Test. In each test trial, actors were presented with a choice between 1/1 and 1/0. Children were provided with no further instructions during test trials. Children were simply informed when it was their turn to play the actor role. Payoff options were counterbalanced so that half of the time the 1/1 was presented on the left, and half of the time it was presented on the right.

Coding. Current Choice was the primary dependent variable, and indicated the choice that an actor made on a focal trial. A choice of 1/1 was coded as '1' and a choice of 1/0 was coded as '0'. *Partner's Previous Choice* indicated the choice that an actor's partner had made on the trial immediately prior to the focal trial (a 1/1 choice was coded as '1,' a 1/0 choice was coded as '0'). *Sex* indicates the sex of the actor (female was coded as '1', male was coded as '0'), *Trial Number* indicates the trial number of the focal trial, and *Age* was the absolute age of the actor.

The covariate *Relationship Quality* was created by asking teachers to rate the strength of the pair's friendship. Teachers were provided with a 7-point likert scale (1="not friends at all"; 4="on average, as good friends as are most children"; 7="best friends"; ?="don't know"). We were able to collect ratings of relationship quality from 68 of our 80 subjects; 10 of the missing ratings were from the oldest children in our sample. As our sample of relationship quality is skewed toward younger children, we performed separate analyses on the subset of children for which relationship quality data were available.

Analyses. Each actor made binary choices between 1/1 and 1/0 payoff outcomes on four different trials. We used multi-level logistic regressions with 'actor identity' as a random effect, controlling for each subject contributing multiple data points. Models for *Current Choice* explore whether actors' choices on focal trials are predicted by their partners' previous choices (*Partner's Previous Choice*), actors' experience within the experiment (*Trial Number*), demographic information (*Age* and *Sex*), and dyadic relationship quality (*Relationship Quality*). Results are presented as Odds Ratios (ORs).

We hypothesized that *Partner's Previous Choice* would predict *Current Choice*, a result consistent with reciprocal altruism. An OR greater than 1.00 would indicate reciprocity by showing that a prior choice (either 1/1 or 1/0) predicts a greater likelihood of the same choice on the subsequent trial. Effects of age and sex have been reported in other studies of prosocial behavior in children (Eisenberg et al., 2006; Silk and House, 2012), with females and older children being more prosocial than males and younger children, so we explored whether *Age* or *Sex* would predict *Current Choice.* An OR greater than 1.00 would indicate that females are more likely to choose 1/1 than are males, or that older children are more likely to choose 1/1 than are younger children (while an OR below 1.00 indicates the opposite). Game theory predicts an "endgame effect" for the last round of an iterated game, because as the game comes to an end individuals should be indifferent to the past behavior of others and act in their own self-interest because there are no future benefits to be obtained by cooperating (Normann and Wallace, 2004; Selten and Stoecker, 1986). We provided no explicit information about when the interaction would end, but children might expect that each subsequent trial had a greater probability of being the last, and might therefore have chosen 1/1 less often as the experiment progresses. An OR below 1.00 for the variable *Trial Number* would suggest an endgame effect by showing that children were less likely to choose 1/1 as the experiment progressed.

We also explored interactions between *Partner's Previous Choice* and *Age*, *Sex*, and *Trial Number* using the interaction terms: *Age X Partner's Previous Choice, Sex X Partner's Previous Choice*, and *Trial Number X Partner's Previous Choice*. We predicted that *Partner's Previous Choice* would interact positively with *Age* (i.e. older children would be more reciprocal than younger children), with an OR greater than 1.00. Endgame effects should also lead to a negative interaction between *Partner's Previous Choice* and *Trial Number* (i.e. an OR less than 1.00), again because children might expect that each subsequent trial has a greater probability of being the last, and thus become more indifferent to the prior behavior of their partners. We had no strong predictions about whether *Partner's Previous Choice* would interact with *Sex*.

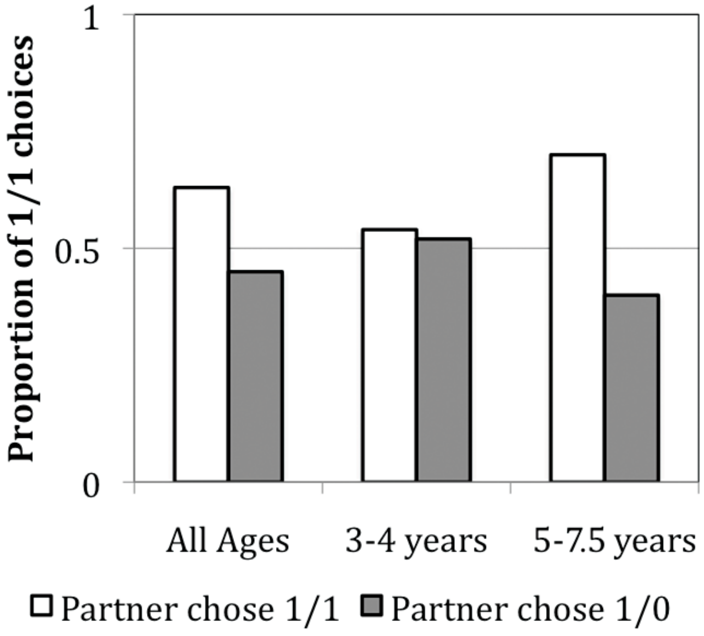
To determine how well these factors (*Partner's Previous Choice, Age, Sex, Trial Number, Age X Partner's Previous Choice, Sex X Partner's Previous Choice,* and *Trial Number X Partner's Previous Choice)* fit the data, using Akaike weights (Burnham and Anderson, 2002; McElreath et al., 2008) we calculated the probability that each of these factors would be present in the model that best fits the data (for more details see Appendix 3). This is an independent measure of how important a particular factor is across different model structures.

Relationship Quality: We had no clear hypotheses about how *Relationship Quality* would predict *Current Choice*, though children were expected to be more prosocial (i.e. more likely to choose 1/1) when paired with closer friends. We also had no predictions about whether *Relationship Quality* would interact with *Partner's Previous Choice* (i.e. whether relationship quality predicted reciprocity), because closer friends might be more likely to interact in the future and thus more reciprocal, but prior studies also suggest that friends may be less likely to immediately reciprocate than non-friends (Silk, 2003). Analyses for *Relationship Quality* were performed separately from the other analyses because the sample from which we received relationship ratings was smaller, and skewed towards younger ages. The procedures for these analyses are identical to those used above, except that we used a reduced number of factors and thus consider fewer models in our analyses (31 models, instead of 127; see Appendix 3).

Results

Across all ages, children chose the 1/1 outcome on 63% of trials in which their partner had previously chosen 1/1, and on 45% of the trials in which their partner had previously chosen 1/0 (Figure 2). Older children are primarily responsible for this pattern. A partner's previous choice had little impact on the behavior of 3-4 year-olds. However, 5-7.5 year-olds chose 1/1 on 70% of trials in which their partners had chosen 1/1, but on only 40% of trials in which their partner had chosen 1/0 (Figure 2). These aggregate data do not control for the non-independence in the data, and we use multi-level logistic regressions to confirm and extend these results.

Figure 2: *Data across all trials for all ages combined, children aged 3-4, and children aged 5-7.5. White bars denote the proportion of 1/1 choices that children made when their partner chose 1/1 on the previous trial. Grey bars denote the proportion of 1/1 choices that children made when their partner chose 1/0 on the previous trial. These data do not control for the fact that individual children were observed multiple times, and for this reason we do not include confidence intervals. See Figure 3 for appropriate confidence intervals.*



First we present the results for regression models of our main effects as odds ratios (*Partner's Previous Choice, Age, Sex,* and *Trial Number*), followed by the results of models including interaction terms (*Age X Partner's Previous Choice, Sex X Partner's Previous Choice*, and *Trial Number X Partner's Previous Choice*). We also present the results of our Akaike weight analyses for each factor in turn, which gives an indication of how important each factor is for interpreting these data.

Model 1 reveals an odds ratio larger than 1.00 for *Partner's Previous Choice* (Table 1), indicating that across all subjects actor's choices of 1/1 are positively predicted by their partner's choices of 1/1 on the previous trial. However, the probability that *Partner's Previous Choice* appears in the best model is relatively low, suggesting that other factors have an important impact on children's behavior in this task.

In Model 2 *Age* displays an odds ratio slightly smaller than 1.00, suggesting that older children do not chose 1/1 more frequently than younger children. *Age* also has a relatively low probability of appearing in the best model. Similarly, in Model 3 *Trial Number* has an odds ratio slightly smaller than 1.00, indicating that children chose 1/1 less frequently as the experiment progressed. The probability of appearing in the best model is higher for *Trial Number* than it is for *Age*, but it is still relatively low. Thus, both age and progress through the experiment are factors that do not strongly predict children's choices of 1/1 on their own, and they are not the most important factors for understanding children's behavior in this task.

In Model 4, *Sex* has an odds ratio greater than 1.00, indicating that females are more likely to choose 1/1 than are males. *Sex* also has a high probability of being included in the best model.

Table 1: Models for the Current Choice, for the full sample. The probability that each factor appears in the best model (out of all 127 models considered) is calculated by summing the Akaike weights for all models that include that factor. Factors with probabilities closest to 1 are the factors most likely to explain the data well, irrespective of exact model structure. Each model provides odds ratios and standard errors for each factor that has been included in the model. Odds ratios larger than 1.00 indicate that the parameter predicts a higher probability of choosing 1/1, while odds ratios less than 1.00 indicate that the parameter predicts a lower probability of choosing 1/1. The last row provides the estimates for the random effect (child id), which are presented as coefficients instead of odds ratios. For each model this parameter's coefficient is substantially larger than the standard error, indicates substantial differences across individual subjects in how they behave in this task.

DV: Current Choice	Probability that variable appears in the best model	Models						
		1	2	3	4	5	6	7
		Odds Ratio (St Err)	Odds Ratio (St Err)	Odds Ratio (St Err)	Odds Ratio (St Err)	Odds Ratio (St Err)	Odds Ratio (St Err)	Odds Ratio (St Err)
Partner's Previous Choice	.44	2.58 (.69)				.15 (.20)	.58 (.72)	3.35 (1.35)
Age	.38		.99 (.15)			.73 (.14)		
Trial Number	.54			.91 (.05)			.86 (.07)	
Sex	.77				1.83 (.60)			2.42 (1.05)
Age X Partner's Previous Choice	.73					1.71 (.41)		
Trial Number X Partner's Previous Choice	.41						1.15 (.13)	
Sex X Partner's Previous Choice	.34							.66 (.36)
Random effect parameter (child ID)		.88 (.24)	1.00 (.23)	1.03 (.23)	.96 (.23)	.81 (.24)	.88 (.24)	.80 (.24)

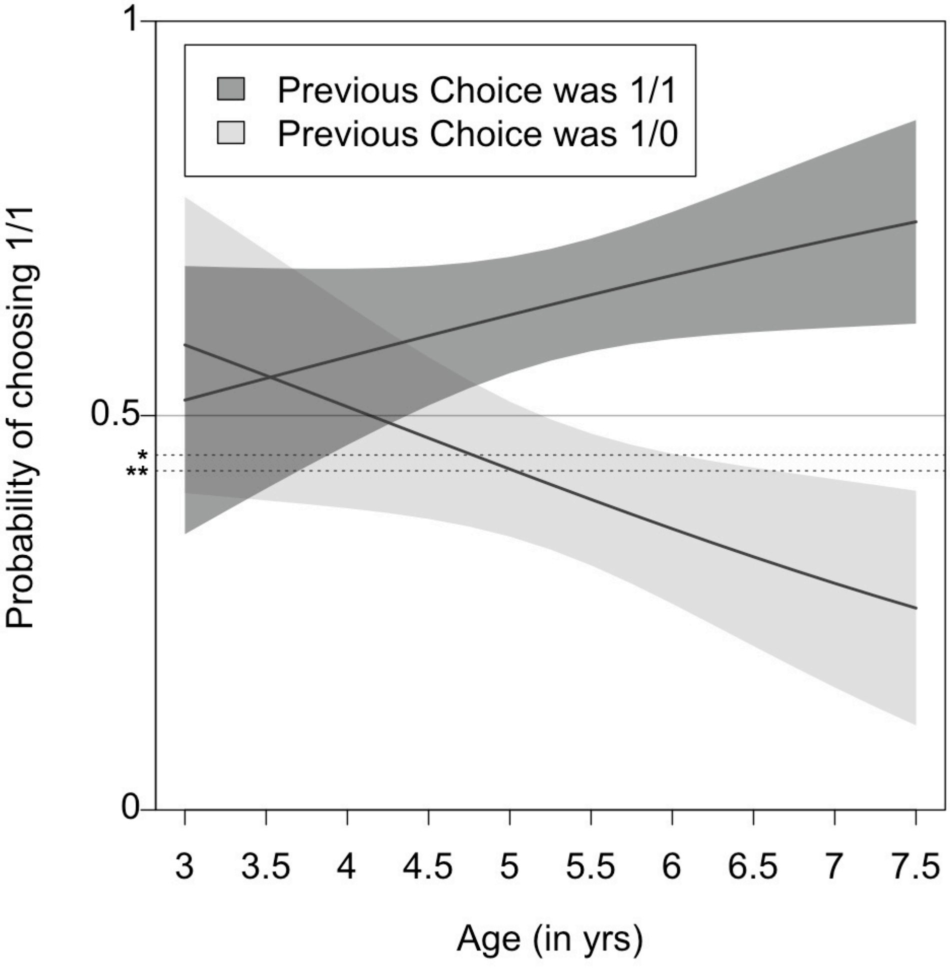
Model 5 suggests that children become more reciprocal with age, as there is an odds ratio larger than 1.00 for the *Age X Partner's Previous Choice* interaction (Table 1). This indicates that with each one year increase in age, children are 1.71 times more likely to choose 1/1 if their partner had previously chosen 1/1. Additionally, the odds ratio for *Age* in Model 5 is smaller than 1.00, indicating that for each one year increase in age, children are 1.37 times more likely to choose 1/0 if their partner had previously chosen 1/0. The probability that *Age X Partner's*

Previous Choice is included in the best model is relatively large, suggesting that this interaction is much more important for understanding children's behavior in this task than is *Partner's Previous Choice* on its own. These results are illustrated in Figure 3, which displays two logistic functions obtained from applying Model 2 independently to the trials in which the actor's partner previously chose 1/1 and 1/0 (Figure 3 is also a representation of the interaction between *Partner's Previous Choice* and *Age* in Model 5). These two samples of data are best modeled by two different functions: one indicating that the probability of actor's choices of 1/1 increase with age (when their partner's previous choice was also 1/1), and one indicating that the probability of actors' choices of 1/1 decrease with age (when their partner's previous choice was 1/0). For comparative purposes, Figure 3 also plots the mean rates of chimpanzees' 1/1 choices after their partner previously chose 1/1 and 1/0, as reported by Brosnan et al. (2009).

In Model 6, the odds ratio for the interaction between *Trial Number* and *Partner's Previous Choice* indicates that as actors progressed through the experiment they became more likely to match their partner's previous choices (Table 1). However, the magnitude of the coefficient and the probability that this factor appears in the best model are both relatively small, suggesting that it is not very important for explaining children's behavior. The odds ratio for the interaction between *Sex* and *Partner's Previous Choice* in Model 7 is smaller than 1.00, indicating that males are more likely to reciprocate their partner's choices than are females, but the relatively large standard error for this coefficient implies that this effect is not very consistent. Supporting this interpretation, the probability that *Sex X Partner's Previous Choice* appears in the best model is relatively low.

Figure 3: Graphical representation of logistic function from Model 2 from Table 2, as independently applied to trials that were preceded by a partner's 1/1 choice (dark grey) and trials that were preceded by a partner's 1/0 choice (light grey). The x-axis represents children's age, and the y-axis represents the model's prediction about children's probability of choosing the 1/1 outcome. The dark and light grey regions denote estimated 95% confidence intervals for the logistic function. Children are estimated to reciprocate 1/1 choices by their partner more than 50% of the time by age 4.5 years, and to reciprocate 1/0 choices more than 50% of the time by 5.5 years. In contrast, chimpanzees never reciprocated their partner's choices more than 50% of the time (Brosnan et al., 2009). Differences between the behavior of children and chimpanzees can be estimated by determining where the confidence intervals no longer overlap with the dotted lines.

** Probability that chimpanzees choose 1/1 when their partner chose 1/1. By 4 years of age children reciprocate 1/1 choices by their partner more than do chimpanzees.*
*** Probability that chimpanzees choose 1/1 when their partner chose 1/0. By 7 years of age children reciprocate 1/0 choices by their partner more than do chimpanzees.*



Relationship Quality: We obtained ratings of *Relationship Quality* for 85% of the dyads, and the majority of these ratings were for younger children. However, the patterns in this sample (Table 2) generally resemble those in the full sample (Table 1). The odds ratios for *Partner's Previous Choice* (Model 8) and *Age X Partner's Previous Choice* (Model 12) are again greater than 1.00, indicating that actors tend to reciprocate the previous choices of their partners, and that this tendency increases as a function of age. However, the odds ratio for *Age X Partner's Previous Choice* in Model 12 is reduced (relative to the odds ratio in Model 5), as is the probability that this factor is included in the best model. In Model 10, the odds ratio for *Relationship Quality* is greater than 1.00 and indicates that actors were more likely to choose 1/1 when they were paired with closer friends. The high probability that this factor is included in the best model suggests that relationship quality has an important impact on prosocial behavior. Including both *Relationship Quality* and *Partner's Previous Choice* in Model 11 only moderately reduces both odds ratios, suggesting that these are largely independent effects. Interestingly, relationship quality also doesn't appear to be positively related to reciprocity in several experiments with captive chimpanzees (see Brosnan et al., 2009).

Model 13 then tests for an interaction between *Relationship Quality* and *Partner's Previous Choice*, which asks whether close friends are more influenced by a partner's previous choices than others. The odds ratio is larger than 1.00 but smaller than the standard error, suggesting a weak effect, and the low probability of being included in the best model suggests this factor is not nearly as important as is *Relationship Quality* on its own.

Table 2: Models for Current Choice, for the sample rated for relationship quality. The probability that each factor appears in the best model (out of all 31 models considered) is calculated by summing the Akaike weights for all models that include that factor. Factors with probabilities closest to 1 are the factors most likely to explain the data well, irrespective of exact model structure. Each model provides odds ratios and standard errors for each factor that has been included in the model. Odds ratios larger than 1.00 indicate that the parameter predicts a higher probability of choosing 1/1, while odds ratios less than 1.00 indicate that the parameter predicts a lower probability of choosing 1/1. The last row provides the estimates for the random effect (child id), which are presented as coefficients instead of odds ratios. For each model this parameter's coefficient is substantially larger than the standard error, indicates substantial differences across individual subjects in how they behave in this task.

DV: Current Choice	Probability that variable appears in the best model	Models					
		8	9	10	11	12	13
		Odds Ratio (St Err)	Odds Ratio (St Err)	Odds Ratio (St Err)	Odds Ratio (St Err)	Odds Ratio (St Err)	Odds Ratio (St Err)
Partner's Previous Choice	.40	1.88 (.55)			1.72 (.51)	.26 (.37)	1.17 (1.12)
Age	.28		1.09 (.18)			.87 (.19)	
Relationship Quality	.71			1.42 (.21)	1.36 (.19)		1.30 (.24)
Age X Partner's Previous Choice	.51					1.47 (.40)	
Relationship Quality X Partner's Previous Choice	.45						1.11 (.26)
Random effect parameter (child ID)		.93 (.25)	1.02 (.25)	.92 (.25)	.86 (.26)	.87 (.26)	.85 (.26)

Discussion

These results demonstrate contingent prosocial behavior in our sample of American 3-7.5 year-olds, with older children being more likely to match the behavior of their partners than younger children. The models predict that in a similar sample by about 4.5 years children will choose 1/1 more than half the time when their partner chose 1/1 during the previous round, and by about 5.5 years children will choose 1/1 less than half of the time when their partner chose 1/0 previously (Figure 3). This suggests that positive reciprocity develops slightly ahead of negative reciprocity, but it is also possible that children simply had a baseline bias towards the prosocial outcome making it appear as though positive reciprocity emerges earlier. Conclusions about the separate ontogenies of positive and negative reciprocity will require further study.

The behavior of human children differs substantially from the behavior of adult chimpanzees in this task. By age 5.5, children reciprocated both 1/1 and 1/0 choices by their partners significantly more than 50% of the time, while chimpanzees never did so. However, it would be premature to conclude that there are differences in the capacity for contingent reciprocity among chimpanzees and human children. There is correlational evidence for reciprocity in grooming and food sharing among wild chimpanzees (Mitani, 2006), and it is possible that reciprocity among chimpanzees is poorly captured by laboratory tasks like this one (see also Melis et al., 2008). Moreover, although we modeled our experiment after Brosnan et al. (2009), the procedures were not identical. For example, the children received verbal instructions, while the chimpanzees did not, receiving numerous training trials instead. It is also possible that developing in captivity has cognitive or behavioral consequences for chimpanzees that makes the behavior of captive animals a poor model for the behavior of wild animals (Boesch, 2008, 2007; but see Tomasello and Call, 2008).

Regardless, our results clearly indicate that humans and chimpanzees differ in how reciprocity shapes their social interactions in a similar context, and this enhances our understanding of the constraints on the development of contingent reciprocity in humans and other animals. Understanding these constraints is necessary for understanding the mechanisms that underlie cooperation across species.

Developmental Effects on Contingent Reciprocity. Our results indicate that children begin to respond contingently when they are about between 4.5 and 5.5 years of age. Unfortunately, few other studies of the development of contingent cooperation span this age range within a single experimental context, making it hard to compare our results with the results from other studies. The correlational study showing that 3-4 year old Japanese children selectively share and help those that most often share and help them suggests that children may practice contingent strategies by this age (Fujisawa et al., 2008)—though contingency is not actually shown. In contrast, the 3-4 year-olds that we tested did not condition their behavior on the previous behavior of their partners. Differences in methodology make it difficult to compare these results directly, but raise a number of possibilities. First, as noted earlier, it is possible that the patterns observed among the Japanese preschoolers are not the product of contingent reciprocity. Second, it is possible that contingent behavioral strategies emerge earlier in naturalistic, everyday settings than in more artificial experimental settings. Third, cultural differences may produce different developmental trajectories among children in the US and Japan.

Birch and Billman (1986) found that 3-5 year old children were more likely to share with others if they had previously been the recipients of others' generosity than if they had not been the recipients of generosity. However, it is not clear whether the youngest children were as likely to "pay it forward" as the oldest children that they tested. Our results are also consistent with Fishbein and Kaminski (1985) finding that 6-11 year olds respond contingently to the behavior of their partners.

Our results are also consistent with the results of Dahlman et al. (2007) who found that 6-8 year-olds were significantly more likely to respond contingently to the behavior of anonymous partners in the Prosocial Game than 3-5 year olds. It is not clear, however, how anonymity influences children's likelihood of reciprocating, so the parallels in the results must be viewed with some caution.

Effects of Sex. In this experiment, females were generally more likely to choose the prosocial option than males, but there was no effect of sex on the likelihood of reciprocation. In other words, females were more likely than males to choose 1/1 when their partner chose 1/1 but also when their partner chose 1/0. This pattern is largely consistent with findings from the literature. Many studies of the development of prosocial behavior have reported that females are more prosocial than males (Fabes and Eisenberg, 1998). In Dictator Games conducted with children, females are more likely than males to donate some amount, and more likely to donate larger amounts (Blake and Rand, 2010; Gummerum et al., 2010, 2008; Harbaugh et al., 2003; Leman et al., 2009). However, sex differences do not emerge in all experimental economic studies conducted with children (Benenson et al., 2007; Harbaugh and Krause, 2000; Lucas et al., 2008; Sally and Hill, 2006; Takezawa et al., 2006). There is little evidence of sex differences in children's reciprocal behavior. (Sutter and Kocher, 2007) found no effects of sex in an anonymous trust game played with subjects aged 8 years to adult. (Dreman and Greenbaum, 1973) found that male subjects, but not female subjects, responded to anonymity by becoming less prosocial. This might suggest an effect of sex on sensitivity to anonymity, but only indirectly suggests a possible sex difference in contingent responses.

Effects of Trial Number. Endgame effects are commonly found in repeated games as rates of cooperation drop as the game progresses toward the last rounds (Normann and Wallace, 2004; Selten and Stoecker, 1986). However, we found little evidence for endgame effects among the children that we tested. Although trial number negatively predicted prosocial behavior (Model 3, Table 1) and positively predicted reciprocity (Model 6, Table 1), the effects of trial number are very weak, and the magnitude of these effects is very small. Moreover, if the weak negative effect of trial number on prosocial choices were evidence of an endgame effect, then we would predict that trial number would also have a negative effect on reciprocity as children become less sensitive to the previous behavior of their partners in the last rounds of the game. Instead, we found that reciprocity increases as the experiment progresses, and on the final trial children reciprocate 1/1 choices 81% of the time (SE=10). This suggests that children are becoming more, not less, reciprocal as they gain experience with the task. Thus, trial number may negatively predict children's probability of choosing 1/1 because they are becoming more inclined to punish selfish behavior by others, not because they are becoming less prosocial due to an endgame effect. Regardless, effects of trial number are substantially weaker than effects of previous choices and actor sex, suggesting that children enter the task already endowed with reciprocal strategies and their responses change little over the course of the experiment.

Effects of Relationship Quality. One might assume that relationship quality is associated with the likelihood of future interactions, and thus stronger relationships should predict higher rates of reciprocity. However, empirical studies of friendship among adults (at least in the West) show that friends are less likely to immediately reciprocate a prosocial act than are non-friends, and immediate repayment by a friend can even viewed negatively, perhaps explaining why friends also sometimes go to the trouble of concealing prosocial acts (Silk, 2003). Interestingly, studies with non-human primates also suggest that reciprocity might be stronger across longer timescales (Jaeggi et al., 2013; Schino and Aureli, 2009). The reasons for such behavior among humans aren't fully understood. It is possible that short-term bookkeeping within a relationship is avoided because it implies that future cooperative interactions are unlikely. Alternatively, a long history of reciprocity within a dyad may reduce the relative value of any particular cooperative action, thus reducing the relative costs of that act and the need to reciprocate small prosocial acts. Regardless of their cause, these patterns among adults fit with our finding that children paired with close friends are typically more prosocial, but not more reciprocal, than those paired with non-friends. Importantly, we also show that effects of relationship quality are distinct from the effects of partner's previous choices, meaning that our evidence for reciprocity in children's behavior is not simply due to friends being highly (but non-contingently) prosocial.

Our analyses suggest that at least some of the fundamental characteristics of (Western-style) friendship in adults also describe friendships among children, as measured by third-party adult raters. This points to a paradigm for exploring the development of the dynamics of friendships among children in a systematic way, a topic that has not been investigated in much detail.

Summary

Despite considerable evidence for reciprocity in human social behavior, we do not fully understand the ontogenetic development of contingent reciprocity in humans. Our results demonstrate the emergence of contingent prosocial behavior and are largely consistent with the limited developmental literature on reciprocity that suggests that reciprocity is predicted by child age but not by child sex (although sex does predict prosocial behavior more generally). The current study also adds to our understanding of the phylogeny of human reciprocity, by illustrating that within similar experimental contexts children engage in contingent prosocial behavior (as do human adults) but captive adult chimpanzees do not. This suggests differences in the reciprocal strategies of humans and our closest living relatives, but the source of these differences will require further systematic study of the conditions under which reciprocity is elicited in both species. Our findings suggest that reciprocity develops in American children by 5.5 years of age within this experimental context. These results provide a useful foundation for future work that explores the nature of this developmental process, and sets the stage for more focused tests of how cognitive changes and cultural acquisition influences contingent reciprocity. Both are necessary for fully understanding the developmental processes that underlie human cooperation, and for understanding how human cooperation differs from that of our close primate relatives.

1 APPENDIX

THE DEVELOPMENT OF CONTINGENT RECIPROCITY IN CHILDREN

Appendix 1

Table S1: Design features of current experiment, and previous experiments of reciprocity in children and chimpanzees.

Study	Population	Age (in years)	Choices were discrete	Rewards	Relationship of partners	Partners were face-to-face
Brosnan et al. (2009)	Captive chimpanzees	n/a	Yes	Food	Social group member	Yes
Current Study	American children	3-7.5	Yes	Stickers	Schoolmate	Yes
Dahlman et al. (2007)	Swedish children	3-8	Yes	Food	Schoolmate	No (anonymous partners)
Levitt et al. (1985)	American children	2.5-3	No	Access to toys	no data	Yes
Fujisawa et al. (2008)	Japanese children	3-4	n/a	n/a	Schoolmate	Yes
Birch and Billman (1986)	American children	3-5	No	Food	Schoolmate	Yes
Fishbein and Kaminski (1985)	American children	6-11	No	Advancement in a board game	Unfamiliar (partner was a stooge)	Yes
Staub and Sherk (1970)	American children	~ 9	No	Food; Access to crayons	Schoolmate	Yes

Appendix 2: Instructions

Children were given the following instructions about how to play the game:

"This game is called the turn-taking game. In this game, (child A's name) will get a turn, and then (child B's name) will get a turn, and then (child A's name) will get another turn, and then (child B's name) will get another turn, and the game will keep going like that."

"[speaking to actor] Every time we play, you will get to choose one of these two cards [experimenter touches each card], but you can only choose one.

"If you choose this card [experimenter touches the card closest to them], you get what's in this blue circle [experimenter points to blue circle] and [recipient's name] gets what's in this red circle [experimenter points to red circle]. *(note: colors may be reversed depending on the child's side)*

"But, if you choose this card [experimenter touches the card furthest from them], you get what's in this blue circle [experimenter points to the blue circle] and [recipient's name] gets what's in this red circle [experimenter points to red circle].

"Also, anything that is on the card that you don't choose goes back in the box [experimenter points to a box containing the washers to be used as payoffs]."

Appendix 3: AIC weights

To determine how well these factors explained our data we created 127 models containing all possible combinations of the seven factors: *Partner's Previous Choice, Age, Sex, Trial Number, Age X Partner's Previous Choice, Sex X Partner's Previous Choice,* and *Trial Number X Partner's Previous Choice*. For our relationship quality analyses we used 31 models combining 5 factors: *Partner's Previous Choice, Age, Relationship Quality, Age X Partner's Previous Choice,* and *Relationship Quality X Partner's Previous Choice.*

We determined how well each of these models fit the data, assuming that factors which frequently appear in the models that best fit the data are most important for understanding children's behavior in this experiment. By evaluating how factors fit the data across these models, we are able to determine whether effects are robust to changes in model structure. This gives us confidence in the importance of particular factors, above and beyond the magnitude and significance of regression coefficients or Odds Ratios associated with particular factors in particular models. The procedure is not biased by a priori predictions about which factors would be important, and thus diminishes any concerns that our results are specific to the particular models that we report.

We determine the goodness of fit of each model using Akaike weights (Burnham and Anderson, 2002; McElreath et al., 2008), which are calculated using the Akaike Information Criterion (hearafter AIC; Akaike, 1973). AIC values incorporate an important trade-off between a model's goodness of fit and the number of parameters that model includes. Adding more parameters inevitably improves the fit of a model, so the 'best' model is the one that displays the best fit to the data with the fewest number of parameters. The model with the lowest AIC is interpreted as the ‘best’ model.

Akaike weights can be thought of as probabilities: the Akaike weight for a model is the probability that this model is the best model out of the population of models being considered. Summing the Akaike weights for all models that include a particular factor generates the probability that the best model (out of those considered) includes this factor.

2

THE ONTOGENY OF PROSOCIAL BEHAVIOR ACROSS DIVERSE SOCIETIES[2]

Introduction

Human cooperation poses one of the great problems of the human sciences (Axelrod and Hamilton, 1981). While all human groups are generally more cooperative than other primates (Silk and House, 2011), the extent and scale of cooperation varies across societies, behavioral domains, and through historical time (Chudek and Henrich, 2011). Evolutionary researchers generally agree that kinship and reciprocity have shaped our inclinations to help others in important ways, but disagree about whether these evolutionary processes are sufficient to explain the levels of cooperation in contemporary human societies or to account for the diversity in cooperation across societies (Richerson and Boyd, 2005). Some researchers have argued that the psychological mechanisms that underlie our concern for the welfare of others and motivate helpful (prosocial) behavior have been shaped by an interaction between genes and culture (Chudek and Henrich, 2011; Henrich et al., 2010a, 2006, 2005; Richerson and Boyd, 2005). In

---

[2] Chapter 2 is in press:

House, B.R., Silk, J.B., Henrich, J., Barrett, H.C., Scelza, B.A., Boyette, A.H., Hewlett, B., McElreath, R., Laurence, S., in press. *The Ontogeny of Prosoical Behavior across Diverse Societies*. Proceedings of the National Academy of Sciences of the United States.

this account, gene-culture co-evolution has shaped social norms, which are acquired during development, and subsequently influence individual behavior and social preferences in ways that affect the relative success of competing social groups (Chudek and Henrich, 2011).

Many models predict that there will be substantial variation in prosocial behavior across social groups, but gene-culture co-evolutionary models make two distinctive predictions about this variation: (i) variation in prosocial behavior across groups will be more pronounced when the costs of cooperation (and incentives to defect) are higher, and (ii) this variation will emerge as children begin to acquire the social norms of their communities. To investigate these predictions, we conducted a study of 3 to 14 year-old children's prosocial choices in six different societies that collectively capture a substantial amount of human environmental and cultural variation. We demonstrate that population differences in costly prosocial behavior (but not non-costly prosociality) first emerge in middle childhood, suggesting that this is when children become sensitive to cultural influences that modulate their willingness to provide costly help to others.

There is substantial variation in the levels of cooperation across societies, and this is reflected in people's willingness to share monetary rewards with others and incur monetary costs to punish unfair behavior (Henrich et al., 2010a, 2006, 2005; Herrmann et al., 2008; Marlowe et al., 2008; Roth et al., 1991). Gene-culture coevolution predicts that population-level variation in sharing and punishment is linked to demographic and economic variables. Consistent with these predictions, individuals from larger societies are more willing to punish stinginess, and members of world religions and more market-oriented groups are more willing to share (Henrich et al., 2006; Marlowe et al., 2008). These relationships may be the result of selection for social norms and institutions (i.e., sets of norms) that maintain cooperation even under conditions in which cooperation has high costs and is thus fragile, as in large societies where many interactions are ephemeral and there are many opportunities for exchange among strangers or anonymous others (Henrich, 2006; Henrich et al., 2006; Marlowe et al., 2008). However, it is unlikely that cultural evolutionary forces generate identical norms and institutions across all societies. Instead, selective forces acting on cultural beliefs and motivations should maintain differences in cooperative behavior across populations that vary in how costly cooperation is used to overcome ecological and economic challenges. An important implication of this is that gene-culture coevolution predicts more systematic population-level variation in prosociality in contexts in which the costs of cooperation are relatively high, and less variation in contexts in which the costs of a prosocial act are lower.

It is not yet clear how population differences in prosocial behavior emerge during the life span. Many developmental studies suggest that prosocial behavior increases substantially across childhood (Eisenberg et al., 2006; Silk and House, 2012), but these findings are largely based on children in Western societies or urban settings ; but see Callaghan et al., 2011; Madsen and Lancy, 1981; Rochat et al., 2009 for some samples from small-scale non-Western societies). This constrains our ability to assess the extent of variation in children's prosocial behavior across populations and to examine the relationship between the cost of helping and the development of population differences in prosocial behavior. To overcome these limitations, we conducted a set of simple choice tasks in which children had the opportunity to deliver benefits to peers, but the personal costs of delivering benefits varied across trials. We conducted these studies with 326 3-14 year-old children in six different populations (Table 1; see Appendix 2 for age distributions), and 120 adults from five of the six populations. Participants were not immediate family members. This sample includes nomadic hunter-gatherers living in the Congo Basin (Aka), one of the last remaining societies of this kind, as well as semi-nomadic agro-pastoralists from Namibia (Himba), slash and burn horticulturalists from Amazonia (Shuar), sedentized foragers from Australia (Martu), marine forager-horticulturalists from Melanesia (Yasawa Island, Fiji), and urban Americans (Los Angeles). We investigated the effects of age and population of origin on children's responses when they were presented with opportunities to provide benefits to others.

*Table 1: Populations in the current study.**

Population	**Location**	**N (female)**		**Primary Subsistence**
		Children	**Adults**	
Los Angeles	United States	75 (34)	28 (23)	urban
Yasawa Island	Fiji	75 (33)	25 (10)	horticulture, marine foraging
Aka	Central African Republic	35 (13)	10 (6)	hunting/gathering
Himba	Namibia	82 (48)	32 (19)	pastoralism, horticulture
Shuar	Ecuador	37 (13)	25 (7)	horticulture
Martu	Australia	22 (10)	---	hunting/gathering

* See Section 2 for additional details about age distributions and demographics across samples.

We used a standardized task to assess the generosity of a subject Actor when it was personally costly for them to deliver rewards to a peer Recipient (Costly Sharing Game, CSG) and when it cost them nothing to confer benefits on Recipients (Prosocial Game, PG). This task was based on methods used by Thompson et al. (Thompson et al., 1997) and Fehr et al. (Fehr et al., 2008), along with several other studies (Brownell et al., 2009; House et al., 2013, 2012; Moore, 2009). Following previous work on chimpanzees (Jensen et al., 2006; Silk et al., 2005) and children (Brownell et al., 2009; House et al., 2012), we included a Social condition in which children were paired with a familiar peer in a face-to-face interaction and an Asocial condition in which no Recipient was present and rewards were not allocated to anyone except for the Actor (see Methods and Appendix 1 for details, and Appendix 1 for variations in methods across sites). If participants were prosocial, they were expected to deliver more rewards in the Social condition than the Asocial condition.

Both the Asocial and Social conditions included the same four choice tasks: two Familiarization (FAM) tasks followed by two Test tasks. In each of the four choice tasks Actors were presented with a pair of different payoff options to choose between (Option 1 and Option 2), where each pair differed in the distribution of payoffs to the Actor and Recipient.

In the Test trials, Actors chose between a prosocial outcome and a selfish outcome. In the CSG (a simplified Dictator Game), Actors chose between one option which provided two real and visible food rewards to themselves and nothing to their partner (2/0) and a second option which provided one reward to themselves and one reward to their partner (1/1). In the PG, Actors chose between one option that provided one reward for themselves and nothing for their partner (1/0) and another that provided one reward to both themselves and their partners (1/1).

The FAM trials were designed to provide Actors with the experience of Recipients obtaining a reward and the experience of Recipients obtaining nothing. In FAM1, Actors chose between one option which provided two rewards to both themselves and to their partner (2/2) and a second option which provided one reward to themselves and one reward to their partner (1/1). In FAM2, Actors chose between one option that provided two rewards for themselves and nothing for their partner (2/0) and another option that provided one reward to themselves and nothing to their partners (1/0). Thus, after both FAM trials Actors had seen a Recipient obtain a reward in one trial and not obtain a reward in another trial.

We can use Actors' choices in the FAM1 Social condition as a measure of participants' comprehension of the basic choice task by investigating the development of children's tendency to select 2/2 (the income-maximizing outcome) over 1/1. The other FAM trials (FAM1 Asocial, FAM2 Asocial, and FAM2 Social) cannot be used for such an analysis because they are potentially confounded by inequity aversion. Children who understand our choice task might choose 2/2 or 2/0 on these trials because it maximizes their income, but they might also choose 1/1 or 1/0 because it minimizes inequity between themselves and the Recipient. Due to this we focus only on FAM1 Social in our analyses because it is not confounded in this way (see Appendix 7 for more details and an analysis of all FAM trials).

Analyses and Results

Unlike chimpanzees tested in a similar version of the PG (Jensen et al., 2006; Silk et al., 2005), children and adults across societies differentiated between the Asocial and Social conditions in both the PG and CSG. In the PG, children chose the prosocial outcome (1/1) more in the Social condition when another child was present to receive rewards (mean number of trials[SE]=.60[.03]) than in the Asocial condition when no one was present to receive rewards (mean[SE]=.51[.03]). In the CSG, the likelihood of choosing the prosocial option (1/1) over the selfish option (2/0) was lower than the likelihood of choosing the prosocial option in the PG, reflecting the increased cost of prosocial behavior in this game. But in the CSG, children still chose the prosocial outcome more in the Social condition (mean[SE]=.34[.03]) than in the

Asocial condition (mean[SE]=.22[.02]). Similarly, adults also selected 1/1 more frequently in the PG Social (mean[SE]=.73[.08]) than in the PG Asocial (mean[SE]=.43[.09]), and more frequently in the CSG Social (mean[SE]=.57[.09]) than the CSG Asocial (mean[SE]=.33[.08]). See Appendix 3 for analysis of the effect of condition and population membership on adults' behavior.

To model the effects of children's age on the likelihood of choosing the prosocial option in the CSG and PG (and the income-maximizing option in FAM1), we centered participants' ages to create an age parameter called Centered Age (CA), and created a second age parameter by squaring CA (CA^2; see Methods section for details). Including both CA and CA^2 permits our regression models to create either monotonic or non-monotonic age functions, allowing for a range of developmental trajectories.

To investigate whether the development of children's choices in the CSG and PG varied across our conditions (Social and Asocial) and populations, we compared how well these choices were fit by a set of generalized linear multilevel logistic regression models (Table 2). These models either included a single age function (CA and CA^2) for all populations and both conditions (Model A), separate age functions for the Social and Asocial conditions but collapsed across populations (Model B), separate age functions for each population but collapsed across conditions (Model C), or separate age functions for each condition for each population (Model D). We also investigated whether the development of children's choices in FAM1 Social varied across populations, and thus focus on Models A & C for this task.

Table 2: Multilevel logistic regression models. Includes details about the Fixed Effect (FE) and Random Effect (RE) parameters included in each model, the developmental hypothesis reflected in each model, and the DIC and DIC weight values associated with each model when it is applied to the CSG, PG, and FAM1 Social. Best-fit models are in bold.

Model	**Parameters included**	**Hypothesis**	**CSG**	**PG**	**FAM1 Social**
			DIC (weight)	**DIC (weight)**	**DIC (weight)**
A	**FE:** CA, CA^2 **RE:** Actor ID	One overall developmental trajectory	765.12 (.00)	904.61 (.06)	321.53 (.05)
B	**FE:** CA, CA^2, Condition, CA*Condition, CA^2 *Condition **RE:** Actor ID	One trajectory for each condition	746.86 (.00)	**899.10** **(.94)**	
C	**FE:** CA, CA^2 **RE:** Actor ID Population | CA, CA^2	One trajectory for each population	747.30 (.00)	914.97 (.00)	**315.58** **(.95)**
D	**FE:** CA, CA^2, Condition, CA*Condition, CA^2 *Condition **RE:** Actor ID Population | CA, CA^2 Condition, CA*Condition, CA^2 *Condition	One trajectory for each condition, for each population	**734.49** **(1.00)**	916.52 (.00)	

All models for the CSG and PG include random effects for Actor identity (Actor ID), to compensate for the fact that each Actor contributed two observations (one from each of the Social and Asocial conditions). Where we modeled age functions for each condition we included a fixed effect parameter for Condition (Social coded as ‘1’; Asocial as ‘2’). As we focused only on the Social conditions for FAM1 we did not consider models that included the Condition parameter (Models B & D), nor did we include Actor ID as a random effect.

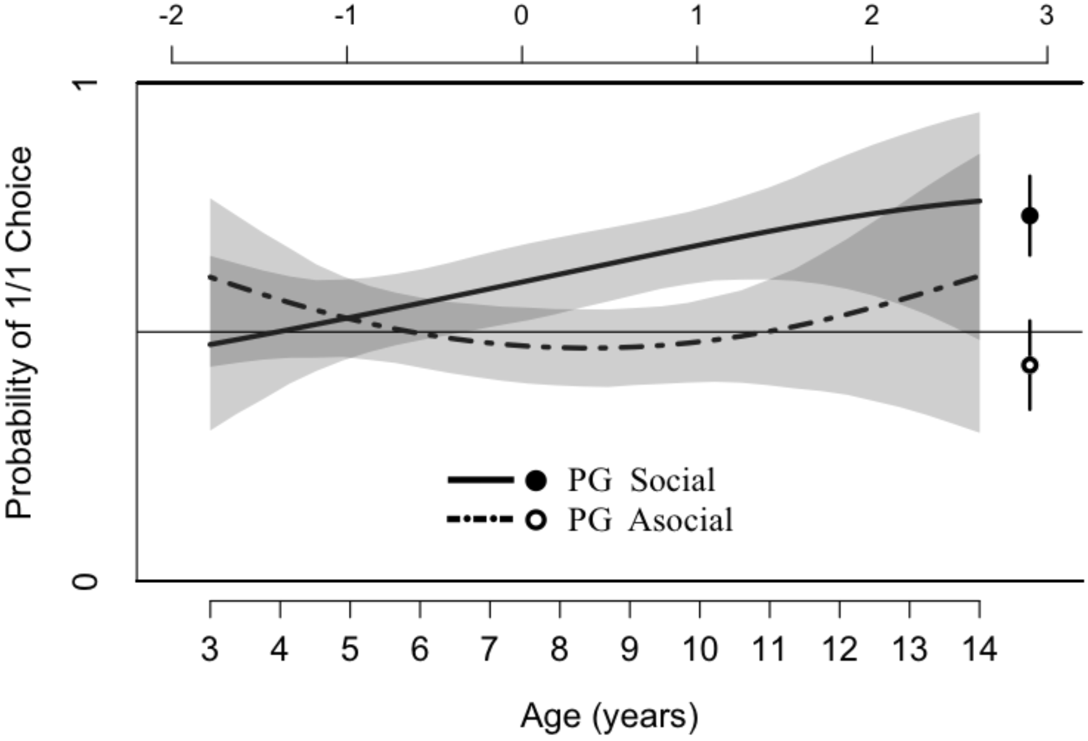
Where we modeled age functions for each population we included Population as a random effect, which estimates parameters for each population in relation to our entire dataset (rather than just the observations from that sample). By using a random effect analysis to estimate parameters for each population using the entire dataset these models are robust to variation in sample sizes across populations, and they also provide a conservative test of population differences by tending to shrink estimates for each population toward the grand mean of the data.

We determined model fit using DIC (Deviance Information Criterion; Lunn et al., 2012) and DIC weights, which adds a more appropriate penalty to the deviance of multilevel models than does AIC (Akaike Information Criterion; Burnham and Anderson, 2002). DIC estimates the out-of-sample prediction error of a model, by penalizing a model for its flexibility in fitting. As a result, smaller values of DIC indicate better-expected out-of-sample predictions (i.e. better predictions about new participants from our populations), and complex models must overcome a large penalty to be deemed better than simpler models. DIC weight is a transformation of DIC that can be thought of as the probability that an individual model is the best out of the set of models being considered. These values allow a group of models to be compared rather than requiring that individual models be accepted or rejected, and they permit the comparison of models with different structures without the concern that more complex models might appear better due to overfitting.

Model B has a DIC weight of .94 for the PG, meaning it has a .94 probability of being the best model for this task. This indicates that development very likely differed across the Social and Asocial conditions, but that differences across populations were likely small. In contrast, Model D is the best fit for the CSG, indicating that development likely varied substantially across the Social and Asocial conditions, and also across our six populations. Model C (without including Actor ID as a random effect) had a .95 probability of being the best model for FAM1 Social, suggesting that here, too, development varied substantially across populations. We obtain the same best-fit models when we include Actor Sex as a covariate in our regression models, and when we focus only on children aged 5-10 (Appendix 4, Table S4[b]).

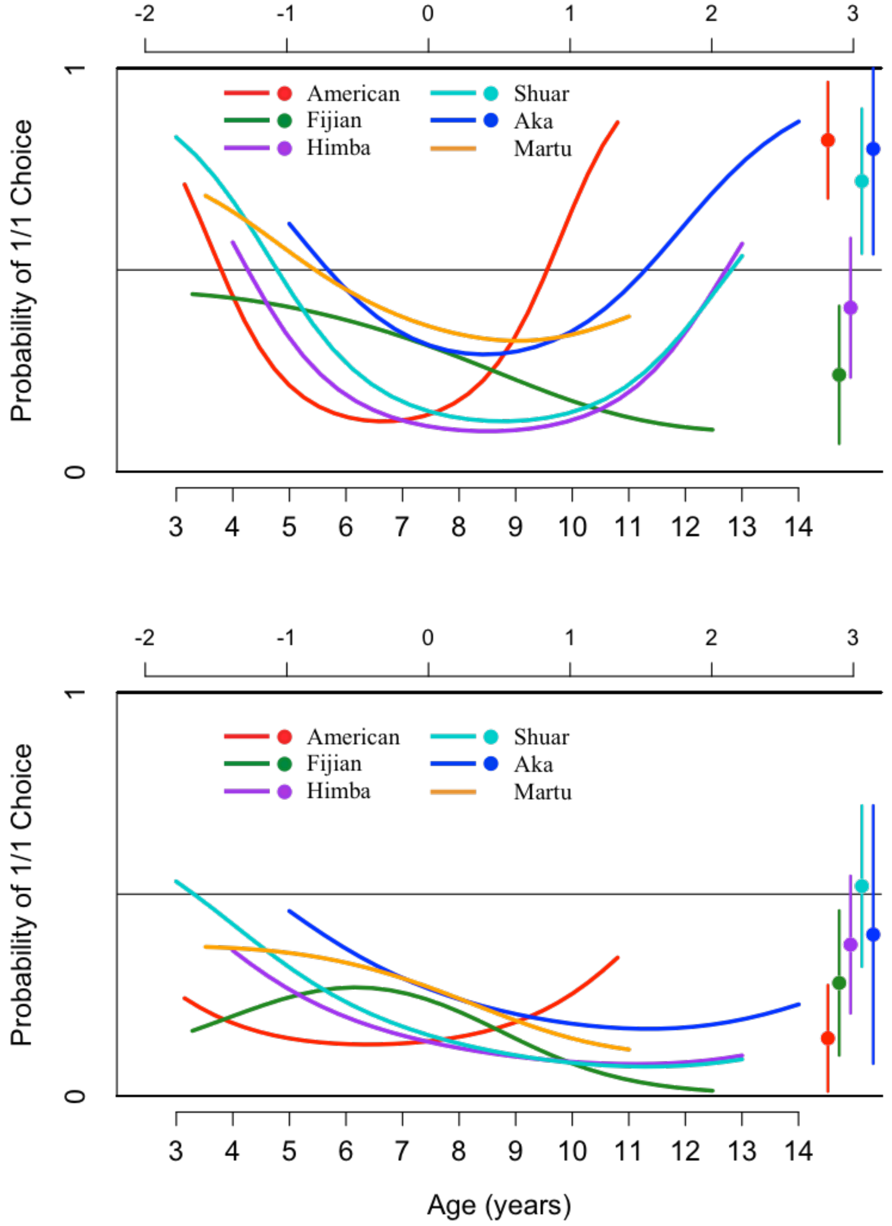
Figure 1 plots how Model B estimates the probability that children select the 1/1 outcome in the PG as a function of age, for both the Social and Asocial conditions. Figure 2 plots how Model D estimates the probability that children from different populations select the 1/1 outcome in the CSG as a function of age, for both the Social condition (Figure 2[a]) and the Asocial condition (Figure 2[b]). Figure 3 plots how Model C estimates the probability that children from different populations select the 2/2 outcome in FAM1 Social as a function of age. Fully specified models used to plot Figures 1-3 are provided in Appendix 4.

*Figure 1: Best-fit model of Actors' choices of 1/1 in the PG. Vertical axis is the estimated probability that children will choose the prosocial (1/1) outcome. Bottom horizontal axis is children's age (in years), and top horizontal axis is the equivalent value of CA. Age functions capture the estimated probability that children will select the 1/1 outcome as a function of age, with estimates extracted from the best-fit model for the PG (Model B, Table 2) for both the Social condition (solid line) and the Asocial condition (dotted line). Wide confidence intervals above age 13 are due to small samples above this age. The dot and hollow circle on the right side of the plot reflects the proportion of 1/1 choices actually made by adults in the PG Social and PG Asocial (respectively). The lines above and below the dot and circle correspond to 95% confidence intervals. **



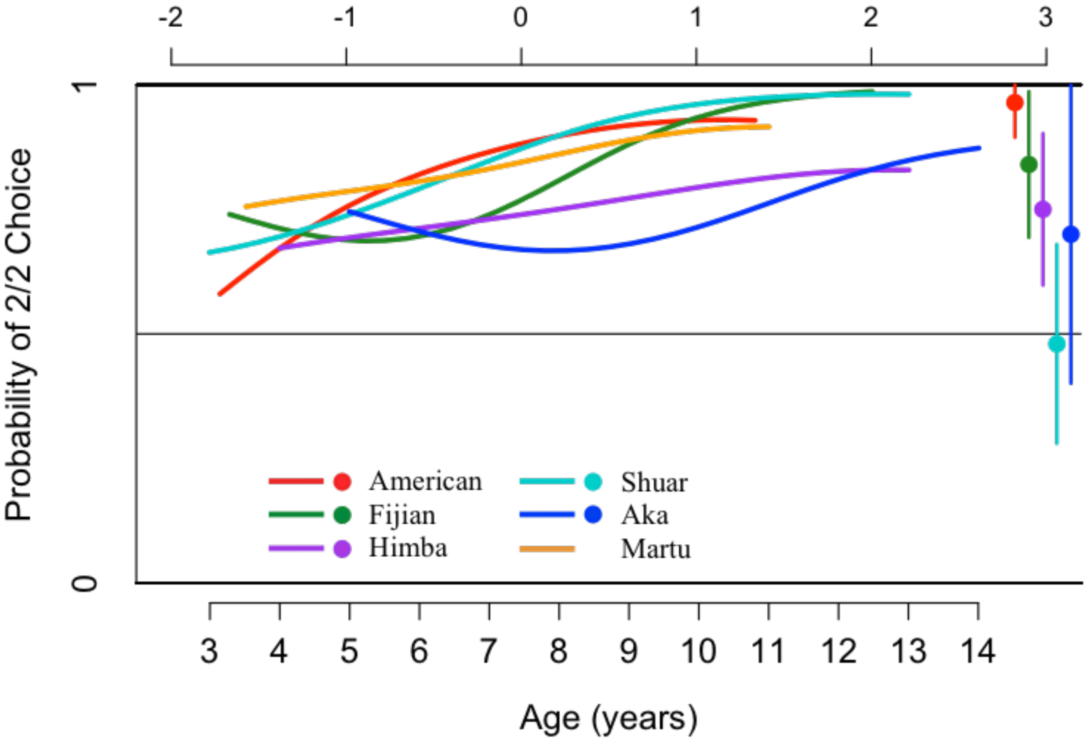
** See Appendix 4 for models and a comparable plot for the CSG.*

*Figure 2: Best-fit model of Actors' choices of 1/1 in the CSG. Vertical axis is the estimated probability that children will choose the prosocial (1/1) outcome. Bottom horizontal axis is children's age (in years), and top horizontal axis is the equivalent value of CA. Colored lines represent estimates for child participants' choices in each population. Dots on the right side of the plot represent data from adults' actual choices in these populations, and lines above and below the dots correspond to 95% confidence intervals. 2(a) plots choices in the CSG Social while 2(b) plots choices from the CSG Asocial. Estimates are all extracted from the best-fit model for the CSG (Model D, Table 2). **



* *See Appendix 4 for models, and Appendix 6 for confidence intervals.*

*Figure 3: Best-fit model of Actors' choices of 2/2 in FAM1 Social. Vertical axis is the estimated probability that children will choose the income-maximizing (2/2) outcome. Bottom horizontal axis is children's age (in years), and top horizontal axis is the equivalent value of CA. Colored lines represent the age functions for child participants' choices in each population. Dots on the right side of the plot represent data from adults' actual choices in these populations, and lines above and below the dots correspond to 95% confidence intervals. Estimates for each population were extracted from the best-fit model for FAM1 Social (Model C, Table 2). **



** See Appendix 4 for models, and Appendix 7 for confidence intervals.*

Figures 1-3 reveal a number of important developmental patterns. Figure 1 depicts the effects of condition in the PG, showing that across all populations there is a general tendency for children to select 1/1 more frequently in the Social condition than in the Asocial condition and that this effect of condition increases with age. In the PG Asocial the probability of selecting 1/1 doesn't deviate from 0.5 (chance) across the entire age range, while in the PG Social the probability of 1/1 choices increases monotonically across the age range, rising above the probability of 1/1 choices in the PG Asocial by about age 6-7.

For the CSG, our model selection procedure indicates that we must consider both condition and population to fully understand children's behavior, and Figure 2 plots separate age functions for each population in both conditions. In the CSG Social (Figure 2[a]) there is considerable conformity across populations prior to middle childhood, with children in all groups becoming less likely to select the 1/1 outcome up until about middle childhood. After this point, the societies begin to diverge, as children in some groups become relatively more likely to select 1/1. The extent of the increase in prosociality among older children varies considerably across populations, but the model's predictions about the behavior of older children roughly corresponds to the actual behavior of adults from the same population in the same task (colored dots on the right side of Figure 2[a]). Variation across groups in adults' behavior in the CSG also matches population-level variation in anonymous Dictator Games (6), and this suggests that children's behavior in the CSG reflects the same kinds of preferences that are captured by other allocation games with adults and children (Appendix 8). See Appendix 5 for additional discussion and analyses showing that these effects are not due to small samples, and for a version of Figure 2(a) that provides means and standard errors for children's choices for each population across three age groups (3-5, 6-8, 9-14 years).

In the CSG Asocial (Figure 2[b]), across populations the probability of selecting 1/1 is below 0.5 for almost the entire age range, and either drops or remains generally flat as age increases. There are some differences in the shape of the age functions across populations, but they are qualitatively similar in that they reveal a common tendency to choose 2/0 across our age range, a pattern that starkly contrasts both with the greater likelihood of 1/1 choices in the CSG Social, and with the substantial population variation in the development of children's likelihood of selecting 1/1 (Figure 2[a]).

Figure 3 depicts the probability of children's choices of 2/2 (the income-maximizing option) in FAM1 Social, which is our proxy measure of children's comprehension of the task. There are some differences in the age trajectories across societies, but the overall pattern is that children are likely to select 2/2, and this likelihood increases with age. Looking across groups, children show systematic preferences for 2/2 over 1/1 by about age 5 (see Appendix 6 for details and confidence intervals). Thus, we assume that children across societies understand this task by at least this age, and when we repeat our model selection procedure with those subjects who selected 2/2 in FAM1 Social we obtain the same qualitative patterns as obtained from the full sample (see Appendix 4, Table S4[b]). We also note that because our methods employ face-to-face interactions and the immediate distribution of real food rewards, our task is likely more easily understood by children than tasks that employ participants that are anonymous or present only in photographs.

These analyses point to four important developmental patterns. First, in the CSG children in all six populations are relatively unlikely to choose the prosocial option as they approach middle childhood. Second, beginning in middle childhood participants in the CSG show population-specific developmental shifts towards adult levels of prosocial behavior in their own groups. Third, in the PG children from all populations show a common shift towards more prosocial behavior with age, but there is markedly less variation across populations in the development of prosocial behavior when generosity is less costly (PG) relative to when it is more costly (CSG). Fourth, there is evidence that in all six societies children understand the choice task substantially prior to middle childhood and the emergence of population variation.

The steep decline in prosocial behavior in the CSG Social among children prior to middle childhood in all six populations (Figure 2[a], also see Appendix 6 Figure S4[g]) suggests a developmental trend toward progressively stronger preferences for self-interested outcomes as children approach middle childhood. This pattern is somewhat surprising because most studies of prosocial behavior indicate that children become more generous as they mature (Eisenberg et al., 2006; Silk and House, 2012). Thus, it is important to consider the possibility that this pattern is the product of young children's confusion about the task or the relative value of the two options that they were presented with. Our analysis of data from FAM1 Social suggests that children understood the choice task by at least age 5, but they may have understood the task even earlier. If our youngest subjects were confused they should have made similar choices in the Asocial and Social conditions. However, there is some evidence that children younger than about 6 years of age (n = 100, 3.0-5.96 years) were more likely to select 1/1 in the CSG Social condition than in the CSG Asocial condition (see Appendix 6, Figures S6[a-g] for age functions with confidence intervals). By showing that children discriminated the conditions, this result suggests that our youngest participants understood the choice task, and it implies that the developmental trajectory we observe prior to middle childhood reflects a shift in children's preferences away from outcomes that benefit others and towards outcomes that benefit themselves.

Discussion

These patterns have important implications for understanding the ontogeny of prosocial behavior and for the study of child development more broadly. Our results suggest that the prosocial behavior of young children may develop through a different process than does the prosocial behavior of older children. This finding highlights the importance of considering non-monotonic developmental patterns, which have been documented for a number of behaviors and cognitive competencies (Siegler, 2004). Very young children are certainly shaped by social learning, but the similarity in the prosocial behavior of our youngest subjects across the very diverse populations we studied suggests that social learning in early childhood does not shape prosocial behavior in a population-specific manner. However, during middle childhood the development of prosociality begins to diverge along population lines, suggesting that children are beginning to become sensitive to society-specific information about how to behave in costly cooperative situations.

The fact that children begin to show increasing rates of prosocial behavior in middle childhood in the CSG Social is consistent with evidence from relatively similar studies showing that Swiss (Fehr et al., 2008) and American (Blake and McAuliffe, 2011) children become substantially more averse to inequity after 7-8 years of age. This increasing shift toward egalitarianism in Western children beginning at age 7-8 is consistent with our sample of children from Los Angeles, but is not as consistent with samples from several of our non-Western populations. This supports the idea that variation in egalitarian motives underlies some of the population-based variation in prosocial behavior in these games. Interestingly, in a recent study of American children, 3-4 year-olds reported that both they and others *should* distribute payoffs equally in a Dictator Game, yet children in the study failed to actually do so until about 7-8 years of age (Smith et al., 2013). This suggests that middle childhood may be when children begin to conform to cooperative social norms, even if they may have learned these norms years prior. Overall, the timing of the shift in the developmental trajectory of prosocial behavior is consistent with claims that middle childhood—a period with unique features in humans that begins around age 6 and ends with sexual maturity (Thompson and Nelson, 2011)—is an important developmental stage across human societies in which children are incorporated into the larger cultural community outside their households (Lancy and Grove, 2011). This would therefore be a particularly important time during development for individuals to conform to local social norms.

Group-specific differences emerged in the CSG but not the PG, suggesting that population-specific influences on the development of prosocial behavior are most pronounced when prosocial outcomes are costly. This fits predictions from gene-culture co-evolutionary models which hypothesize that social norms and institutions will be most influential when group-beneficial behavior is costly and, therefore, more difficult to maintain. Further work should explore specific cultural beliefs and institutions that influence cooperative behavior, and how their acquisition and application shapes children's behavior across development.

Previous work on the ontogeny of prosocial behavior in Western subjects has suggested a trajectory of increasingly prosocial behavior throughout childhood. By tracing the ontogeny of prosociality across a wide age range and in diverse populations, our study shows that this picture is incomplete in several important respects and suggests a more complex role of culture in the ontogeny of prosocial behavior. While there is an important phase of prosocial development prior to middle childhood that appears to be largely independent of society-specific information, it is one characterized by low and perhaps decreasing rates of costly prosociality in our choice task. Beginning in middle childhood, costly prosociality generally increases but the extent of the increase is highly variable and moves toward population-specific levels of mature adult prosociality, developmental diversity that conforms to a distinctive set of predictions derived from gene-culture co-evolution models. We note that in daily life there is substantial cooperation in all of these groups, and our study likely does not capture all of the factors that influence prosociality (e.g. institutions, social norms, evolved biases). We caution that behavior in this study may not be sufficient to predict naturalistic prosocial behavior by individuals in these groups. However, our data show that population membership is one factor that influences cooperative behavior, and this influence emerges in middle childhood.

The fact that our youngest participants (<5-6 years) were relatively more likely to engage in costly prosociality than immediately older children (7-9 years) is a surprising finding, though it is consistent with the considerable evidence that children aged three years and younger act prosocially (Birch and Billman, 1986; Dunfield et al., 2011; Hay et al., 1999; Sommerville et al., 2013; Svetlova et al., 2010; Thompson et al., 1997; Vaish et al., 2009; Warneken and Tomasello, 2009; Warneken, 2013). Indeed, prosociality has been found as early as 25 months of age in choice tasks similar to the one used here (Brownell et al., 2009). A focused investigation of very young children's understanding of these tasks is needed, but our primary results concerning the emergence of population variation are not tied to this issue, as our results show that task comprehension clearly precedes the emergence of population variation in prosociality.

We also note that the specific age predictions that emerge from our models should be interpreted in light of our methods. Differences between our results in the CSG and those obtained with other tasks based on the Dictator Game may be partly due to the fact that, unlike most of these other studies, interactions in our study were not anonymous (as this was unnecessary for our primary questions about population variation). Our subjects' behavior could have been influenced by reciprocity, reputational concerns, and other factors.

The emergence in middle childhood of population differences in costly prosociality, together with a population-independent pattern of monotonically increasing non-costly prosociality, suggests that human prosocial behavior develops through a complex interaction with acquired local culture. Our findings contribute to ongoing discussions of the processes that underlie both uniformity and diversity in social behavior across societies, and highlight the importance of expanding the scope of developmental studies to encompass a wider range of extant human diversity. This is particularly important given the growing evidence for considerable population variation in experimental studies of human behavior (Henrich et al., 2010b).

Materials and Methods

Setup: Actors and Recipients were seated across from one another, with a primary experimenter seated on one side. At some sites a secondary experimenter observed from nearby. Using an apparatus based on prior studies (Fehr et al., 2008), two 8.5" x 14" paper trays were placed on the floor between the Actor and Recipient (see Appendix 1, Figure S1 and Figures S2[a-b]). Each tray had one red circle and one blue circle printed on it. For each trial, payoffs were placed in the circles and the Actor was permitted to choose one of the two trays. In the Social condition, Actors received the payoff in the circle closest to them on the tray that they chose, while recipients received the payoff in the circle closest to them on the same tray. In the Asocial condition, when there was no Recipient, Actors still only received the payoff in the circle closest to them on the tray that they chose (payoffs in the other circle were retrieved by the experimenter). Payoffs were real, visible, food items and immediately edible (see Appendix 1 Table S1[b] for rewards used across sites).

Procedure: Each Actor was presented with two FAM trials (FAM1 and FAM2) and two Test trials (CSG and PG) in each of two different conditions (Social and Asocial), for a total of eight different trials. In the Asocial condition, Actors made choices without a Recipient obtaining rewards (see Figure S1[b]), while in the Social condition a Recipient was seated across from them and received rewards (see Figure S1[c]). In each condition FAM trials were always presented before Test trials. FAM trials were always presented in the same order (FAM1 then FAM2), but the side of presentation for each payoff was counterbalanced across subjects. The order of the Test trials (PG and CSG) was also counterbalanced, as was the order of the two conditions (Social and Asocial). Before all four FAM trials the Actor was given the full set of instructions. Instructions were not given during the Test trials (CSG and PG). See Appendix 1 for protocols and scripts. Data were recorded live on paper datasheets by the experimenter, as video recording was not reliably available at all sites. Datasheets were coded twice (once by a researcher naïve to hypotheses) and inconsistencies in coding were resolved prior to analysis.

Analysis: We used multilevel logistic regressions to analyze our binary outcome variable: whether or not participants selected the 1/1 payoff distribution in the CSG and PG, or the 2/2 payoff in FAM1. We center Participants' Age (PA) to create an age parameter called Centered Age (CA), and we create a second age parameter by squaring CA:

$$\textbf{Centered Age (CA)}=[\textbf{PA}-(\textbf{mean of PA})] \,/\, (\textbf{standard deviation of PA})$$

$$\textbf{Centered Age}^2\ (\text{CA}^2)=\{[\textbf{PA}-(\textbf{mean of PA})] \,/\, (\textbf{standard deviation of PA})\}^2$$

We analyzed the data in the R Environment for Statistical Computing (R Development Core Team, 2013). We fit the models using Stan (Stan Development Team, 2013), a Hamiltonian

Monte Carlo sampler. Results are based on 5000 samples each from 4 chains, after 5000 adaptation steps in each. Convergence was assessed by both trace plots and the R-hat Gelman & Rubin statistic. Model code was generated and DIC calculated using glmer2stan (McElreath, 2013), a convenience package for Rstan. We analyzed the data using both uninformative (flat) priors, as well as weakly informative variance priors, without any substantive change in inferences.

Acknowledgements: For comments and suggestions on the project we would like to thank Nancy Eisenberg, Dan Fessler, Daniel Haun, Keith Jensen, Laurie Santos, Felix Warneken, and two anonymous reviewers. Funding for this research was provided by a grant from the UK's Arts and Humanities Research Council (for the AHRC Culture and the Mind project) and from the Hang Seng Centre for Cognitive Studies, University of Sheffield. JH thanks the Canadian Institute for Advanced Research (CIFAR). BH was supported by the Interdisciplinary Relationship Science Program funded by NSF IGERT.

2 APPENDICES

THE ONTOGENY OF PROSOCIAL BEHAVIOR ACROSS DIVERSE SOCIETIES

Appendix 1: Additional details about methods

Instructions for Experimenters

Rewards for Subjects. The payoffs used are small food items, such as candies. Before being included in gameplay, these items are occluded from view (to avoid distracting attention). Keeping them out of sight in a bag or pocket allows them to be easily retrieved when loading the apparatus at the start of each trial. When the Actor selects a particular payoff distribution, the experimenter removes the unselected payoffs, and encourages the subjects to take their candies. *[Note: candy was not used as the reward at all sites, see Table S1(b)]*

Apparatus. The choices will be presented to children on two large paper trays. On each tray there are two circles (blue on the actor's side, red on the recipient's side). Food rewards will be placed inside the circles during the Familiarization and Test trials.

Trials. Within each session, the order of trials is identical across conditions, but the sides on which the payoffs are presented are reversed between Asocial and Social (e.g., if in the Prosocial Game the 1/1 payoff is on the left in the Social condition, it will be on the right in the Asocial condition). Across sessions, the order of Test trials is counter-balanced. The order of familiarization trials is kept constant across sessions. [The order of familiarization trials is kept constant because pilot testing with adults suggested that subjects gain a clearer understanding of the instructions when Familiarization #1 comes first. This is because both Actor and Recipient could receive a payoff in Familiarization #1, whereas in Familiarization #2 the Recipient cannot receive anything. This can cause confusion.]

Participant roles. To maximize the number of pairings that we can create, most of the children will participate as an Actor and then as a Recipient (except for the Recipient in the first pairing, who will never be an Actor). Children will play the Actor role first, then become the Recipient in a pairing with a new Actor. With a set of 6 children (A, B, C, D, E, and F) 5 pairings would be constructed as follows (see Table S1[a]):

Table S1(a): Pairing of child participants.

Pairing (study session #)	Actor	Recipient
1	**B**	A*
2	C	**B**
3	D	C
4	E	D
5	F**	E

*A begins as Recipient, never plays Actor role (because Actors must be naïve to the task)

**F begins as Actor, but never plays Recipient role (because there are no naïve Actors left)

Note that the Actor in each pairing is always a naïve subject, who has not participated in the game yet. If it is a concern that the first recipient ('A' in the above example) plays the game fewer times than the other subjects, the first recipient could later play with the last Actor ('F' in the above example), who also plays fewer times than the other subjects. However, this pairing should NOT be included as data in the sample, because 'A' was not a naïve subject when playing the role of Actor. If problems with pairings occur, it is ok for a child to play Recipient in multiple pairings after he/she has been an Actor, but not vice versa. Genetically-related siblings should not be paired together, but other pairings are ok.

General Procedure/Setup. Familiarization trials and Test trials are run in a block. First the Familiarization trials are run, and then the Test trials. There will be a set of Familiarization and Test trials for the Social condition, and a separate set of Familiarization and Test trials for the Asocial condition. The ordering of the Social block of trials and Asocial block of trials will be randomized. One experimenter will explain the game & distribute the candies; if possible, a second experimenter sitting away from the children (preferentially out of Actor's line of sight) will record responses live. The children can be seated at a table or on the floor (Figure S1[a]).

Figure S1(a): Experimental setup

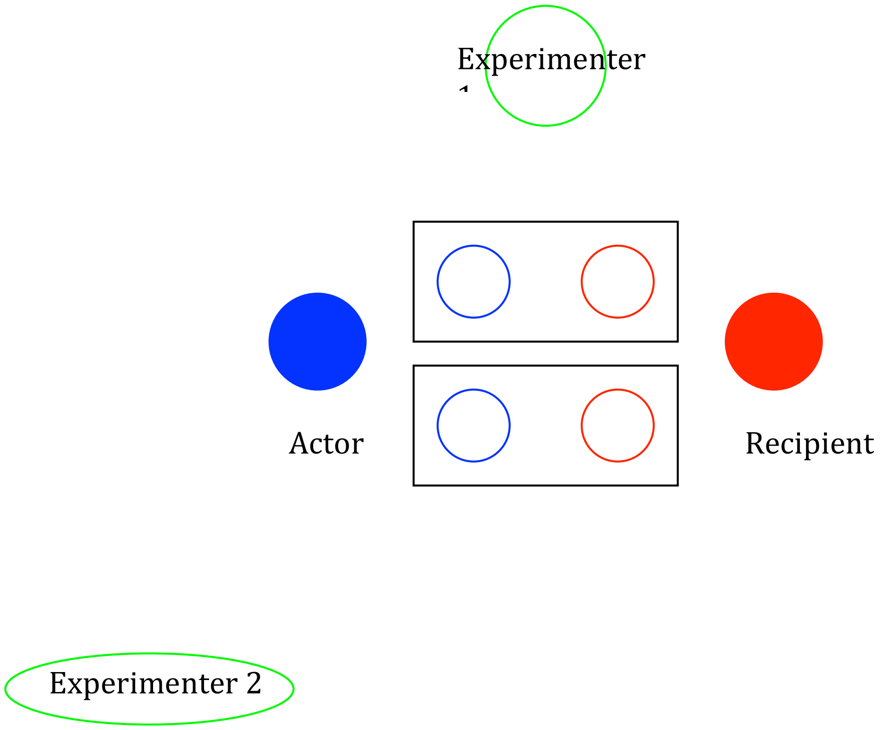


Figure S1(b): Asocial condition

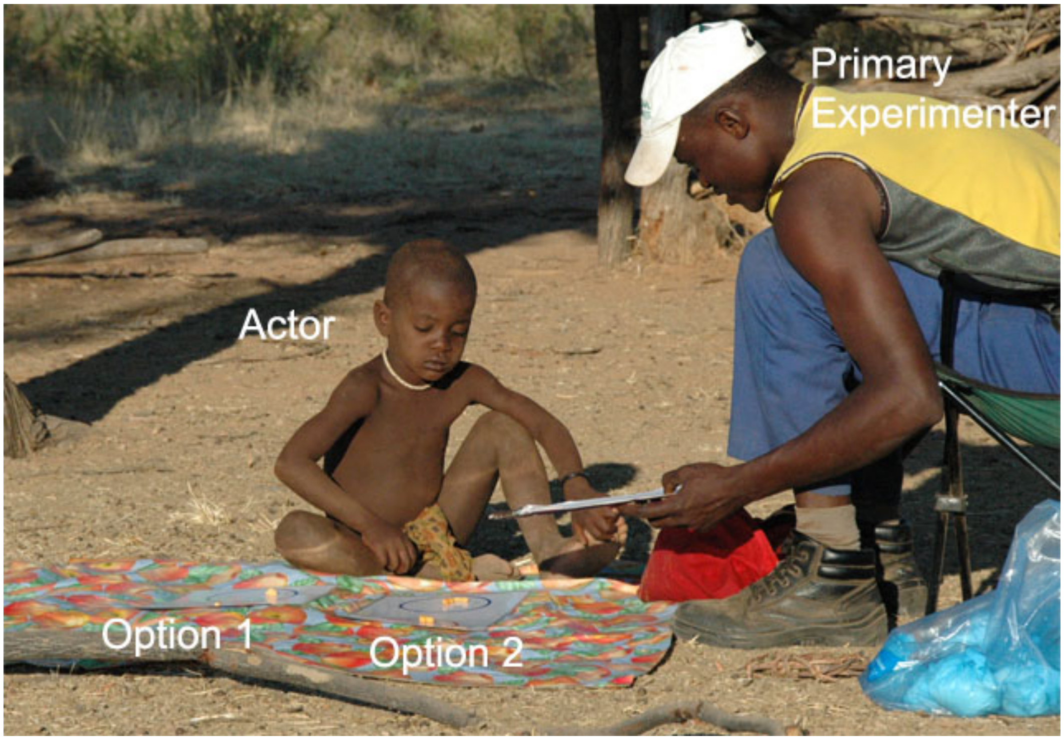
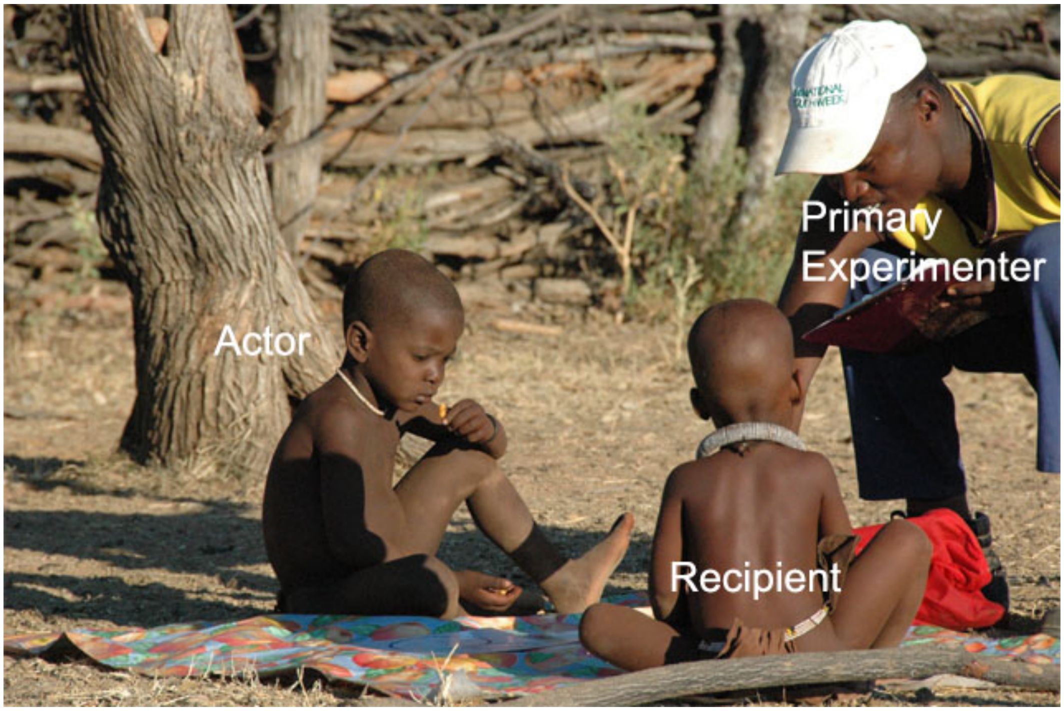


Figure S1(c): Social condition



***[Scripts:** we provide the example scripts distributed to all experimenters, which were translated as needed for each site. Text in brackets was instructions for experimenters, and was not read to participants. Where scripts refer to candies, this was modified as needed to reflect whatever rewards were used at each site (see Table S1(b) for rewards used at different sites).]*

Script for child subjects

“Okay! On each of these papers *[trays]*, there are two circles, a blue circle and a red circle.” **[Point to blue circle and red circle.]**

For Familiarization #1 and #2 (both Social and Asocial conditions)

[Move Left tray in front of child.]

“I put a CANDY in the blue circle which is next to you. Now I put another candy in the blue circle.” **[Place candy on blue circle.]**

“I put a candy in the red circle which is next to [name of Recipient]. Now I put another candy in the red circle.” **[Place candy on red circle.]** *{Note: in Familiarization #2 the experimenter says ‘I don’t put anything in the red circle next to [name of Recipient]’}* *{Note: in the Asocial condition the experimenter replaces ‘next to [name of Recipient]’ with ‘on the other side’}*

[Move Right tray in front of subject.]

“I put a CANDY in the blue circle which is next to you. “**[Place candy on blue circle.]**

“I put a CANDY in the red circle which is next to [name of Recipient].” **[Place candy on red circle.]** *{Note: in Familiarization #2 the experimenter says ‘I don’t put anything in the red circle next to [name of Recipient]’}*

{Note: in the Asocial condition the experimenter replaces ‘next to [name of Recipient]’

with 'on the other side'}

[Point to the Left tray:]

"If you choose this paper *[tray]*, you get whatever is in the blue circle and [name of Recipient] gets whatever is in the red circle." *{Note: in the Asocial condition the experimenter replaces '[name of Recipient]' with 'nobody'}*

[Point to the Right tray:]

"If you choose this paper *[tray]*, you get whatever is in the blue circle and [name of Recipient] gets whatever is in the red circle." *{Note: in the Asocial condition the experimenter replaces '[name of Recipient]' with 'nobody'}*

[Now sit back and look at the subject, not at the trays.]

"NOBODY gets what is on the other paper *[tray]*. The candy goes back into the bag. You can choose whichever paper *[tray]* you want, but you can only choose one paper *[tray]*. Now, you can make your choice. Go ahead."

"Okay, now let's do some more."

For PG and CSG (both Social and Asocial conditions)

[Left tray: Place candies on blue circle. Then, place candies on red circle.]

[Right tray: Place candy on blue circle. Then, place candy on red circle.]

“Now, you can make your choice. Go ahead.”

[**Remove remaining candies and put them back in the bag. If the subjects do not take their candies, encourage them to do so.]**

Script for adult subjects

“This is a game that we usually play with children, but we’re also curious about how adults will play it. Would you be willing to participate?”

“On each of these papers *[trays]*, there are two circles, a blue circle and a red circle.

[Point to blue circle and red circle.]

For Familiarization #1 and #2 (both Social and Asocial conditions)

[Left tray: Place candies on blue circle. Then, place candies on red circle.]

[Right tray: Place candy on blue circle. Then, place candy on red circle.]

[Point to the Left tray:]

“If you choose this paper *[tray]*, you get whatever is in the blue circle and [name of Recipient] gets whatever is in the red circle.” *{Note: in the Asocial condition the experimenter replaces ‘[name of Recipient]’ with ‘nobody’}*

[**Point to the Right tray:]**

"If you choose this paper *[tray]*, you get whatever is in the blue circle and [name of Recipient] gets whatever is in the red circle." *{Note: in the Asocial condition the experimenter replaces '[name of Recipient]' with 'nobody'}*

[Now sit back and look at the subject, not at the tray.]

"Remember, NOBODY gets what is on the paper *[tray]* that you don't choose. That candy goes back into the bag."

"Ok, go ahead and make your choice."

[**Remove remaining candies and put them back in the bag. If the subjects do not take their candies, encourage them to do so.]**

"Okay, now let's do some more."

For PG and CSG (both Social and Asocial conditions)

[Left tray: Place candies on blue circle. Then, place candies on red circle.]

[Right tray: Place candy on blue circle. Then, place candy on red circle.]

"Now, you can make your choice. Go ahead."

[**Remove remaining candies and put them back in the bag. If the subjects do not take their candies, encourage them to do so.**

Table S1(b): *Methodological variations across sites*

Group	**Child recruitment**	**Child pairs familiar?**	**Child Rewards**	**Reward rare on site?**	**Adult recruitment**	**Adult pairs familiar?**	**Adult rewards**	**Reward rare on site?**
Aka	On an opportunity basis in local community (forest camps).	Yes. Subjects drawn from same group of families living together. Children knew each other since infancy.	Hard candy	Yes	On an opportunity basis in local community (forest camps).	Yes, all drawn from same camp.	Hard candy	Yes
American / West Los Angeles	At preschools and primary schools.	Yes. Subjects drawn from same school classes.	Cheese crackers	No	Undergraduate sections of Anthropology courses at UCLA	Yes. Subjects drawn from same school classes.	Hershey's kisses	No
Fijian (Yasawa)	At schools and on an opportunity basis in three small villages.	Yes. All subjects lived in same villages; children of school age were paired with classmates.	Hard candy	Fairly rare	Residents of same villages	Yes	hard candy	Fairly rare
Himba	At schools and on opportunity basis in extended family compounds.	Yes.	Cheese snack	Yes	Adults were drawn from the same community as child subjects.	Yes; members of the same community	Cheese snack	Yes
Martu	At community school.	Yes, drawn from same school.	Crackers	No	n/a	n/a	n/a	n/a
Shuar	At village school.	Yes, all from same village.	Animal cracker cookies	No	Residents of same village	Yes, all from same village. Some were kin.	Animal cracker cookies	No

Appendix 2: Additional demographic data

Figure S2: Age distributions for the different populations in our sample (N=326).

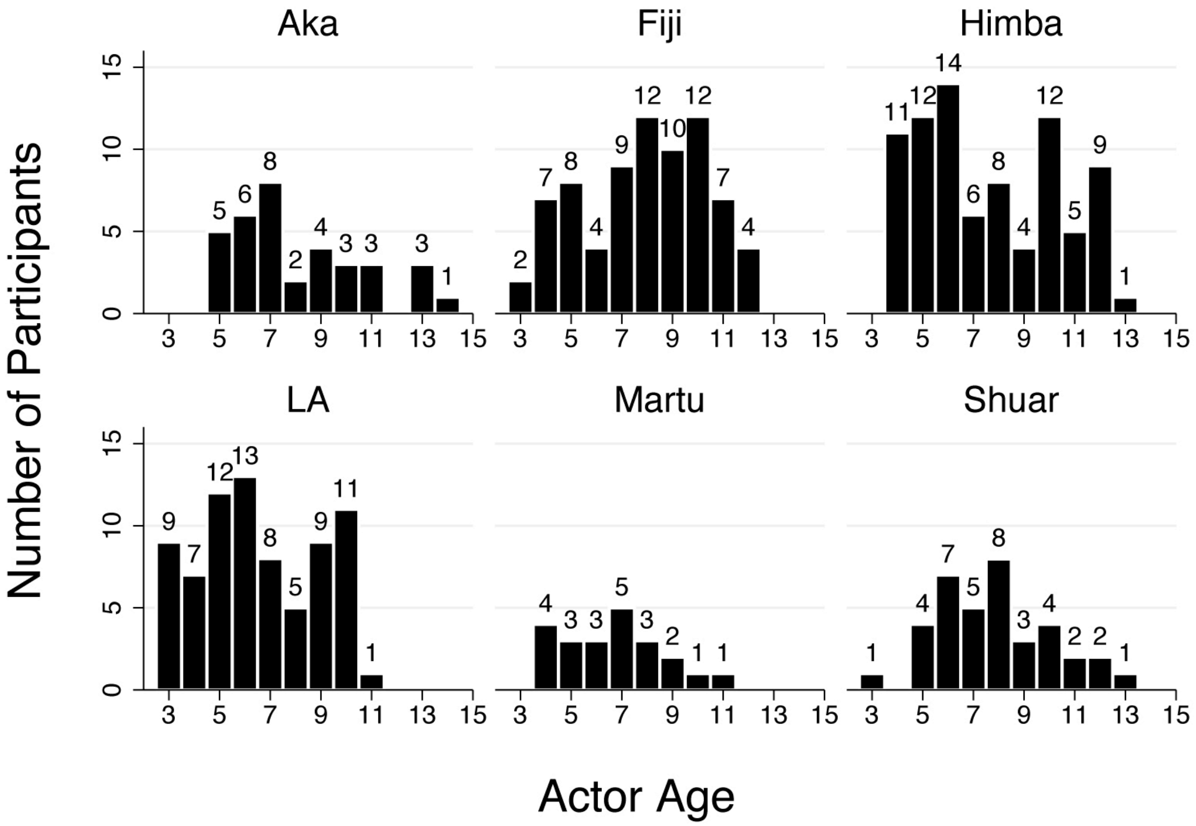


Table S2(a): Additional demographic information about field sites.

Group	**Nation/Region**	**Language Family**	**Environment**	**Economic Base**	**Residence**	**Researcher**
Aka	Central African Republic	DiAka (Bantu C10)	Tropical forest	Foraging	Nomadic	Boyette / Hewlett
American	United States / Los Angeles	English	Urban	Wage work	Sedentary	House
Fijian	Fiji / Yasawa Island	Oceanic	Coastal tropical island	Horticulture / marine foraging	Sedentary	Silk / Henrich
Himba	Namibia (Kaokoland)	Herero	Semi-arid mountain savanna	Agro-pastoralism	Semi-nomadic	Silk / Scelza
Martu	Western Australia	Martu Wangka	Desert	Foraging/Wage Work	Semi-nomadic	Scelza
Shuar	Ecuador/Amazonia	Jivaroan	Tropical forest	Horticulture, some wage work	Sedentary	Barrett

Table S2(b): Details about school attendance and change in how children interact with peers across development.

Group	Age children first attend school	Children's interactions with peers (early childhood; <5 years)	Children's interactions with peers (middle childhood; 6 years to puberty)	Children's interactions with peers (adolescence; >puberty)
Aka	Most children do not go to school. Those that do start as early as 5-years-old. None were attending school when the study was conducted.	Children stay in camp, but interact with anyone who is around. Mostly they interact with peers <5 years at this age, as older children will leave camp during the day to play and work.	After age 5, children spend on average twice as much time away from camp in the forest foraging and playing. At least half of the time spent in the forest they are out of sight of adults, with mixed-age, mixed sex groups of children (ages 7-18 years).	Aka adolescents spend a greater percentage of their time with other adolescents than younger children spend with only same-age peers, but they are also usually with adults or children younger than age 5, either working or visiting.
American	2-3 for preschool; 5 kindergarten; 6 primary school.	Children are primarily in the home, interacting with siblings or neighbors.	Children also interact with many unfamiliar peers in school settings.	Children interacting with many peers in school, and also increasingly in non-school activities.
Fijian	5: kindergarten; 6 primary school.	Children primarily in the home, interacting with siblings or neighbors in the village.	Children also interact with peers in school, which includes many of the same children from their village, in addition to children from neighboring villages whom they are likely already familiar with.	Children interacting with many peers in school. At age 14/15 many children leave the village to continue schooling on a larger island, and there will interact with unfamiliar peers.
Himba	5, but very few attend school.	Children primarily in the home, interacting with siblings or children from neighboring compounds.	Children are primarily in the home, interacting with siblings, cousins and neighboring children. Expected to perform household labor from about age 5, including collecting water, grinding maize, herding goats, and watching young siblings.	Children interact with those in their compound and neighboring compounds and may travel independently. Girls may go to live with a spouse after menarche (although marriage can occur much earlier, change in residence usually doesn't happen until menarche). First birth is in mid to late teens.
Shuar	5 years	Children visit households all over the village.	Lots of interaction across households within village; visits to other villages (e.g. for sports) starts in middle childhood.	Children interact with peers both within and outside the village.
Martu	4 years	Children primarily in the home, interacting with siblings, cousins and neighboring children.	Children at school or at home or neighboring home within the community.	Children interact with those in their household and community and often travel to other communities to see relatives or friends.

Appendix 3: Effect of condition and population for adults.

If Actors were prosocial then they should have chosen 1/1 more in the Social condition than in the Asocial condition. Here we explore whether adults' 1/1 choices in the Prosocial Game (PG) and the Costly Sharing Game (CSG) varied across condition and population. In Table S3(a) we use a model-selection procedure similar to that used with children (see Table 3 Main text, Appendix 4), and in Table S3(b) we present the best-fit models for the CSG and PG.

Table S3(a): Multilevel logistic regression models used in model selection, comparable to model selection procedure used with child data. Below we provide details about Fixed Effect (FE) and Random Effect (RE) parameters included in each model, the hypothesis reflected in each model, and the DIC and DIC weight values associated with each model when it is applied to the CSG, and PG. Best-fit models are in bold.

Model	Parameters included	Hypothesis	CSG	PG
			DIC (weight)	DIC (weight)
A	**FE:** none **RE:** Actor ID	No differences across conditions or populations	335.42 (.00)	332.6 (.00)
B	**FE:** Condition **RE:** Actor ID	Differences across conditions	323.6 (.00)	**311.87** **(.70)**
C	**FE:** none **RE:** Actor ID Population	Differences across populations	327.93 (.00)	334.72 (.00)
D	**FE:** Condition **RE:** Actor ID Population | Condition	Differences across conditions and populations	**302.41** **(1.00)**	313.53 (.30)

As with children (Table 3 in main text, Appendix 4), the best-fit model for adults was Model D for the CSG, and Model B for the PG (Table S3[a]). Model D also had a .3 probability of being the best model for the PG, and the DIC values for Models B and D are similar, suggesting that there is a greater likelihood of differences between populations of adults than between populations of children, but it is still substantially more likely that there are population differences in the CSG than in the PG.

Table S3(b): Best-fit models for the CSG and PG for adults.

DV: Chose Prosocial Outcome (1/1) **N=120**	**Model 1**	**Model 2**
	CSG	**PG**
	Coef. **(95% CI)**	**Coef.** **(95% CI)**
Intercept	-2.11 (-4.07 : -0.56)	-1.69 (-2.59 : -0.82)
Condition	1.32 (0.20 : 2.76)	1.39 (0.83 : 1.99)
Variance parameters	**SD** **(95% CI)**	**SD** **(95% CI)**
Actor ID | Intercept	0.83 (0.47 : 1.29)	0.52 (0.41 : 1.02)
Population | Intercept	2.15 (0.43 : 3.05)	
Population | Condition	1.53 (0.45 : 2.54)	

The overall pattern is that, despite a greater likelihood of between-population differences in the CSG than in the PG, the estimate for the Condition parameter is greater than zero for both. This indicates that adults were more likely to select the 1/1 outcome in the Social condition than in the Asocial condition, for both the CSG and the PG.

Appendix 4: Best-fit models for PG, CSG, and FAM1 for children.

Table S4(a): Table 3 from the main text. N=326 for each task.

Model	Parameters included	Hypothesis	CSG	PG	FAM1 Social
			DIC (weight)	DIC (weight)	DIC (weight)
A	**FE:** CA, CA^2 **RE:** Actor ID	One developmental trajectory	765.12 (.00)	904.61 (.06)	321.53 (.05)
B	**FE:** CA, CA^2 Condition, CA*Condition, CA^2 *Condition **RE:** Actor ID	One trajectory for each condition	746.86 (.00)	**899.10** **(.94)**	
C	**FE:** CA, CA^2 **RE:** Actor ID Population | CA, CA^2	One trajectory for each population	747.30 (.00)	914.97 (.00)	**315.58** **(.95)**
D	**FE:** CA, CA^2 Condition, CA*Condition, CA^2 *Condition **RE:** Actor ID Population | CA, CA^2 Condition, CA*Condition, CA^2 *Condition	One trajectory for each condition, for each population	**734.49** **(1.00)**	916.52 (.00)	

Table S4(b): Additional model selection routines, comparable to those from Table S4(a). Here we run the same procedures, but either (a) include Actor Sex as a parameter in the model (which removes three subjects whose sex was not recorded), (b) limit the dataset to those Actors 5-10 years of age, or (c) limit the dataset to those Actors who selected the 2/2 (income-maximizing) outcome in FAM1 Social. The best-fit models are the same as in Table S4(a).

		(a) Including Actor Sex as a Fixed Effect parameter (N=326)		(b) Limiting analysis to Actors who are 5-10 years of age (N=235)		(c) Limiting analysis to Actors who selected 2/2 in FAM1 Social (N=259)	
Model	Hypothesis	CSG	PG	CSG	PG	CSG	PG
		DIC (weight)	DIC (weight)	DIC (weight)	DIC (weight)	DIC (weight)	DIC (weight)
A	One developmental trajectory	761.59 (.00)	896.16 (.04)	530.17 (.00)	650.31 (.02)	560.32 (.00)	714.92 (.00)
B	One trajectory for each condition	743.87 (.00)	**889.99** **(.96)**	517.66 (.02)	**642.12** **(.98)**	555.07 (.00)	**696.63** **(1.00)**
C	One trajectory for each population	745.97 (.00)	904.89 (.00)	521.48 (.00)	653.08 (.00)	546.72 (.01)	722.06 (.00)
D	One trajectory for each condition, for each population	**731.94** **(1.00)**	907.71 (.00)	**509.54** **(.98)**	653.45 (.00)	**537.34** **(.99)**	708.98 (.00)

Table S4(c): Fully specified best-fit models from Table S4(a). The model for FAM1 focuses only on the Social condition, so does not include parameters for Condition or for Actor ID (as each Actor only contributed one observation to this dataset).

DV: Chose 1/1 **N = 326**	**Model 3**	**Model 4**	**Model 5**
	CSG	**PG**	**FAM1 Social**
	Coef. **(95% CI)**	**Coef.** **(95% CI)**	**Coef.** **(95% CI)**
Intercept	-1.59 (-2.69 : -0.49)	-0.63 (-1.31 : 0.05)	1.53 (0.88 : 2.23)
CA	-0.71 (-1.63 : 0.15)	-0.50 (-1.04 : 0.02)	0.63 (-0.07 : 1.39)
CA^2	-0.68 (-1.57 : 0.19)	-0.24 (0.25 : 0.70)	.00 (-0.57 : 0.59)
Condition	0.14 (-0.59 : 0.89)	0.52 (0.09 : 0.95)	
CA * Condition	0.16 (-0.48 : 0.77)	0.40 (0.08 : 0.76)	
CA^2 * Condition	0.67 (0.05 : 1.29)	-0.12 (-0.41 : 0.20)	
Variance parameters	**SD** **(95% CI)**	**SD** **(95% CI)**	**SD** **(95% CI)**
Actor ID | Intercept	1.02 (0.70 : 1.25)	0.52 (0.28 : 0.81)	
Population | Intercept	0.66 (0.32 : 1.24)		.67 (0.33 : 1.24)
Population | CA	0.65 (0.32 : 1.17)		0.74 (0.35 : 1.34)
Population | CA^2	0.66 (0.32 : 1.21)		0.57 (0.29 : 1.01)
Population | Condition	0.60 (0.31 : 1.05)		
Population | CA * Condition	0.57 (0.30 : 1.01)		
Population | CA^2 * Condition	0.56 (0.30 : 0.98)		

Table S4(d): Estimated parameters for the PG, extracted from the best-fit model for this task (Model B, Table S4[a]). These are the estimates plotted in Figure 1 in the main text.

DV: Chose 1/1 **N = 326**	**Model 6**	**Model 7**
	PG Social	**PG Asocial**
	Coef. **(95% CI)**	**Coef.** **(95% CI)**
Intercept	0.41 (0.08 : 0.73)	-0.11 (-0.41 : 0.19)
CA	0.31 (0.08 : 0.55)	-0.10 (-0.33 : 0.14)
CA^2	0.01 (-0.21 : 0.25)	0.13 (-0.09 : 0.34)

Model 7 indicates that for the PG Asocial none of the parameters are clearly different from zero, suggesting that there was little change as a function of age in the probability that children chose 1/1 in this task. In contrast, Model 6 shows us that in the PG Social the coefficient for CA was substantially greater than zero, suggesting that children became more likely to chose 1/1 with increasing age.

Table S4(e): Estimated parameters for each of our six populations in the CSG Social condition, extracted from the best-fit model for the CSG (Model D, Table S4[a]). These are the estimates plotted in Figure 2(a) in the main text, and Figures S4(a-f).

DV: Chose 1/1 N = 326	**Model 8**	**Model 9**	**Model 10**	**Model 11**	**Model 12**	**Model 13**
	American	**Fijian**	**Himba**	**Shuar**	**Aka**	**Martu**
	Coef. (95% CI)	**Coef. (95% CI)**	**Coef. (95% CI)**	**Coef. (95% CI)**	**Coef. (95% CI)**	**Coef. (95% CI)**
Intercept	-1.87 (-2.81 : -1.02)	-0.80 (-1.63 : 0.00)	-2.11 (-3.07 : -1.24)	-1.82 (-2.91 : -0.85)	-0.81 (-1.79 : 0.12)	-0.61 (-1.75 : 0.59)
CA	0.96 (0.24 : 1.76)	-0.59 (-1.25 : 0.00)	-0.66 (-1.25 : -0.10)	-0.86 (-1.81 : 0.09)	-0.61 (-1.67 : 0.38)	-0.51 (-1.78 : 0.67)
CA^2	1.53 (0.76 : 2.35)	-0.17 (-0.85 : 0.46)	0.78 (0.24 : 1.38)	0.80 (0.07 : 1.59)	0.74 (0.07 : 1.50)	0.31 (-0.90 : 1.47)

All Models report a substantial positive estimate for CA^2, with the exceptions of Models 9 and 13, where the estimate for CA^2 is near zero. A near-zero estimate for CA^2 suggests a monotonic age trajectory, while substantial positive estimates for CA^2 point to a quadratic shape with a positive bend (a u-shape). The estimates for the intercept for all of the populations are substantially negative, indicating a low probability of 1/1 choices at this point on the centered age vector (i.e. 7.42 years of age), across all populations. This suggests qualitatively different age trajectories across our different populations, with some showing increasing rates of 1/1 choices as a function of age, while others show little change or decreases in the rates of 1/1 choices with age. These results are clearly discernible from Figure 2(a) in the main text (also Figure S4[a-f]).

Table S4(f): Estimated parameters for each of our six populations in the CSG Asocial condition, extracted from the best-fit model for the CSG (Model D, Table S4[a]). These are the estimates plotted in Figure 2(b) in the main text, and Figures S4(a-f).

DV: Chose 1/1 N = 326	**Model 14**	**Model 15**	**Model 16**	**Model 17**	**Model 18**	**Model 19**
	American	**Fijian**	**Himba**	**Shuar**	**Aka**	**Martu**
	Coef. (95% CI)	**Coef. (95% CI)**	**Coef. (95% CI)**	**Coef. (95% CI)**	**Coef. (95% CI)**	**Coef. (95% CI)**
Intercept	-1.89 (-2.69 : -1.16)	-1.18 (-1.89 : -0.51)	-1.88 (-2.68 : -1.16)	-1.77 (-2.66 : -0.99)	-1.02 (-1.83 : -0.27)	-1.03 (-1.90 : -0.14)
CA	0.34 (-0.24 : 0.93)	-0.64 (-1.26 : -0.06)	-0.69 (-1.22 : -0.19)	-0.87 (-1.61 : -0.16)	-0.70 (-1.50 : 0.04)	-0.66 (-1.61 : 0.20)
CA^2	0.41 (-0.19 : 1.01)	-0.64 (-1.33 : -0.04)	0.16 (-0.34 : 0.66)	0.13 (-0.51 : 0.78)	0.17 (-0.40 : 0.73)	-0.26 (-1.17 : 0.56)

Most of these models include estimates for CA^2 that are close to zero, with the exception of the negative estimate in the model of our Fijian sample (Model 15), which suggests an inverted u-shape. Half of the models' estimates for CA are close to zero (Model 14, 18, and 20), while the others are negative. Additionally, all of the models include a large negative estimate for the intercept. Overall, these results suggest that in the CSG Asocial there is some variation in how children's behavior develops across our populations, but across groups children are consistently unlikely to select the 1/1 outcome, and they either stay consistently unlikely to do so or become less likely to select 1/1 with increasing age.

Table S4(g): Estimated parameters for each of our six populations for FAM1 Social condition, extracted from the best-fit model for this dataset (Model C, Table S4[a]). These are the estimates plotted in Figure 3 in the main text, and Figures S5(a-f).

DV: Chose 1/1 N = 326	**Model 20**	**Model 22**	**Model 23**	**Model 24**	**Model 25**	**Model 26**
	American	**Fijian**	**Himba**	**Shuar**	**Aka**	**Martu**
	Coef. (95% CI)	**Coef. (95% CI)**	**Coef. (95% CI)**	**Coef. (95% CI)**	**Coef. (95% CI)**	**Coef. (95% CI)**
Intercept	2.08 (1.27 : 3.04)	1.24 (0.52 : 2.00)	1.07 (0.41 : 1.76)	2.03 (1.04 : 3.35)	0.74 (-0.14 : 1.59)	1.80 (0.77 : 3.06)
CA	0.83 (0.06 : 1.83)	1.01 (0.29 : 1.96)	0.28 (-0.19 : 0.77)	1.32 (0.18 : 3.13)	-0.10 (-1.07 : 0.77)	0.65 (-0.43 : 1.98)
CA^2	-0.10 (-0.82 : 0.62)	0.59 (-0.10 : 1.48)	0.04 (-0.43 : 0.54)	0.40 (-0.53 : 1.61)	0.34 (-0.28 : 1.08)	0.24 (-0.74 : 1.28)

All models report a substantial positive estimate for Centered Age (CA), with the exception of Models 23 and 25 in which the estimate for CA is close to zero. Also, all models report an estimate for CA^2 that is close to zero. Additionally, the estimates for the intercepts are positive, and all are substantially above zero with the exception of the model for the Aka sample (Model 25). This suggests a high probability of children selecting 2/2 at the intercept value of CA (i.e. 7.42 years of age), across at least most of our populations. These results suggest some variability in the age function across these populations, but all of these functions are shifted towards high probabilities of choosing 2/2, and most involve either an increasing slope to the age function (i.e. increasing change in prosociality as a function of age) or a relatively unchanging slope.

Appendix 5: Addressing concerns about sample size

Our population samples are varied both in their total size and in their distribution across the ages of 3-14 years. This is an unavoidable feature of naturalistic samples drawn from small-scale societies, but in a number of ways our analyses are well-equipped to deal with small samples. By utilizing age as a continuous variable we avoid much of the noise created when aggregating observations into broad age bins. For example, grouping children into 3 year-olds and 4 year-olds is problematic because a 4.1 year-old is likely to be more similar to a 3.9 year-old than to a 4.9 year-old. Using regressions also allows us to make informative predictions about the behavior of children even at the peripheries of our age distributions, where our samples are small or even missing, because the behavior of a 3 year-old is not completely independent from the behavior of a 4 year-old. Furthermore, our model structures reduce concerns about small sample sizes, by combining data across populations and estimating our age functions using interactions and random effects. By using random effects the model is considering each population as a sample drawn from a larger distribution, and observations from one population still inform the estimation of the age parameters for the other populations. This reduces concerns about differences in sample size across populations because all the observations in the sample are used when estimating parameters for each population.

Truncated samples: We can further reduce concerns about the effect of small samples along the periphery of our age distribution by truncating our sample to all participants aged 5-10 years. This covers the densest parts of our age distributions for each of our samples, and all of our populations include participants with these ages. In Tables S4(a) and S4(b) we compare the results of our primary model selection comparison when applied to all of our subjects, and to the subset of our sample aged 5-10 years. The results are qualitatively the same.

Age bins: We can also reduce concerns about sample size by aggregating subjects into broad age categories with approximately comparable numbers of participants. Here we use three age bins: 3-5 years, 6-8 years, and 9-14 years. This approach is noisy and imprecise, but by presenting our data in a more traditional format with means and standard errors we hope to alleviate concerns that our models produce spurious results. In Figure S3(a), we re-present a version of Figure 2[a] from the main text, and in Figure S3(b) we present another version of this figure that uses means and standard errors to plot the same data (children's responses in the CSG Social and PG Social).

Figure S3(a): Similar to Figure 2(a) from main text. Best-fit models used to plot age functions for CSG Social (colored lines) and PG Social (black dotted line). Dots on the right-hand side of the graph represent the proportions of adults who selected 1/1 (colored dots for the CSG, solid black dot for the PG), and lines above and below dots correspond to 95% confidence intervals. The vertical axis corresponds to the estimated probability that children will choose the prosocial (1/1) outcome. Horizontal axis corresponds to children's age (in years).

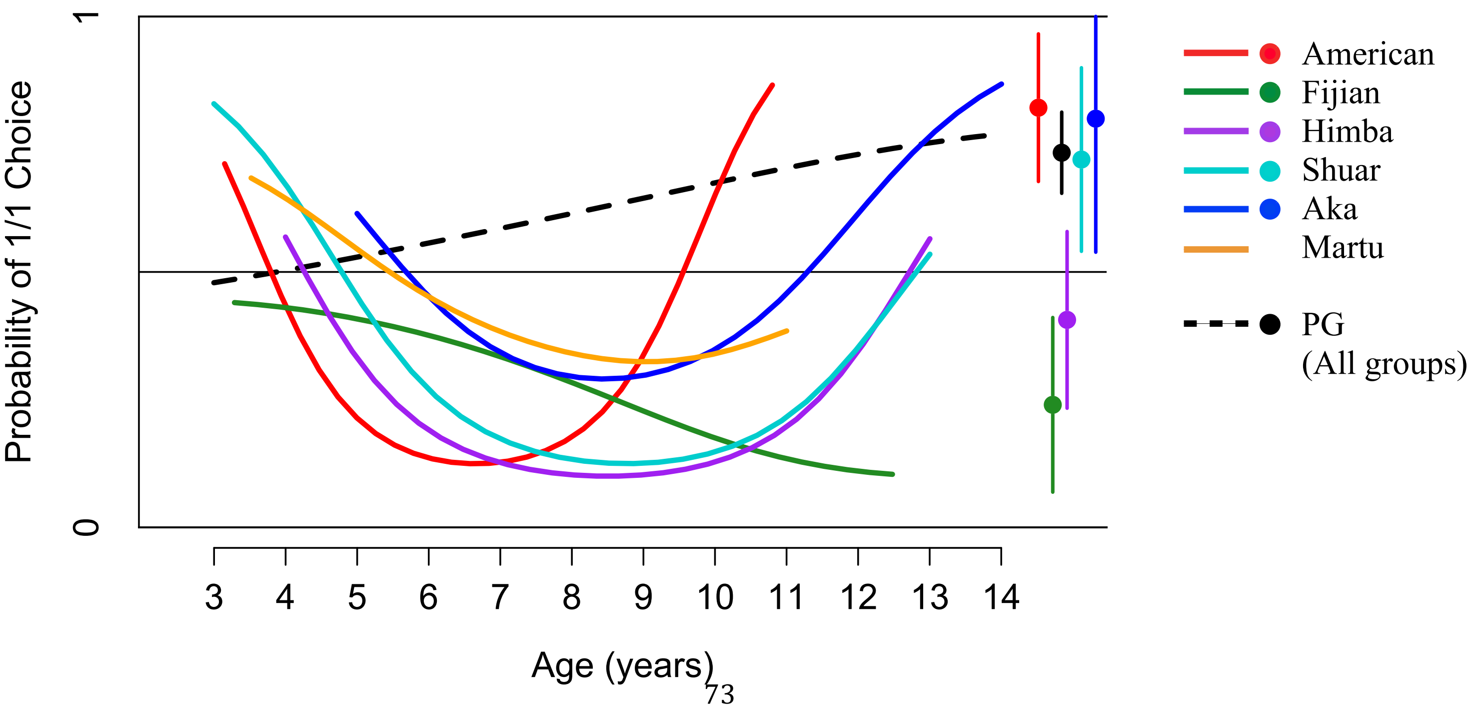
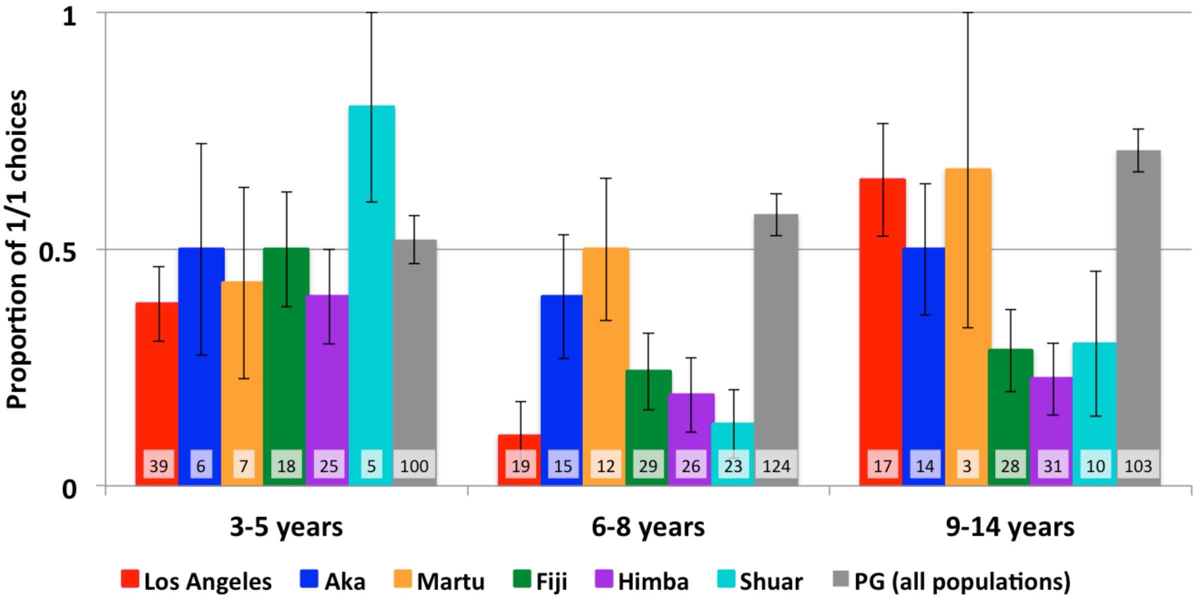


Figure S3(b): Proportions of children who select 1/1 outcome, for three age categories that have approximately equal numbers of participants: 3-5 years, 6-8 years, and 9-14 years. Colored bars represent means for each population in the CSG Social, grey bars plot means for all populations together in the PG Social. Error bars reflect the standard error of the mean, and the values at the base of each bar reflect the number of participants included in that bar.



We can see similar overall developmental patterns in Figure S3(a) and Figure S3(b). Both plots indicate relative similarity across populations in early childhood (3-5 years). This is followed by lower overall probabilities/proportions of 1/1 choices and emerging population variation in middle childhood (6-8 years). Finally, in late childhood/early adolescence (9-14 years) we can see even greater population variation, along with higher probabilities/proportions of 1/1 choices in some groups (relative to 6-8 year-olds).

In general, our primary results concerning the origins of population variation hold across different methods for representing our data. This shows that our results are not an artifact of our methods. Furthermore, given the many benefits of using multilevel regressions, this strongly argues for their use when exploring development across different populations.

Appendix 6: 95% Confidence intervals for Figure 2 from main text.

Figures S4(a-f): See Tables S4(e) and S4(f) for associated models. The vertical axis on each plot corresponds to the estimated probability that children will choose the prosocial (1/1) outcome. The bottom horizontal axis corresponds to children's age (in years), and the top horizontal axis corresponds to equivalent values of CA. The solid line and the colored shaded area in each plot correspond to the estimated probability of 1/1 choices and the 95% confidence interval for the Social condition. The filled dot on the right side of the plot corresponds to adults' choices in the Social condition, and the hollow dot corresponds to adults' choices in the Asocial condition. The dotted line and the grey shaded area in each plot correspond to the estimated probability and 95% confidence interval for the Asocial condition. The darker shaded area is where the confidence intervals overlap.

Figure S4(a): Los Angeles

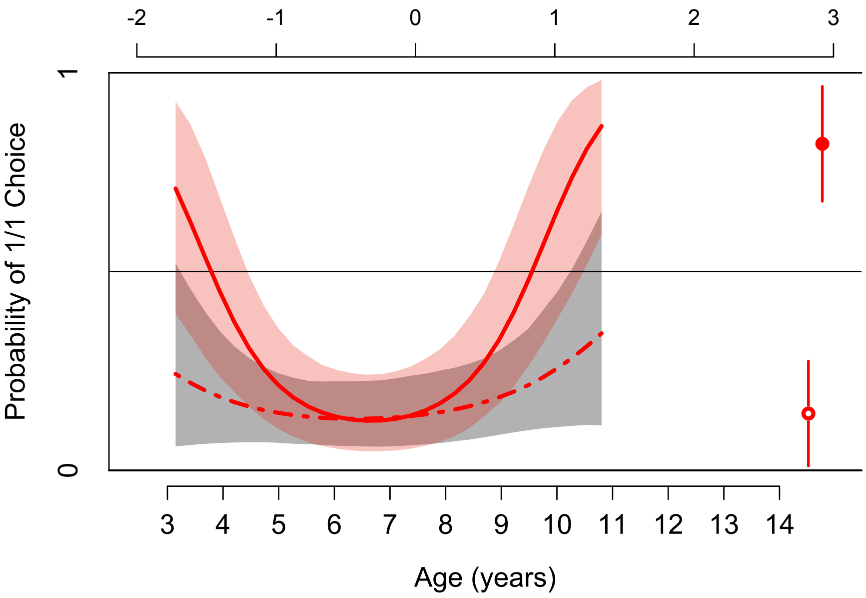


Figure S4(b): Fiji

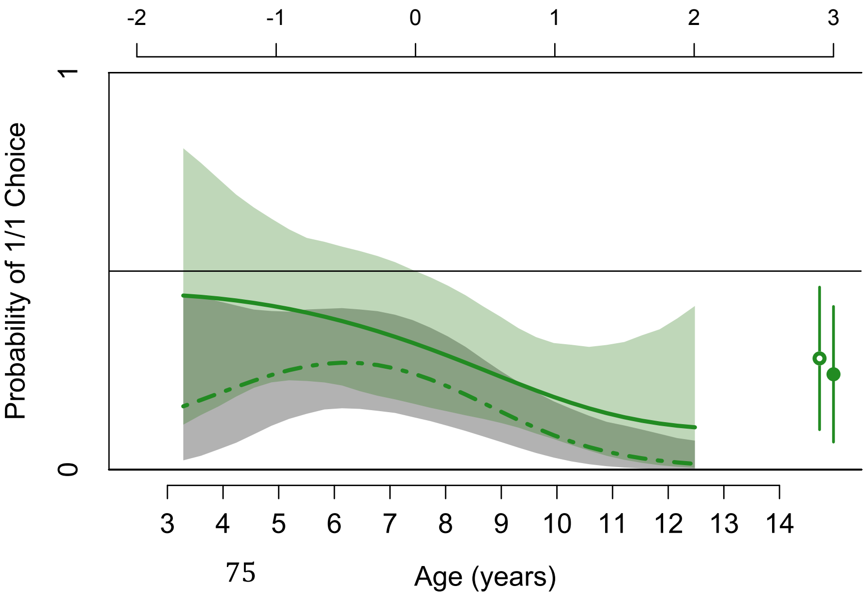


Figure S4(c): Himba

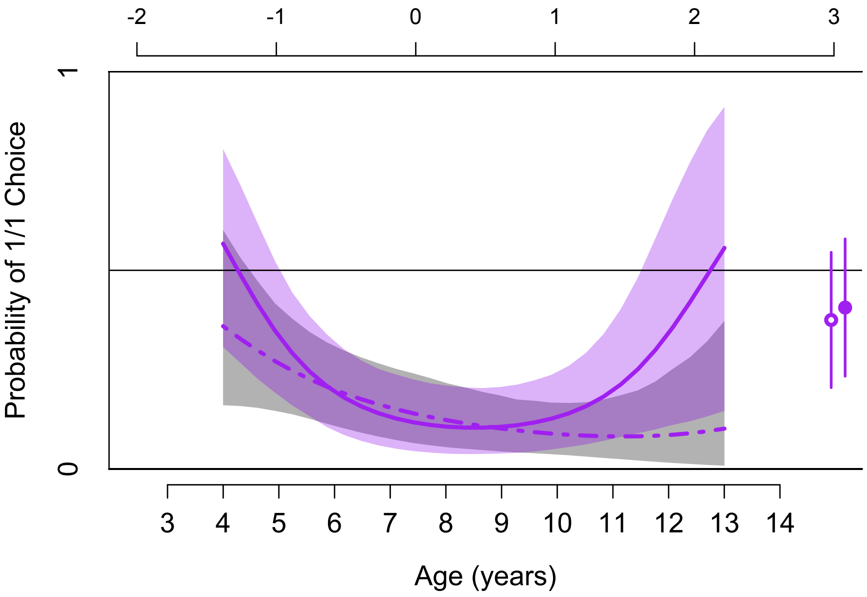


Figure S4(d): Shuar

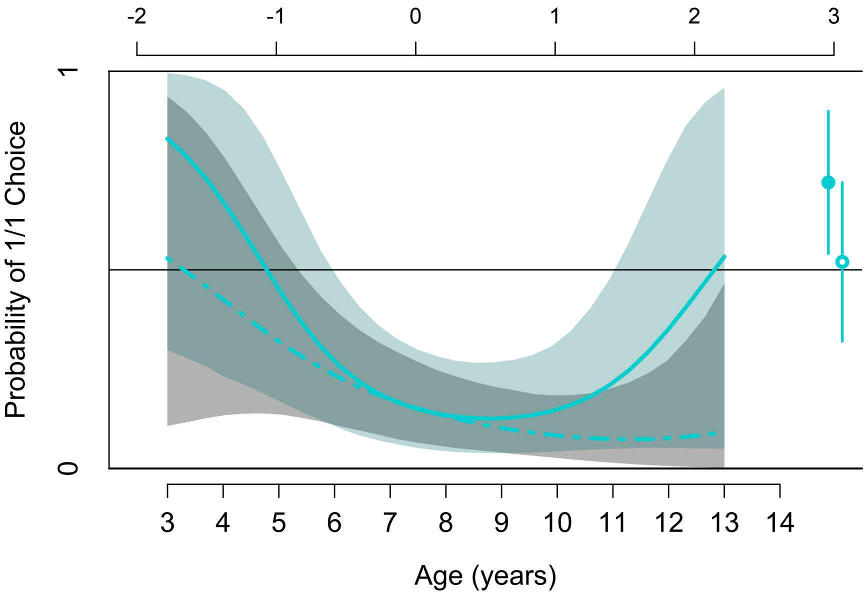


Figure S4(e): Aka

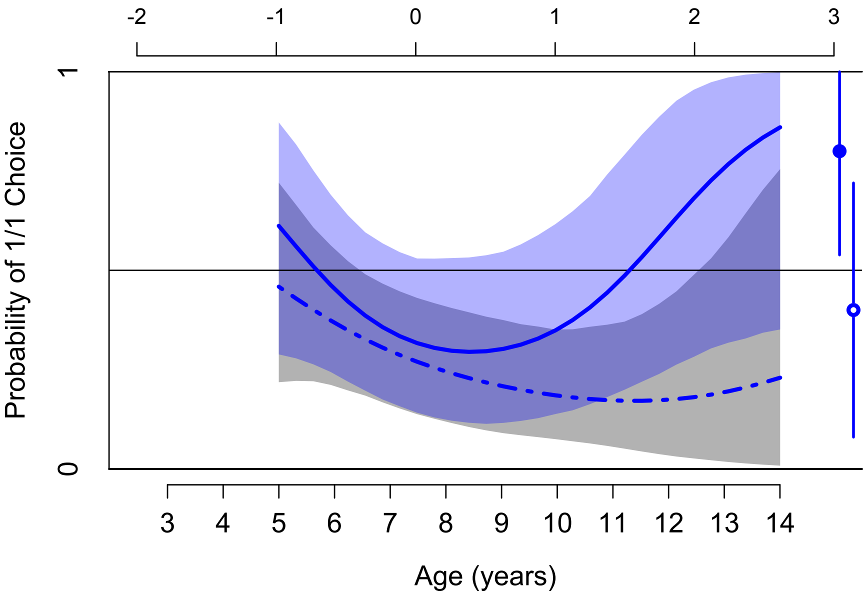


Figure S4(f): Martu

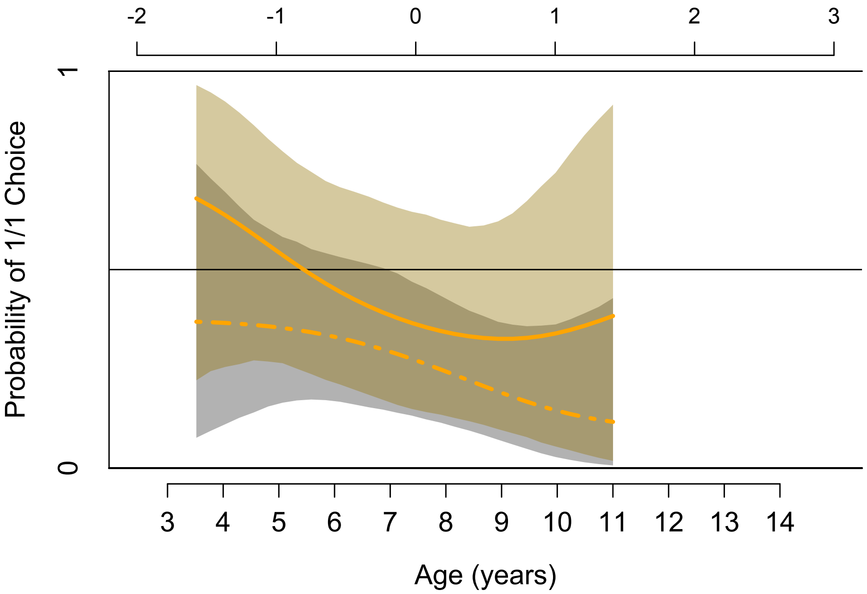
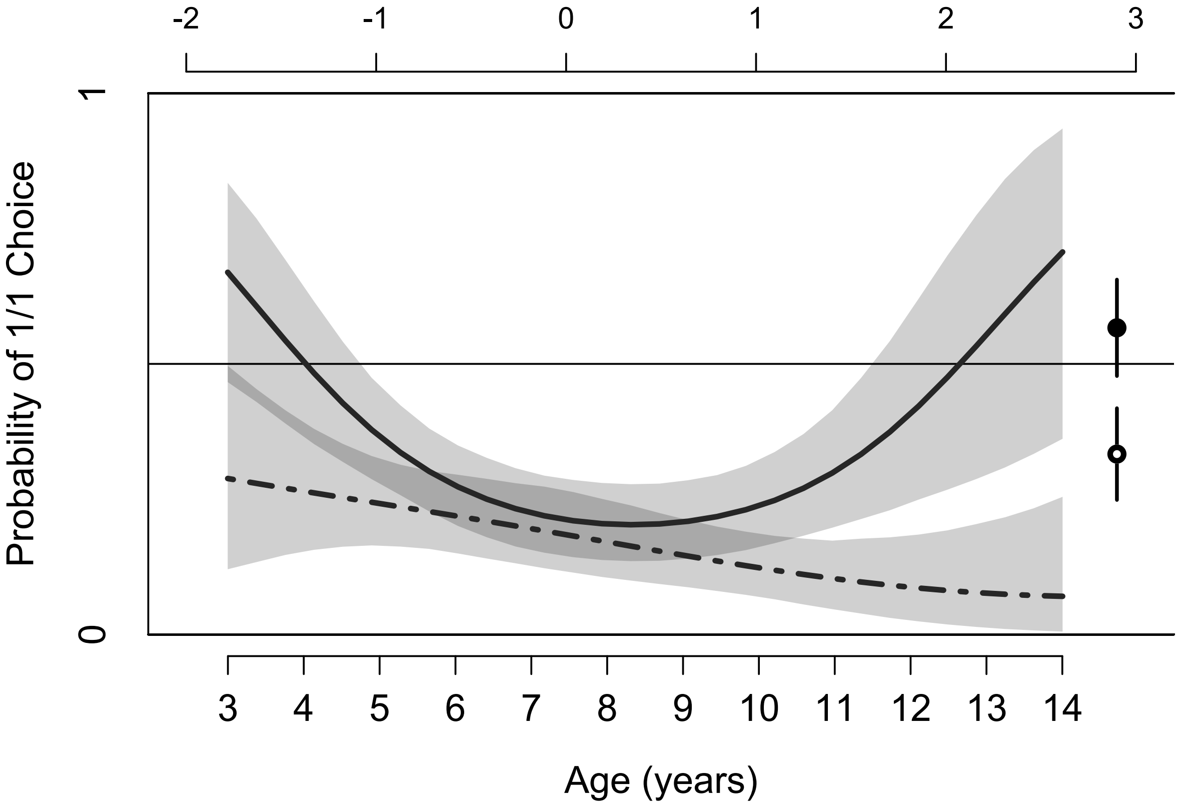


Figure S4(g): Age functions for the CSG, collapsing across all populations. Vertical axis is the estimated probability that children will choose the prosocial (1/1) outcome. Bottom horizontal axis is children's age (in years), and top horizontal axis is the equivalent value of CA. Age functions capture the estimated probability that children will select the 1/1 outcome as a function of age, with estimates extracted from Model 2 for the CSG (Table 4 main text, Table S4[a] above) for both the Social condition (solid line) and the Asocial condition (dotted line). Dots on the right side of the graph reflect the proportion of 1/1 choices actually made by adults.



Appendix 7: Analysis of Familiarization Trials.

The choices that participants made during familiarization trials provide some insight about their comprehension of the experimental task. However, some of these trials allow multiple interpretations that could lead Actors to select either outcome, even if they correctly understand the task. In particular, for three of the four trials a participant might choose one outcome if they wish to maximize personal payoff, but they might choose the other outcome if they wish to minimize inequity in payoff between themselves and the recipient. Only one trial avoids this confound between payoff maximization and inequity aversion (Familiarization #1 Social). We included four familiarization trials across the Social and Asocial conditions:

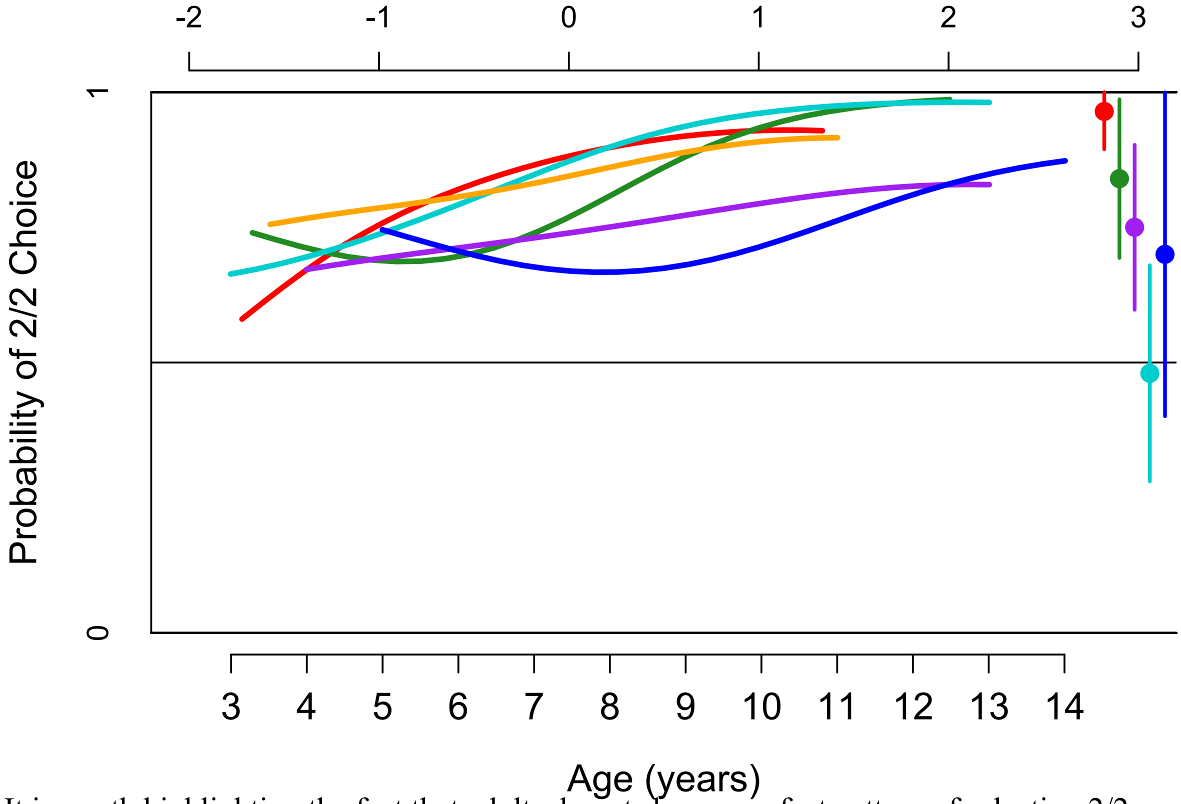
Table S5: Details about Familiarization trials.

Trial	Payoffs	Rationale for choice
Familiarization #1 (Social Condition)	1/1 vs. 2/2	2/2 maximizes payoff to self *without increasing inequity*
Familiarization #2 (Social Condition)	1/0 vs. 2/0	2/0 maximizes payoff to self, 1/0 minimizes inequity
Familiarization #1 (Asocial Condition)	1/1 vs. 2/2	2/2 maximizes payoff to self, 1/1 minimizes inequity
Familiarization #2 (Asocial Condition)	1/0 vs. 2/0	2/0 maximizes payoff to self, 1/0 minimizes inequity

Only the first trial, Familiarization #1 Social, affords an income-maximizing trial that doesn't increase inequity. All three other trials result in the Actor receiving a higher payoff than the Recipient. If the Actor and Recipient are both present (Social condition), then choosing 2/0 over 1/0 increases inequity between the Actor and the Recipient, because at the end of the trial the Actor's payoff has been increased by two while the Recipient's payoff has not increased. If the Actor had selected 1/0 over 2/0 then there would be less inequity, as the Actor's payoff would only increase by one. This also applies if the Actor chooses 2/0 over 1/0 when no recipient is present (Asocial condition), because the Actor's payoff increases by two while the Recipient's payoff does not increase (the physical presence of the Recipient being irrelevant in this case). This argument also applies to the 2/2 vs. 1/1 choice in the Asocial condition. To avoid this potential confound between self-interest and inequity aversion, we focus only on Familiarization #1 (FAM1) Social.

In Figure S5(a) (Figure 3 in main text) we plot the age functions for this trial, with the outcome measure reflecting the probability that participants will select 2/2 (coded as '1') over 1/1 (coded as '0'). It is clear from Figure S5(a) that children generally prefer 2/2 over 1/1, and that this tendency generally increases with age across the six populations. Figures S7(b-g) plot these functions separately for each population, including estimated 95% confidence intervals for the functions. These confidence intervals reveal that for all of the populations, by about age 5 children show evidence of selecting 2/2 with a probability above 0.5. This indicates that they show a systematic bias towards the self-maximizing outcome, and provides strong evidence that they comprehend the task. The age function for the Aka (one of our smaller samples) only briefly deviates above 0.5 probability between ages 5-7 and 11-13 years, but the pattern is very consistent with the other populations.

Figure S5(a): Age functions for all six populations for the FAM1 Social trial (same as Figure 3 from main text). Each colored line corresponds to children from a separate population, and the dots on the right hand side correspond to the mean choices of adults from each population in the same task ***(****lines above and below each dot corresponds to 95% confidence intervals)****. The vertical axis on each plot corresponds to the estimated probability that children will choose the income-maximizing (2/2) outcome. The bottom horizontal axis corresponds to children's age (in years), and the top horizontal axis corresponds to equivalent values of CA.*



It is worth highlighting the fact that adults do not show a perfect pattern of selecting 2/2 over 1/1. It is highly unlikely that adults were unable to comprehend the nature of the task, meaning that we must not require perfect rates of choosing the income-maximizing outcome for deciding whether or not children understood the experiment. It is possible that during these trials participants may sometimes "overthink" the task (plausible as the task is very simple especially for adults and adolescents), or they may be testing what happens when they select an obviously 'incorrect' outcome. They may also simply be exploring the range of available choices.

Figures S5(b-g): 95% Confidence intervals for age functions in Figure 3 from main text. See Table S4(g) for associated models. The vertical axis on each plot corresponds to the estimated *probability that children will choose the income-maximizing (2/2) outcome. The bottom horizontal axis corresponds to children's age (in years), and the top horizontal axis corresponds to equivalent values of CA. The solid line and the colored shaded area in each plot correspond to the estimated probability and 95% confidence interval. The dots on the right side of each plot correspond to adults choices, and the lines above and below them correspond to 95% confidence intervals.*

Figure S4(a): Los Angeles

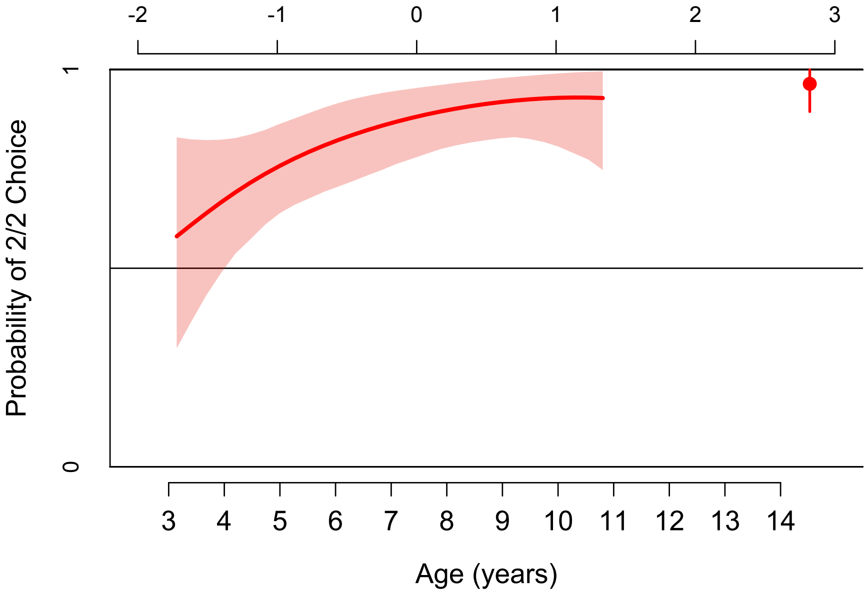


Figure S4(b): Fiji

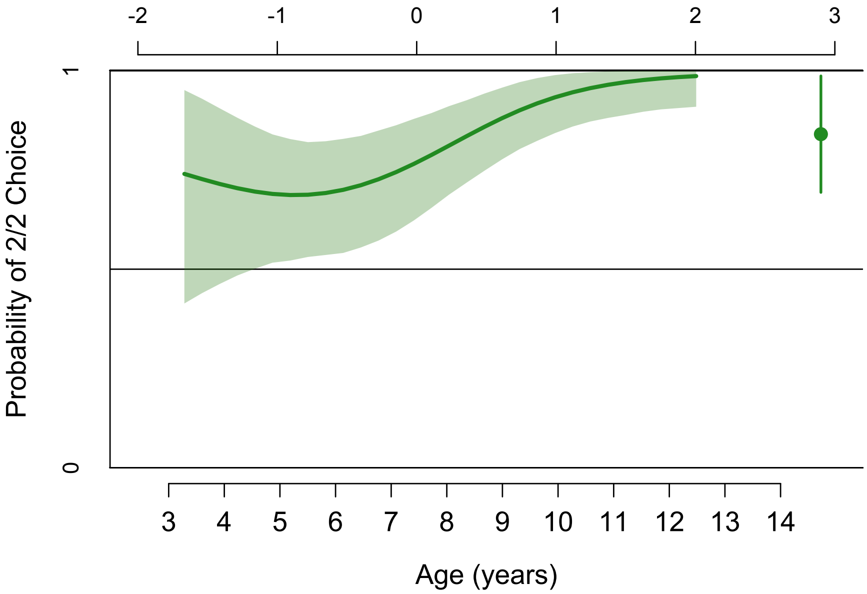


Figure S4(c): Himba

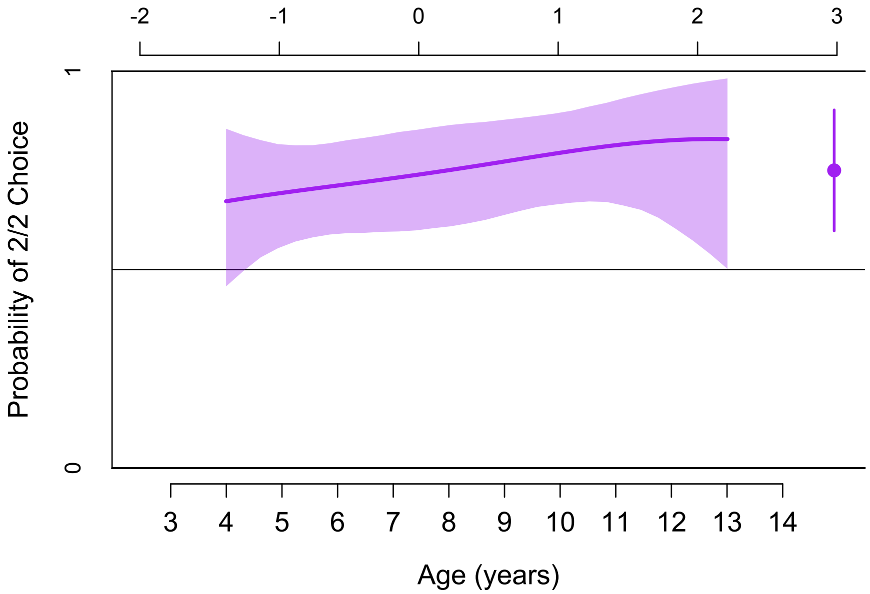


Figure S4(d): Shuar

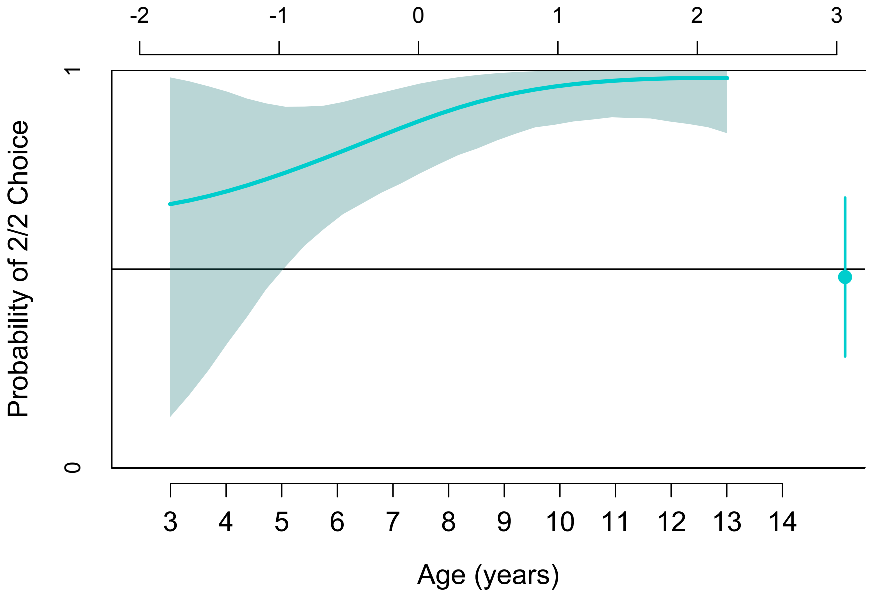


Figure S4(e): Aka

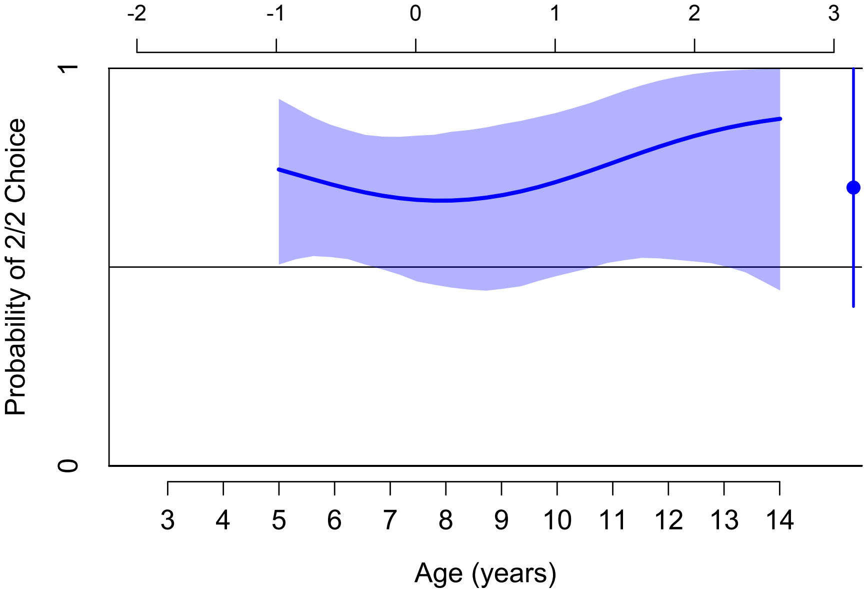
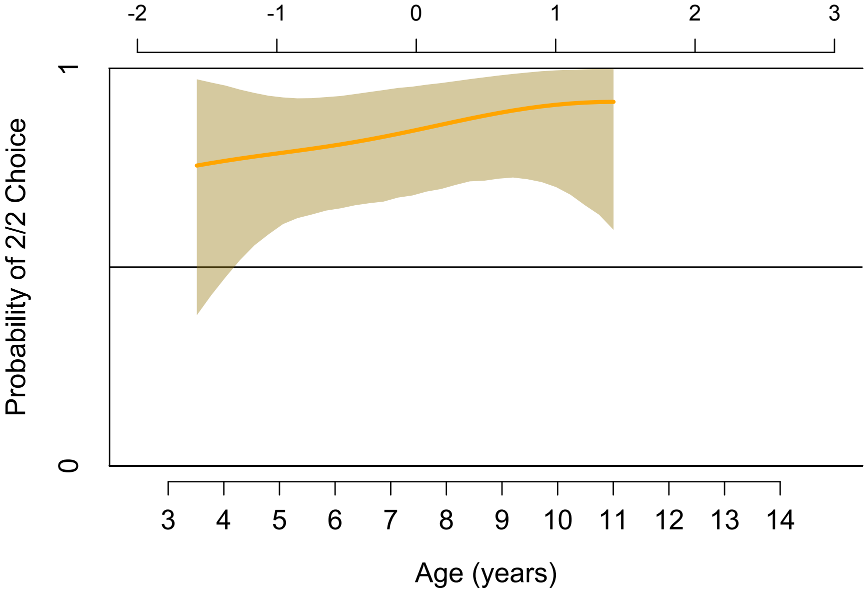


Figure S4(f): Martu



Comparison across all Familiarization Trials: For completeness, we also compare the patterns of development across our societies in the three other familiarization trials: Familiarization #2 Social, Familiarization #1 Asocial, and Familiarization #2 Asocial.

As already noted, we hesitate to draw conclusions from these trials because they are confounded: participants might select the income-maximizing outcome because it resulted in the highest payoff for them, or they might select the income-minimizing outcome because it reduced inequity between the participant and the recipient. If subjects were motivated by inequity aversion on these trials, then the probability of selecting income-maximizing outcomes (2/2 or 2/0) should fall among older children. This is because a growing number of studies have found increasing rates of inequity aversion among children aged about 8 years and older (22, 23, 24; references from main text). However, we should not see such a pattern in Familiarization #1 Social, as inequity aversion does not come into play here. Additionally, we might expect to see a greater drop in the probability of choosing 2/2 or 2/0 for those populations that are most prosocial in the CSG (behavior that may also be driven partly by inequity aversion).

Figures S5(h-k): The vertical axis on each plot below corresponds to the estimated probability that children will choose the income-maximizing (2/2) outcome. The bottom horizontal axis corresponds to children’s age (in years), and the top horizontal axis corresponds to equivalent values of CA.

***Fig.** S5(h) Familiarization #1 (Social)*

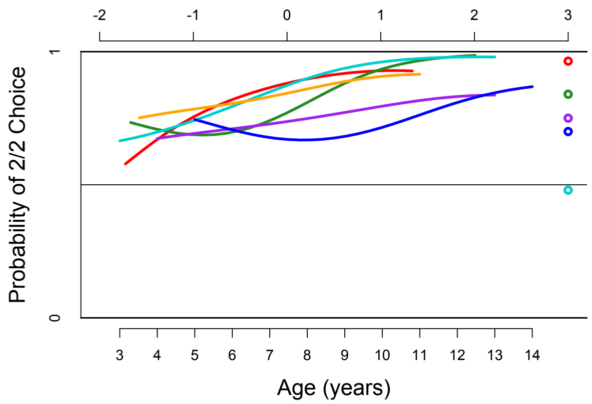


***Fig.** S5(i) Familiarization #2 (Social)*

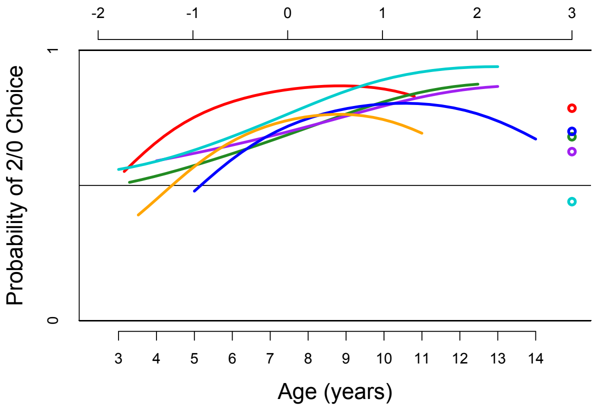


***Fig.** S5(j) Familiarization #1 (Asocial)*

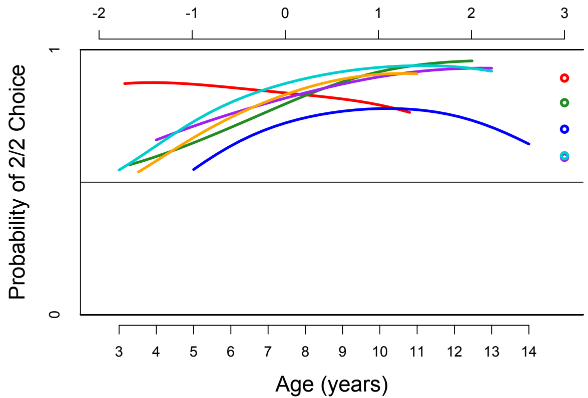
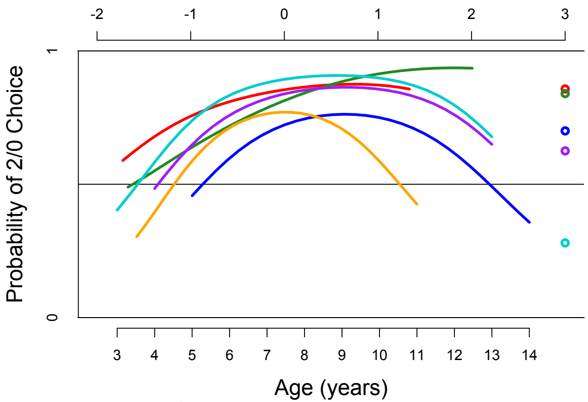
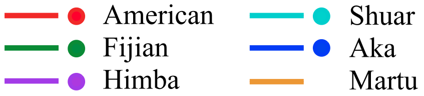


***Fig.** S5(k) Familiarization #2 (Asocial)*





We do see some evidence of this overall pattern. We only observe relative drops in the probability of income-maximizing choices among older children in Figures S7 (i-k), where we would have predicted it. We also see larger drops among the populations that show a stronger shift towards prosociality among older children in the CSG: the U.S. and Aka, with some evidence of drops also among the Martu, Himba, and Shuar. Notably, the only population with no obvious drop in the probability of 2/2 and 2/0 choices is Fiji, the population that showed little to no evidence of a shift towards prosocial choices in older children in the CSG.

Finally, we again note that the distribution of adults' choices in these familiarization trials generally reflect the distribution of predicted choices among adolescents, though there is more variability than in the CSG. This suggests that older children and adolescents were not simply confused, and it is plausible that they were making purposeful choices, just not ones focused exclusively on income-maximization. However, the fact that adults also made variable choices in the familiarization trials reinforces the need for caution when interpreting behavior such as that represented in Figures S7(h-k), and we thus hesitate to draw strong inferences from these trials.

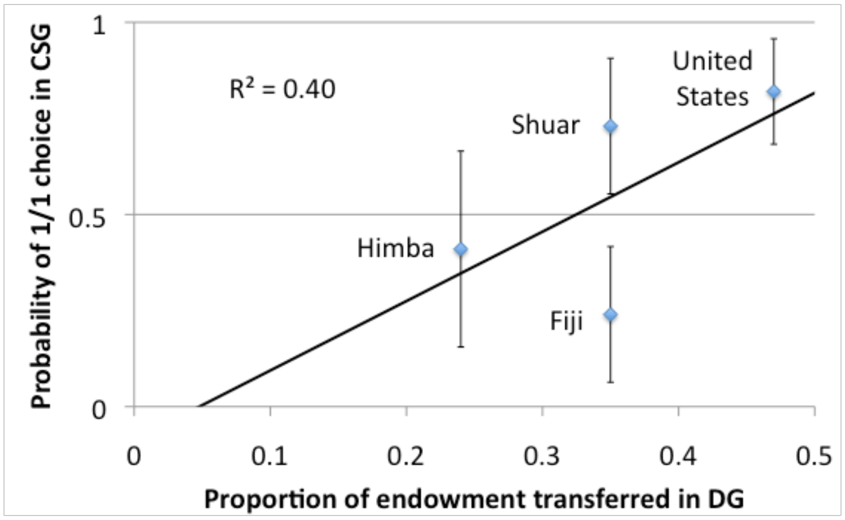
Appendix 8: The behavior of adults in the Costly Sharing Game (CSG) and the anonymous Dictator Game (DG).

Table S6: Adults' behavior in the CSG and the anonymous DG (data drawn from (Henrich et al., 2010a); ‡ data collected as part of the current study, see Appendix 9 for methods).*

	Proportion of 1/1 choices in CSG (Social Condition)	**Proportion of endowment shared in anonymous Dictator Game**
	Mean (SE) N	**Mean (SE) N**
United States	**.82 (.07) 28**	**.47 (.03) 15 ***
Shuar	**.73 (.09) 26**	**.35 (.04) 21 ***
Himba	**.41 (.13) 32**	**.24 (.04) 32 ‡**
Fiji	**.24 (.09) 25**	**.35 (.03) 35 ***

Figure S6: The proportion of adults' 1/1 choices in the CSG (vertical axis) are plotted against the behavior of adults from the same populations in the Dictator Game (horizontal axis). Values are drawn from Table S6.

SI Figure S6:



The positive relationship in Figure S6 between offers in the DG and the probability of a 1/1 choice in the CSG suggests that both tasks are measuring similar underlying prosocial tendencies (and similar variation in those tendencies across populations.

However, a better approach to exploring the relationship between behavior in the DG and CSG would us a regression model that limits estimates to possible outcomes (e.g., doesn't estimate probabilities of 1/1 choices in the CSG of <0 or >1), and which accounts for uncertainty in the measures of both the DG and the CSG. To do this we use regressions to first estimate behavior in the DG within each of our four populations (Table S7). We then use the posterior distributions of the mean DG offers to predict CSG behavior in each population (Table S8). This procedure provides a conservative estimate of the strength of association between DG offers and CSG behavior, by accounting for uncertainty in the average DG offer (Figure S7 for model structure).

Figure S7: Model structure. In this model, $\boldsymbol{y_{ik}}$ *is the i-th DG offer from population k, and* $\boldsymbol{x_{jk}}$ *is the j-th CSG choice from population k (*$\boldsymbol{\pi_{jk}}$ *is the expected value of* $\boldsymbol{x_{jk}}$*). Fitting this model using Hamiltonian Monte Carlo allows us to integrate over uncertainty in* $\boldsymbol{\mu_k}$ *(the mean DG offer for each population) when estimating* $\boldsymbol{\beta}$*.*

$$\boldsymbol{y_{ik} \sim \mathrm{Normal}(\mu_k, \sigma)}$$

$$\boldsymbol{x_{jk} \sim \mathrm{Bernoulli}(\pi_{jk})}$$

$$\boldsymbol{\log \frac{\pi_{jk}}{1 - \pi_{jk}} = \alpha + \beta \mu_k}$$

Table S7: Estimated DG Offers for each population. These largely correspond to the mean DG offers in Table S6, though the Monte Carlo sampling means that the estimated values here differ somewhat from the actual values. The actual values fall within the confidence intervals of our estimates.

DV: Proportion of Endowment Given in Dictator Game N=103	**Coef. (95% CI)**
Intercept: United States	.48 (.40 : .57)
Intercept: Shuar	.39 (.32 : .45)
Intercept: Himba	.27 (.20 : .33)
Intercept: Fiji	.30 (.25 : .36)

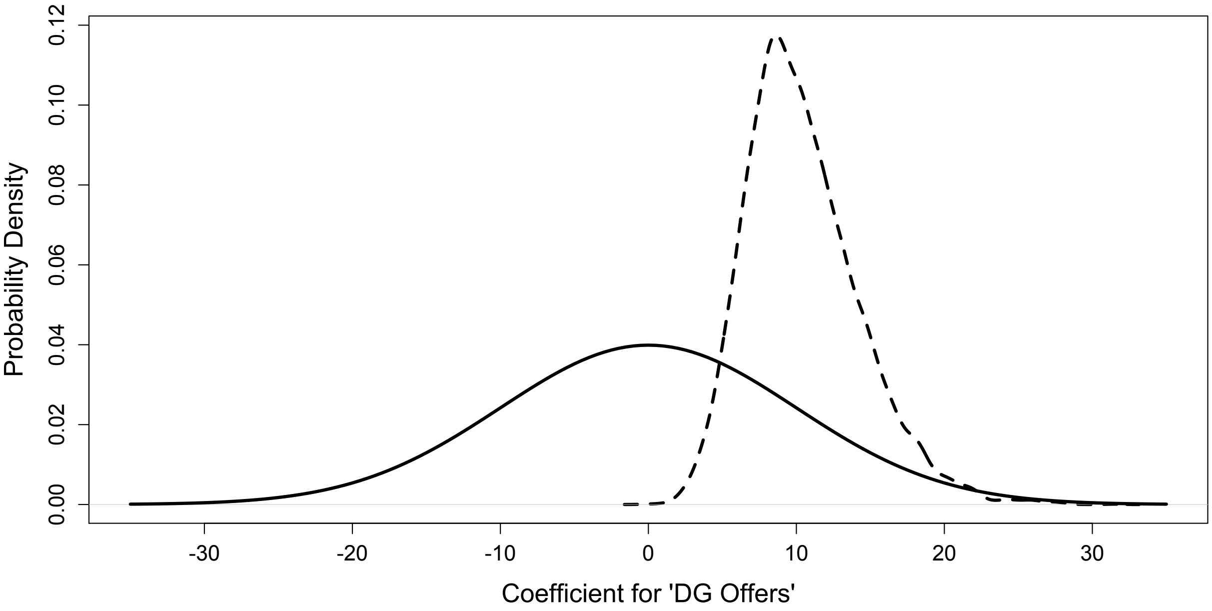
Table S7: Estimated DG Offers for each population, and estimated variation both within and across populations. These estimates largely correspond to the mean values presented in Table S6, though the sampling approach means that the estimated values here differ somewhat from the actual values (though the actual values fall within the estimated confidence intervals of our estimates).

DV: Probability of a 1/1 choice in the Costly Sharing Game N=120	**Coef. (95% CI)**
Intercept	-3.48 (-6.55: -1.28)
DG Offers	10.43 (4.32 : 19.10)

The parameter *DG Offers* has a positive estimate likely to be above zero, again suggesting a positive relationship between offers in the DG and choices in the CSG, but now incorporating uncertainty from both the DG and CSG measurements. Model results are based on 2000 samples from four chains, and we can plot both the assumed prior distribution (solid line, Figure S8) and the actual observed distribution (dashed line, Figure S8) of coefficient values for

DG Offers across the 8000 samples modeled. Figure S8 shows that the actual distribution of coefficients for DG Offers is likely to be positive, as the density of the probability distribution that is near and below 0 is very small.

Figure S8: Prior distribution (solid line) and Posterior/observed distribution (dashed line) for the parameter DG Offers in Table S7.



Appendix 9: Methods for the anonymous Dictator Game conducted with the Himba.

Participants: 32 individuals participated (16 women). Subjects received a small gift (500 kg sugar) as a show-up fee, in addition to payoffs obtained as part of the study.

Experimental task: Subjects were informed that they would be allocated an endowment of 10 NAD (worth $1.30), and could keep the full amount or give some amount to another anonymous individual within their community (the recipient). One investigator laid 10 coins down on a wooden pallet, counting as they laid them down one by one. In Familiarization trials, subjects were asked a series of questions about hypothetical allocation decisions to make sure participants understood the game. In the Test trials, 10 coins were counted out on the pallet, and once the coins were in place, the subject was allowed to make a decision about how to allocate the coins to themselves and the recipient.

Anonymity: Both subjects and recipients were kept fully anonymous. Additionally, investigators were blind to the allocation decisions made by particular subjects. Before subjects made their choices, the investigators moved away from the pallet and stood with their back to the subject at a distance of about 5-8 meters. The subject then placed the coins that they wanted to keep in their pocket, and placed coins that they wanted to allocate to the anonymous recipient in an opaque plastic bag, which was half-filled with popcorn or rice to muffle the sound and obscure the weight of coins being placed in the bag. The subject tied up the bag and deposited it in a box beside the pallet. No identifying information, except the subject's sex, was placed on the bag. Recipient payoffs, transferred in the same opaque bags, were distributed three weeks later to community members at a public gathering open to all those living in the area.

3

TASK DESIGN INFLUENCES PROSOCIALITY IN CAPTIVE CHIMPANZEES (PAN TROGLODYTES) [3]

Introduction

The literature on social behavior in chimpanzees (*Pan troglodytes*) shows clearly that cooperation is common in these animals. They cooperate when patrolling territorial boundaries and attacking neighboring groups (Watts and Mitani, 2001); collaboratively hunting small prey (Boesch and Boesch, 1989); sharing meat and other foods (Gilby, 2006; Mitani and Watts, 2001; Nishida and Turner, 1996); exchanging grooming for other valuable resources (de Waal, 1997; Mitani, 2009, 2006); and jointly guarding mates (Watts, 1998). This rich record of cooperation drawn from observational studies does not fully answer the question of what social preferences motivate this behavior, however. Studying naturalistic interactions is necessary to understand chimpanzee social behavior, but understanding the motivations for this behavior requires studies that control parameters that are theoretically relevant to cooperative mechanisms and the evolutionary mechanisms that would favor them, such as the relative benefits and costs that

---

[3] Chapter 3 is under review:
House, B.R., Silk, J.B., Lambeth, S., Schapiro S.J., under review. *Task Design Influences Prosociality in Captive Chimpanzees (Pan troglodytes)*

cooperative acts have for individuals and other group members. Experiments with captive animals offer opportunities to control these variables.

Some experiments present captive chimpanzees with choices that have different material payoffs for themselves and others, and the choices animals make reveal their underlying preferences. One such study allowed animals to select between two different payoff outcomes, both composed of fully visible food items (Silk et al., 2005). One outcome (the “1/1” option) delivered a food reward to the animal making the selection (the Actor) and an identical reward to a familiar group member (the Recipient). The other outcome (“1/0”) delivered one reward to the Actor, but nothing to the Recipient. The selection between two payoff distributions like these is often referred to as the Prosocial Choice (PC) task, and it offers Actors three choices: (a) select the 1/1 outcome; (b) select the 1/0 outcome; or (c) do nothing. The key measure in this experiment was whether subjects chose 1/1 more than 1/0. None of the chimpanzees from two different captive facilities differentiated between the Control and Test trials, suggesting that they were indifferent to the welfare of others. These results were replicated in the same populations using a slightly different protocol (Vonk et al., 2008), and in different populations (Jensen et al., 2006; Yamamoto and Tanaka, 2010), all with procedures that used visible food rewards.

One exception to this pattern of results with the PC task is a recent study in which Actors were presented with two tokens rather than a mechanism that delivered food items (Horner et al., 2011). One token could be traded with an experimenter to deliver a 1/1 outcome, while the other could be traded to deliver a 1/0 outcome. No food rewards were directly visible to Actors as they were hidden in opaque wrapping. Actors selected the 1/1 outcome more frequently in the Test condition than in a Non-Social Control condition, implying that animals were prosocial in this study. The use of non-visible food rewards could be important, as the visibility of food rewards has previously been shown to substantially impact animals' choices in experimental tasks (Boysen et al., 2001).

Other kinds of experimental designs lead to different conclusions about chimpanzee social preferences. In a body of experiments called Instrumental Helping (IH) tasks Actors had two options: (a) assist a conspecific or (b) do nothing. In these studies chimpanzees seem to be sensitive to outcomes obtained by others because they provide help to Recipients in a variety of different ways: unlocking a door obstructing access to food rewards (Warneken et al., 2007), releasing food rewards so they slide within reach of a Recipient (Melis et al., 2011), pulling a handle to help a Recipient move a food reward within reach (Greenberg et al., 2010), or transferring a tool to a partner who needs it to obtain a food reward (Yamamoto et al., 2012, 2009). In these studies Actors do not obtain food rewards for themselves when they provide help to a conspecific, and it is possible that this explains why prosociality is observed in IH tasks but not PC tasks (Yamamoto and Tanaka, 2009b). Recent research suggests that the mere presence of desirable food does not influence subject animals' willingness to provide help (Cronin, 2012; Melis et al., 2011), but Melis et al. (2011) noted that animals may be more influenced by obtaining food for themselves than by the mere presence of food rewards.

In a third body of experiments referred to as Inequity Aversion (IA) studies, chimpanzees may refuse to participate in a task if they receive a lower payoff for their participation than a conspecific partner, and it has been argued that such behavior is evidence for an aversion to inequity (Brosnan et al., 2010). Inequity aversion might stabilize cooperation by incentivizing individuals to reject inequitable offers made by their partners in favor of other alternatives (Brosnan, 2011; Yamamoto and Takimoto, 2012). If chimpanzees are averse to inequity then their social preferences must be sensitive to payoffs obtained by others. However, one study was unable to replicate results using similar procedures (Bräuer et al., 2009), and controversy remains both over how to interpret this phenomenon (Henrich, 2004) and the specific methods and contexts necessary for eliciting an aversion to inequity in chimpanzees (Bräuer and Hanus, 2012; Price and Brosnan, 2012).

Differences in methodology, task demands, and rewards obtained across these bodies of research make it very difficult to reconcile divergent findings about chimpanzee sociality. PC tasks typically allow animals to obtain food rewards for themselves while delivering food to others, while in IH tasks Actors do not obtain food at the same time as do Recipients. The mere presence of food does not inhibit subjects' prosocial behavior (Melis et al., 2011), but it is possible that seeing rewards that animals can obtain for themselves obscures their prosocial behavior. This might also explain why one PC study using tokens and non-visible rewards found evidence of prosociality (Horner et al., 2011) where other PC studies did not. A second difference is that PC tasks present animals with a larger number of discrete choices (three choices) than do IH tasks (two choices). The complexity of the tasks may affect the chimpanzees' performance.

Here, we explore prosocial behavior in captive chimpanzees using a common procedure that manipulates payoff outcomes for Actors and Recipients in ways similar to the PC and IH tasks. We investigate whether chimpanzees are less prosocial when (a) tasks are more complex and (b) when animals obtain rewards for themselves, two hypotheses that may explain differences in observed levels of prosocial behavior across studies. No prior study has used a common research design to explore and compare findings from the PC, IH, and IA paradigms. Previous studies have explored whether the mere presence of food rewards reduces chimpanzee prosociality (Melis et al., 2011), but found no evidence that this underlies the differences in results from different research methods. We expand on this work by exploring whether animals are less prosocial when they concurrently obtain food rewards for themselves, in a manner that bridges prior work by Melis et al. (2011) and Jensen et al. (2006). Study 1 presents captive chimpanzees with a two-choice prosocial task comparable to IH tasks, and tests whether animals prefer outcomes that benefit others and/or are averse to inequity. We also test whether animals are averse to prosocial outcomes that result in disadvantageous inequity. Study 2 presents chimpanzees with a three-choice prosocial task more analogous to previous PC tasks. In both studies, we manipulate whether Actors obtain payoffs for themselves at the same time that they deliver payoffs to Recipients.

We also designed our tasks to be more naturalistic by testing animals with unconstrained access to their social group, allowing them substantial freedom of choice over when to participate and with whom to interact in the experimental trials. There is substantial evidence that partner choice is an important component of chimpanzee social interactions, both in the wild (Boesch et al., 2006; de Waal, 1997; Langergraber et al., 2007; Mitani, 2009, 2006; Nishida and Turner, 1996) and in laboratory experiments (Melis et al., 2006). Allowing animals complete freedom of choice over when to participate and with whom creates a more naturalistic social interaction than typically allowed when animals are isolated from their social groups. Such opportunities for partner choice could be more likely to elicit prosocial behavior, given prior evidence that animals engage in partner choice in similar tasks (Melis et al., 2006).

Study 1

Subjects were socially housed chimpanzees at the Michale E. Keeling Center for Comparative Medicine & Research (KCCMR) in Bastrop, TX. The apparatus was novel to all participants, but was similar to a familiar enrichment device. All animals in the group had free access to the apparatus, and individuals thus varied substantially in the number of observations they contributed to the dataset (ranging from 1 to 124 observations; median: 31; see Appendix 1 Table S1 for details). Actor animals were able to deliver payoffs (pieces of apple) to themselves and/or other group members by using an apparatus consisting of two plastic food bins (the Actor bin and Recipient bin) that were anchored to the enclosure more than a full arm span apart (ensuring that the Actor could obtain rewards from only the Actor bin, (and not the Recipient bin). The apparatus could only be operated by one animal at a time.

Payoffs for Actors and Recipients varied across trials (Table 1). If Actors were prosocial they would have pulled more frequently on trials where the Recipient obtained a reward (0/1; Table 1) than on trials where the Recipient obtained nothing (0/0), and more than on trials where the Recipient obtained nothing but a food item was still visible (0/0[1]). 0/1 most closely conforms to the payoff outcome for IH tasks. If Actors desired the food payoffs then regardless of their social preferences they would have pulled the handle at near-ceiling levels when they themselves obtained rewards (1/0 and 1/1), but if chimpanzees are averse to disadvantageous inequity Actors would have pulled the handle less when Recipients obtained a greater payoff than they did (1/3).

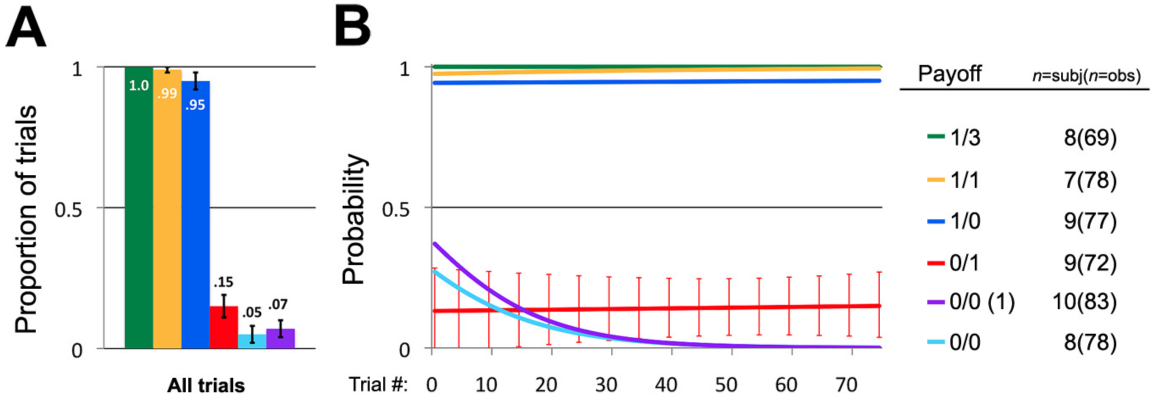
Table 1: Payoff distributions used in Study 1. We label these distributions using the convention of "(actor's payoff) / (recipient's payoff)."

Payoff distributions	Payoff for Actor	Payoff for Recipient
0 / 0	Zero	Zero
0 / 1	Zero	One reward
1 / 0	One reward	Zero
1 / 1	One reward	One reward
1 / 3	One reward	Three rewards
0 / 0 (1)	Zero	Zero (one inaccessible reward placed next to the bin)

Results

The outcome variable was the binary parameter *Actor Pulled Handle* that captured whether the Actor pulled the handle (coded as "1") or did not pull (coded as "0"). We used multi-level logistic regression models to explore how various parameters predicted the outcome variable, and included Actor identity as a random effect to control for non-independence of observations. *Actor's Trial Number* codes for the total number of test trials that Actors had previously participated in (across all six payoff distributions). Animals pulled the handle at different rates across the six payoff distributions (Figure 1A). Actors pulled the handle much more frequently when food was placed in the Actor bin (1/3, 1/1, 1/0) than when no food was placed in the Actor bin (0/1, 0/0, 0/0[1]). Actors were two to three times as likely to pull when Recipients received rewards (0/1) than when they did not (0/0 and 0/0[1]).

Figure 1: Data from Study 1. (1A) Proportions of all trials during which actors pulled the handle to operate the apparatus, for each of payoff distribution. Error bars reflect one SE of the mean but do not control for non-independence, though models suggest there is little between-subjects behavioral variation (see ESM Table S7). Numbers of subjects and observations are listed on the right. (1B) Logistic function modeling the effect of Actor's Trial Number on the probability that actors will pull the handle. These probabilities are comparable to the proportions from Figure 1A. Y-axis represents the probability that actors will pull the handle. X-axis represents the number of trials that an actor has received previously (across all payoff distributions). Error bars around the red function for 0/1 are estimated 95% confidence intervals, and control for non-independence.

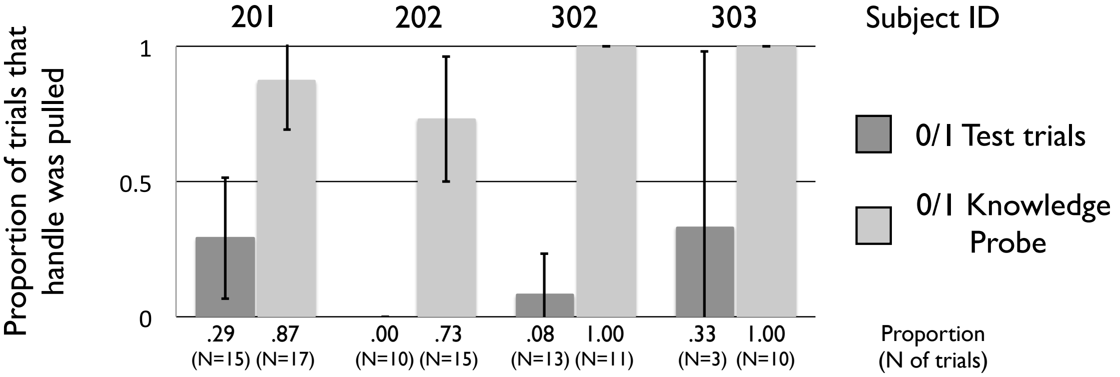


Our regression model (see Appendix Table S7) uses 0/0 as a reference point and asks whether Actors pull the handle more or less frequently for the five other payoff distributions than they do for 0/0. The coefficient for 0/0(1) is positive but smaller than its standard error (***β***=.34, SE=.67), suggesting that animals were not substantially more likely to pull the handle when a food item was present but inaccessible. The coefficient for 0/1 is positive and larger than its standard error (***β***=1.22, SE=.61), indicating that Actors were more likely to pull the handle when a food item was present *and* placed inside the Recipient bin. The coefficient for 0/1 (Appendix Table S7) translates to an Odds Ratio of 3.38 (SE: 2.07), indicating that animals were 3.38 times more likely to pull the handle for 0/1 than they were for 0/0 (see Appendix Table S2 for analysis of the robustness of this result). The coefficients were all positive and much larger than their standard errors for 1/0 (***β***=5.96, SE=.77) and 1/1 (***β***=7.42, SE=1.16), and Actors pulled on every

1/3 trial. This indicates that Actors were very likely to pull when they received food. The random effect parameter reflects the variation in the outcome measure across individuals, and its coefficient is smaller than its SE suggesting that subjects were largely similar in their behavior (coefficient=.30, SE=.41; see Appendix Table S7). *Actor's Trial Number* predicted animals' behavior to different degrees across trial types (Figure 1B, see Appendix Table S2 for models). For 1/1, 1/0, 1/3, and 0/1, the Actor's likelihood of pulling the handle changed little as trial number increased. For 0/0 and 0/0(1)—the two trial types in which neither Actor nor Recipient ever obtained payoffs—the probability of pulling for 0/0 and 0/0(1) dropped substantially over the course of the study, eventually approaching zero.

In the Knowledge Probe we explored whether Actors were more likely to pull the handle when the apparatus was modified to allow them to directly access food from the Recipient bin (four animals were willing to be isolated for this test; see Methods for details). Animals were indeed more likely to pull the handle for 0/1 in the Knowledge Probe than they had been in the Test trials (Figure 2). For 1/1, animals pulled the handle on every trial in both the Knowledge and Test trials, and for 0/0 Actor animals pulled the handle at comparably low rates in both the Knowledge and Test trials (see Appendix Figure S2).

Figure 2: Data from Knowledge Probe (Study 1). Proportions of trials during which Actors pulled the handle to operate the apparatus during the Test trials (dark grey bars) and in the Knowledge Probe (light grey bars), for the 0/1 payoff distribution only. Each pair of bars corresponds to one of the Actors who participated in the Knowledge Probe. Error bars reflect 95% confidence intervals.



Discussion

Actors were relatively more likely to pull on trials in which food was in the Actor bin. On those trials in which no food was placed in the Actor bin, Actors were also relatively more likely to pull the handle on trials in which food was placed in the Recipient bin. Animals also demonstrated their comprehension of the task by holding constant in their tendency to pull the handle when Recipients obtained a benefit (0/1), while also becoming less likely to pull the handle when neither Actor nor Recipient benefitted (0/0 and 0/0[1]). This shows that Actors were attentive to how their choices impacted group members, and learned to ignore trials that didn't deliver benefits to conspecifics or to themselves. Additionally, Actors were more willing to pull the handle in the Knowledge Probe than in the Test trials when food was placed only in the Recipient bin (0/1), showing that they understood that pulling the handle caused food in the Recipient bin to become available. These results indicate that animals showed some tendency to deliver benefits to conspecifics, and they also provide three independent confirmations that animals in this study understood the causal relationship between pulling the handle and making food accessible to Recipients. Additionally, Study 1 suggests that captive chimpanzees display behavior that confers prosocial outcomes within a task similar to the one used in IH tasks.

In Study 2 we explore whether chimpanzees also behave prosocially when we modify the same apparatus to create a situation that more closely resembles the PC task (Brosnan et al., 2009; Silk et al., 2005; Yamamoto and Tanaka, 2010). We also explore whether chimpanzees will be more prosocial in the PC task if they do not obtain food for themselves concurrently with delivering food to conspecifics, as proposed by Melis et al. (2011). Jensen et al. (2006) previously tested this hypothesis and found no support for it, but using a different method than Melis et al. (2011). Here we extend these findings by exploring whether we can obtain more evidence of prosociality in chimpanzees than did Jensen et al. (2006) by using our apparatus from Study 1, which (like the apparatus used by Melis et al.; 2011) elicits non-zero rates of helping behavior.

Study 2

Participants were drawn from three different social groups at KCCMR, each containing 4-7 animals. During trials, Actors were faced with a choice between two of the same 2-bin mechanisms used in Study 1, arranged side by side (see Appendix Figure S2). When an Actor pulled the handle that operated one of the mechanisms the device automatically retracted the other mechanism's handle, preventing Actors from operating more than one mechanism within a single trial. All animals in the social group again had free access to the apparatus, and there was again substantial variation in the number of observations that the Actors contributed to the dataset (ranging from 23 to 304 observations; median: 139; see Appendix Table S4 for details).

Payoffs for Actors and Recipients (pieces of banana) again varied across trials (Table 2). If subjects preferred 1/1 over 1/0 (Reward condition), but not 1/1 over 1/0(1) (Reward [food balanced] condition), this would suggest that they were simply biased by the greater quantity of food. The No Reward condition was equivalent to the Reward condition except Actors obtained no food rewards for themselves regardless of which option they selected, and the Control condition evaluated animals' comprehension of the task by providing them with choices that only vary in their payoffs for themselves.

Table 2: Payoffs distributions used in Study 2. We label these distributions using the convention of "(actor's payoff) / (recipient's payoff)."

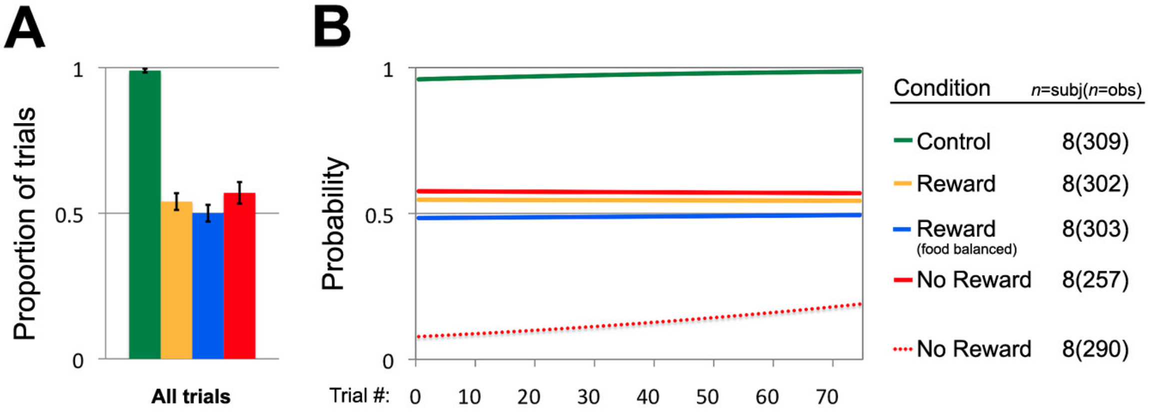
Condition	Payoff Distribution	Description
Reward	1/1 vs. 1/0	Prosocial outcome vs. Non-Prosocial outcome. Rewards for actor.
Reward (food balanced)	1/1 vs. 1/0(1)	Prosocial outcome vs. Non-Prosocial outcome. Rewards for actor. *One food item next to the recipient bin for the 1/0 payoff.*
No Reward	0/1 vs. 0/0	Prosocial outcome vs. Non-Prosocial outcome. No rewards for actor.
Control	1/1 vs. 0/1	Payoff for self vs. No payoff for self.

Results

The primary outcome variable was the binary parameter *Actor Chose Prosocial Outcome*. For the Reward, Reward (food balanced), and No Reward conditions, prosocial choices (1/1, 1/1, and 0/1) were coded as "1", and non-prosocial choices (1/0, 1/0(1), and 0/0) were coded as "0". For the Control condition, 1/1 was coded as "1" and choices of 0/1 as "0". Our primary analyses exclude trials in which Actors selected neither of the payoff distributions, and we separately analyze Actors' choices to "do nothing." Multi-level logistic regressions controlled for the non-independent observations, and we investigated how *Actor's Trial Number* predicted animals' behavior in each of the four conditions. Contact the corresponding author for raw data.

Animals behaved differently across the payoff distributions, but only displayed a substantial preference for one mechanism over the other when their own payoffs were at stake (Control condition; Figure 3A). For trials in which only the Recipient's payoffs were at stake, Actors chose the prosocial outcome slightly more than 50% of the time for the Reward and No Reward conditions, but exactly 50% of the time for the Reward (food balanced) condition. Across conditions, the estimated probability that Actors chose the prosocial outcome was very stable as trial number increased, and the random effect parameter was small, suggesting that there was little behavioral variation across subjects (Figure 3B, see Appendix Table S5 for regression models). However, increasing trial number did lead to a greater probability that Actors would do nothing (i.e. pull neither of the handles) in the No Reward condition (Figure 3B; see Appendix Table S6 for regression model).

Figure 3: Data from Study 2. (3A) Proportion of trials during which actors selected the prosocial outcome (for Reward, Reward [food balanced], and No Reward) or the self-maximizing outcome (for Control). Numbers of subjects and observations are listed on the right. (3B) Logistic functions modeling the probability of actors' choices as a function of Actor's Trial Number. Solid lines depict the estimated probabilities that actors will select the prosocial or self-maximizing outcome, which are comparable to the proportions from Figure 3A. Y-axis represents the probability that actors will select the prosocial or self-maximizing outcome, and the X-axis represents the number of trials that an actor has received previously. The dotted line depicts the estimated probability that actors will do nothing in the No Reward condition.



Discussion

These results suggest that chimpanzees understood the apparatus, because from the beginning of the study subjects systematically chose the outcome that provided rewards for themselves in the Control condition. Overall, animals were slightly more likely to select the prosocial outcome over the non-prosocial outcome in the Reward and No Reward conditions, but rates of selecting the prosocial outcome did not differ across these two conditions. Actors were not more likely to select the prosocial outcome in the Reward (food balanced) condition. This implies that in the Reward and No Reward conditions animals were not biased toward the prosocial outcome per se, but they were biased toward the location containing greater quantities of food, a phenomenon that has been reported previously in chimpanzees (Boysen et al., 2001, 1996; Silk et al., 2005).

General Discussion

We observed that captive chimpanzees acted to confer benefits on others in some laboratory contexts, and our findings suggest that methodological differences may help explain discontinuities in results across studies. These data are thus consistent with studies of wild animals, and with the results of laboratory studies showing prosocial behavior in IH studies (Greenberg et al., 2010; Melis et al., 2011; Warneken et al., 2007; Yamamoto et al., 2009) and one PC study (Horner et al., 2011). However, it is crucial to distinguish between prosocial behavior and prosocial preferences. Data from these kinds of laboratory studies (and also naturalistic interactions) cannot demonstrate the presence of psychological mechanisms based on prosocial preferences because these data do not disambiguate animals' various possible motivations. This problem has been highlighted recently by Heyes (Heyes, 2012), who points out that prior evidence for prosocial behavior (Horner et al., 2011) could be a product of conditioning—not a psychology motivated to deliver benefits to conspecifics. Chimpanzees may be adaptively prosocial even if they lack psychological mechanisms that value the welfare of others independently of their own welfare. This may be true in the current study as well as all previous studies in the literature. However, we emphasize that the psychological motives underlying prosocial behavior in chimpanzees can only be studied after first developing methods that consistently elicit it.

Although chimpanzees acted prosocially in Study 1 the rates of animals' prosocial choices were modest, which starkly contrasts with the prosocial acts frequently observed in laboratory interactions between humans (Camerer, 2003; Henrich, 2006; Henrich et al., 2005; Melis and Semmann, 2010). It is thus important to consider the different motives that could have generated the results. In Study 1 Actors were 10% more likely to pull the handle on 0/1 trials than 0/0 trials, implying that Actors were weakly influenced by the payoffs obtained by Recipients in this task. This could be motivated by reciprocity given that our subject animals were permitted unconstrained partner choice in non-anonymous interactions with long-term social partners. Detailed analyses of reciprocity would require additional data, as free partner choice resulted in particular pairs of Actors and Recipients being disproportionately represented in the dataset, reducing the necessary variation in Actor/Recipient pairings.

Our results are not consistent with the hypothesis that chimpanzees are less prosocial when they obtain rewards for themselves (Cronin, 2012; Melis et al., 2011), because in Study 2 Actors were equally non-prosocial in both the No Reward or the Reward condition. This replicates similar findings by Jensen et al. (Jensen et al., 2006) showing that animals were not more prosocial in a No Reward condition, but does so using an experimental apparatus that elicited low levels of prosocial behavior in Study 1. Instead, our finding that animals were more prosocial in Study 1 than in Study 2 suggests that more complicated experimental tasks may be less likely to elicit prosocial behavior in captive chimpanzees than are simpler tasks. In Study 1, animals showed a weak tendency to deliver prosocial outcomes when presented with a single apparatus that they could either use or not use (similar to IH tasks), but animals showed no evidence of prosociality in Study 2 where they had to select between two different apparatuses (similar to PC tasks). In Study 1, greater experience with the task led Actors to be less willing to pull the handle and operate the apparatus for the 0/0 and 0/0(1) payoff distributions, but not for 0/1. Actors thus learned to reduce their pulling when neither they nor a Recipient obtained a reward, but not when a Recipient obtained a payoff. This would predict that greater experience with the No Reward condition from Study 2 (0/1 vs. 0/0) should lead Actors to reduce their choices of 0/0 but not 0/1. The likelihood of selecting the prosocial outcome in the No Reward condition should have increased as the study progressed, but this did not happen. If anything, with greater experience Actors simply became more likely to do nothing.

Subject animals in Study 1 showed in a number of ways that they understood how to use the apparatus to access food in the Recipient bin, and the causal affordances of the apparatuses in Study 2 were identical to those of the apparatus used in Study 1. Thus, it is unlikely that the reduction in prosociality in Study 2 was due to an inability to understand that pulling the handle made food available to the Recipient, and it was more likely due to the introduction of an additional apparatus (i.e. a second payoff choice). The addition of a second apparatus and a third behavior alternative may have distracted the animals in Study 2 and prevented them from attending to the Recipient's payoff, and this implies that a particular kind of task complexity (i.e. number of discrete choices) may be particularly problematic for laboratory studies of prosociality in captive animals. Our results suggest that the different results obtained across PC and IH tasks is not due to the causal complexity of the experimental task, and it is rather due to differences in the number of alternative choices the task presents to subject animals.

These results are interesting to consider in light of recent findings showing that chimpanzees displayed a preference for prosocial outcomes in a variant of the PC task in which Actors obtain food when delivering food to Recipients (Horner et al., 2011), a result that appears to contradict those of the Reward condition that we used in Study 2 and also the results of prior studies using the PC task (Brosnan et al., 2009; Jensen et al., 2006; Silk et al., 2005; Vonk et al., 2008; Yamamoto and Tanaka, 2010). Our results showing more prosociality in Study 1 relative to Study 2 suggests that the most important methodological difference across these PC tasks from the literature is more likely to be task complexity, rather than the presence of visible food rewards. The use of tokens might be crucial, and differences in the use of this methodology could underlie controversial inconsistencies in how chimpanzees' behave in recent studies of the Ultimatum Game (Jensen et al., 2007; Proctor et al., 2013). Future research should explicitly test whether the use of tokens impacts prosocial behavior, in the same way that we have here tested other methodological hypotheses. One final possibility is that differences between the populations tested may play a role.

Our results are also inconsistent with studies finding inequity aversion in captive chimpanzees (Brosnan et al., 2010). If Actors in the current study had been averse to disadvantageous inequity, we would have expected them to pull the handle less for 1/3 trials than for 1/1 trials, a pattern similar to one reported when capuchin monkeys (*cebus apella*) were presented with a forced-choice between 1/1 and 1/3 (Fletcher, 2008). However, our chimpanzee Actors appeared insensitive to the difference between these payoff distributions, even though animals showed that they understood the apparatus used in Study 1. It has been argued that it is necessary for subject animals to be engaged in an experimental task for a response to inequity to be elicited (Brosnan et al., 2010), and consistent with this Actors in our studies are engaged in a task. However, the structure of our task is distinctly different from the standard task that elicits inequity aversion, in that there is no token exchange and the subject animal directly creates the inequitable outcome by delivering the rewards to themselves and a conspecific (Brosnan et al., 2010). It remains possible that the use of tokens and the role of the experimenter is required for the phenomenon to hold, but it remains a tantalizing theoretical question why laboratory elicitation of inequity aversion in chimpanzees might be constrained to these particular task designs (Bräuer and Hanus, 2012; Price and Brosnan, 2012).

These results suggest that asymmetries in evidence for prosocial behavior across different experimental tasks may be due to asymmetries in task complexity, but many other differences in methodology might also influence chimpanzee prosociality in laboratory tasks, and future work should explore additional possibilities. However, a larger issue still looms: despite the plausible need for cooperation in both captive and wild chimpanzee social groups, evidence for chimpanzee prosociality in the laboratory is still weak relative to evidence from the wild. A primary benefit of laboratory studies is that they permit the investigation of mechanisms and motivations underlying chimpanzees' naturally-occurring prosocial behavior, but to do this requires experimental tasks that capture prosocial behavior in such a way that costs and benefits can be manipulated in a controlled manner. The methods used here describe one way to approach this problem, while also employing a task that more closely emulates the freedom for partner choice that exists in the wild. Making experiments progressively more naturalistic should be a goal of laboratory work, and it will allow more powerful generalization of laboratory results to the behavior of wild chimpanzees.

Methods

This study protocol was approved by the Animal Research Committee of the University of California, Los Angeles (ARC approval permit #2011-036-01). Animals were not food-deprived, and had ad libitum access to food and water throughout the study. Supplementary fruits and vegetables were provided to animals once daily. Animals were housed in social group enclosures with both exterior and interior areas, and all animals participated in a behavioral management program to ensure their mental health and well-being. This program includes daily environmental enrichment procedures in which animals are provisioned in both indoor and outdoor areas with enrichment devices for resting and climbing (e.g., grass ground cover, multi-level wooden platforms, hammocks, hanging tires, brachiation bars, utility poles, ropes and cargo nets, barrels, and large culverts), along with additional foraging apparatuses (e.g., kong toys filled with food) which were changed regularly. Animals did not suffer during the study, which did not interfere with their normal activities except when animals were isolated during the Knowledge Probe in Study 1. The Knowledge Probe was only begun when animals did not show stress at being isolated, it lasted about 30 minutes, and it was halted if animals began to show any signs of distress. Animals were isolated in the exterior areas of their home enclosure, and had unconstrained access to light, food, water, and enrichment. No animals were harmed or sacrificed.

Study 1

Participants: Many participants had previously taken part in cognitive experiments, and some in prior experiments on prosocial behavior (Brosnan et al., 2009; Silk et al., 2005). Animals were housed in one all-female or one all-male social group of 6 or 7 animals whose membership had been stable for several years, and had access to food and water throughout the day.

Apparatus: Actor animals were able to deliver payoffs to themselves and/or other group members by using an apparatus consisting of two plastic food bins that were anchored to the enclosure more than a full arm span apart. When a handle was pulled by the Actor both bins pivoted upwards, allowing animals inside the enclosure to obtain food placed in the bins, but the distance between the bins prevented one animal from pulling the handle and obtaining rewards from both bins (see Appendix Figure S1). Only one handle was provided, and when the handle was released both bins returned to their original position and rewards were not accessible, ensuring that the Actor could obtain rewards from only the Actor bin (and not the Recipient bin).

Task comprehension: During pilot testing animals were willing to pull the handle and able to retrieve rewards from the Actor bin, so instead of using a formal training procedure we evaluated whether animals' behavior changed as a function of experience (i.e. trial number). These results indicate that Actors understood that the food in both the Actor bin and the Recipient bin was obtainable, and that they conditioned their behavior on the presence of this food (see Study 1 Results). To further assay animals' understanding of the task, at the conclusion of Study 1 we also conducted a Knowledge Probe where we explored whether Actors were more likely to pull the handle when the apparatus was modified to allow them to directly access food from the Recipient bin (see below).

Testing procedure: Testing consisted of multiple blocks of randomized trials, each with a different payoff distribution. A trial did not start unless (1) 30 seconds had passed from the start of the previous trial, (2) a potential Actor was near the testing area, (3) the apparatus was at rest in the No Access position (see Appendix Figure S1), and (4) any payoffs remaining from the previous trial had been removed. A trial lasted no less than 30 seconds, and ended if the animal did not pull within that period or pulled before all payoffs were placed. The experimenter waited until Actors looked in his direction as he held up the payoff and placed it in the appropriate bin (if the trial called for zero payoffs the experimenter only touched the bin). Individuals varied substantially in the number of observations they contributed to the dataset (ranging from 1 to 124 observations; median: 31; see Appendix Table S1 for details).

Knowledge Probe: At the conclusion of Study 1 we modified the apparatus with a longer handle to allow Actors to pull the handle and climb down to obtain food from the Recipient bin. We then isolated animals and presented them with trials using the 0/1, 1/1, and 0/0 payoff distributions using this modified apparatus. If Actors understood that pulling the handle caused food in the Recipient bin to become accessible, they should be more likely to pull the handle in the Knowledge Probe than in the Test trials for the 0/1 payoff distribution, but not for the 1/1 and 0/0 payoff distributions. We conducted the Knowledge Probe at the end of Study 1 to avoid training animals that they could directly obtain food for themselves by pulling the handle during 0/1 trials, which could appear as prosocial behavior if the Testing trials followed the Knowledge Probe. Four of the subject animals (two male, two female) were willing to be isolated for the Knowledge Probe, and we ended each session when the subject became distressed at being separated from their social group.

Analyses: The outcome variable was the binary parameter *Actor Pulled Handle* that captured whether the Actor pulled the handle (coded as "1") or did not pull (coded as "0"). We used multi-level logistic regression models to explore how various parameters predicted the outcome variable, and included Actor identity as a random effect to control for non-independence of observations. *Actor's Trial Number* codes for the total number of test trials that Actors had previously participated in (across all six payoff distributions). Contact the corresponding author for raw data.

Study 2

Participants: The composition of two groups from which participants were drawn had been very stable for several years, while the third group was newly formed (this group contributed only one participant).

Procedure: The procedure was identical to that of Study 1, but a familiarization procedure was employed because the apparatus was more complicated. Prior to testing, subjects received 40 counterbalanced trials from the Control condition (1/1 vs. 0/1 choice). Testing consisted of blocks of eights trials, each containing four pairs of trials that used the same basic payoff distributions but counterbalanced their side of presentation (e.g. one trial would load 1/0 on the left and 1/1 on the right, and the second would load 1/1 on the left and 1/0 on the right). Within each block, the order of these 8 trials was randomized.

At the start of a trial the experimenter first loaded the left Recipient bins followed by the Actor bins (see Appendix Figure S2). The Test Phase began when the experimenter inserted the handles into the enclosure, making them accessible to the Actor (Appendix Figure S2). There was again substantial variation in the number of observations that the Actors contributed to the dataset (ranging from 23 to 304 observations; median: 139; see Appendix Table S4 for details).

Analyses: The primary outcome variable was the binary parameter *Actor Chose Prosocial Outcome*. For the Reward, Reward (food balanced), and No Reward conditions, prosocial choices (1/1, 1/1, and 0/1) were coded as “1”, and non-prosocial choices (1/0, 1/0(1), and 0/0) were coded as “0”. For the Control condition, 1/1 was coded as “1” and choices of 0/1 as “0”. Our primary analyses exclude trials in which Actors selected neither of the payoff distributions, and we separately analyze Actors’ choices to “do nothing.” Multi-level logistic regressions controlled for the non-independent observations, and we investigated how *Actor’s Trial Number* predicted animals’ behavior in each of the four conditions. Contact the corresponding author for raw data.

Acknowledgements

We thank all of the staff at the Michale E. Keeling Center for Comparative Medicine & Research for their effort and support during this study. BH also thanks Lydia Hopper, Lara Wood, Sarah Price, Lisa O’Bryan, and Lisa Reamer for many suggestions and comments.

3 APPENDICES

TASK DESIGN INFLUENCES PROSOCIALITY IN CAPTIVE CHIMPANZEES (PAN TROGLODYTES)

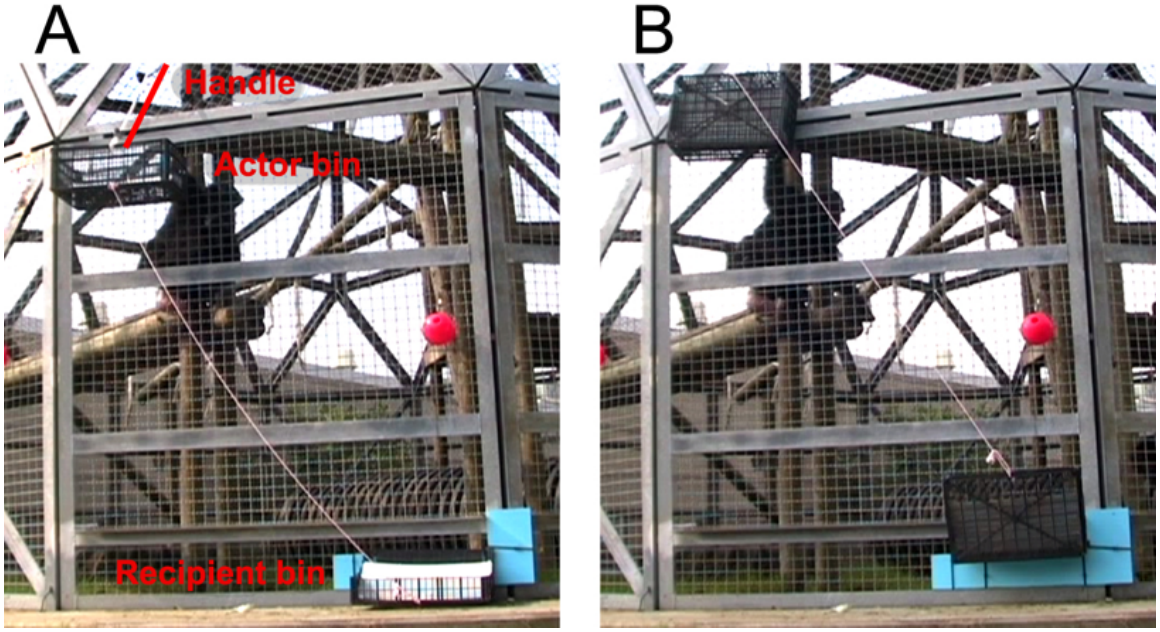
Appendix 1

Figure S1: Apparatus used in Study 1.

(1A): The apparatus in the No Access position, with bins at rest where animals could not obtain any food.

(1B): The apparatus in the Access position, with bins pivoted upward so that food rewards were in reach of animals inside the enclosure.

Pulling the handle moved the apparatus from No Access (1A) to Access (1B). While pulling the handle, the actor could acquire food placed in the Actor bin, but they could not reach food in the Recipient bin. If the actor released the handle both bins returned immediately to the No Access position and any food remaining in the Recipient bin rolled back out of reach.



Appendix 2: Further discussion of Knowledge Probe for Study 1

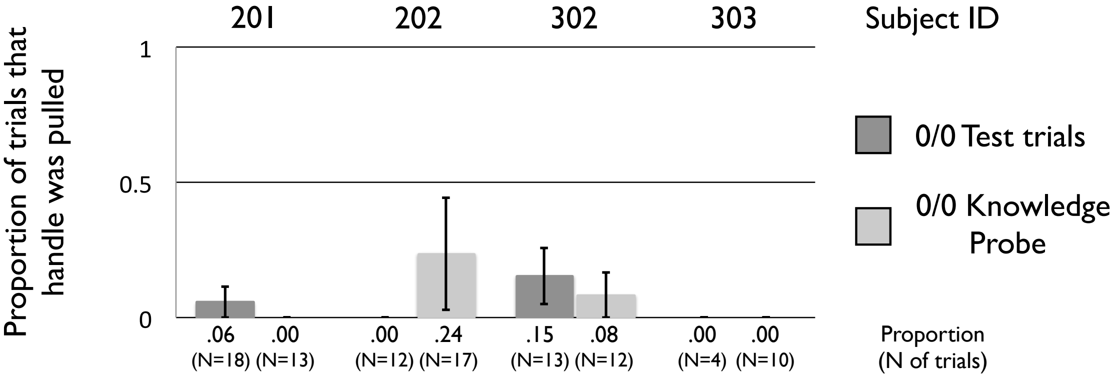
In the Knowledge Probe the apparatus was modified with a longer handle, so that Actor animals could both pull the handle and directly obtain food for themselves from the Recipient bin. Thus, though Actors could not directly obtain food from the Recipient bin in the Test trials, they could do so in the Knowledge Probe.

Actor animals were isolated from their social group so that they could be certain that no other group member would obtain food from the Recipient bin if they pulled the handle. Actor animals were only tested if they did not become distressed by being isolated, and the session was ended when they did become distressed. Four animals were willing to be isolated, and we collected some data from each of these animals.

Animals were presented with three different payoff distributions in the Knowledge Probe: 0/1, 1/1, and 0/0. If Actor animals understood that pulling the handle made food in the Recipient bin accessible, they should be more willing to pull the handle in the Knowledge Probe than the Test trials when food is placed only in the Recipient bin (i.e., the 0/1 payoff distribution). This is because pulling the handle allows them to obtain food for themselves in the Knowledge Probe, but not in the Test trials. However, for the 1/1 payoff distribution animals should be near-ceiling in their willingness to pull the handle both in the Test and Knowledge Probe, because in both sets of trials they obtain food for themselves. Similarly, in the 0/0 payoff distribution animals should be near-floor in their willingness to pull the handle in both the Test and Knowledge Probe, because in both sets of trials they cannot obtain food for themselves.

Results showed that these four subject animals were indeed more likely to pull the handle for the 0/1 payoff distribution in the Knowledge Probe than in the Test trials (see Figure 2 in main text). Additionally, animals pulled the handle for the 1/1 payoff distribution on every trial in both the Knowledge Probe and Test trials. Actor animals also pulled the handle at very low rates for the 0/0 payoff distribution in both the Knowledge Probe and Test trials, and at comparable rates across both sets of trials (see Figure S2).

Figure S2:



Overall, these results indicate that the four subject animals tested in our Knowledge Probe clearly understood how the apparatus worked. Actors switched from non-zero but relatively low rates of pulling the handle for the 0/1 payoff distribution in the Test trials, to very high rates of pulling the 0/1 payoff distribution in the Knowledge Probe (where Actors could reach the previously inaccessible food in the Recipient bin). These animals thus understood how pulling the handle causally made food in the recipient bin to become accessible.

In contrast, animals' behavior did not differ substantially across the Knowledge Probe and Test trials for the 1/1 and 0/0 payoff distributions. Actor animals pulled the handle on every trial for the 1/1 payoff distribution, and pulled at low rates for the 0/0 payoff distribution in both Knowledge Probe and Test trials. It is true that subject 202 did pull the handle more frequently on 0/0 Knowledge trials than 0/0 Test trials, but moving from the Test trials to the Knowledge

Probe has a much greater impact on subject 202’s rates of pulling the handle for the 0/1 payoff distribution (main text Figure 2) than the 0/0 payoff distribution (ESM Figure S2, above). Thus, the overall pattern is largely consistent: all four animals understood the nature of the task and comprehended that pulling the handle allowed food in the Recipient bin to become accessible.

Appendix 3

Figure S3: Apparatus used in Study 2.

(1a): left Actor bin

(2a): right Actor bin

(1b): left Recipient bin

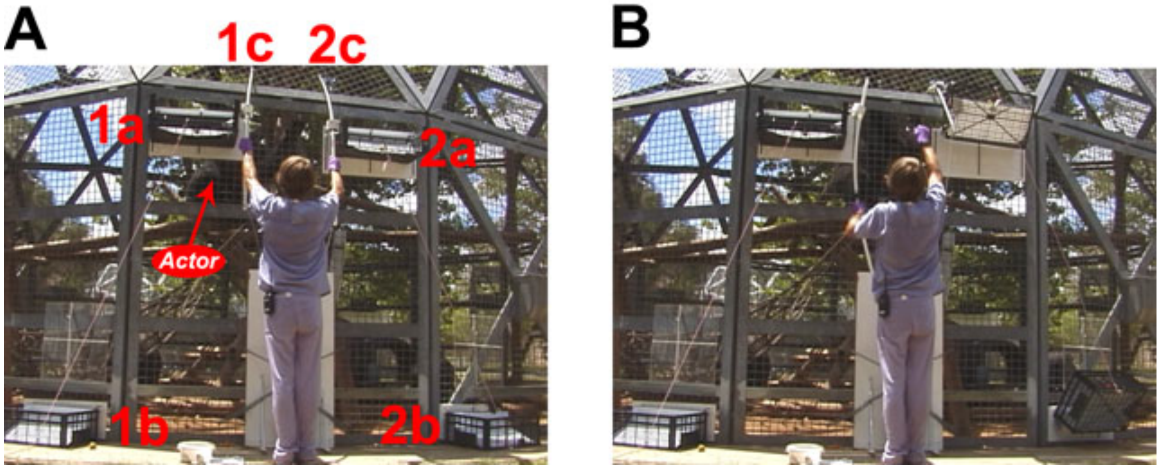
(2b): right Recipient bin

(1c): left handle

(2c): right handle;

(A): The apparatus in the No Access position, before a handle has been pulled by the actor.

(B): The apparatus in the Access position, after the right handle has been pulled by the actor, which retracts the left handle so that it becomes inaccessible (i.e., it is now impossible move the left Actor and Recipient bins into Access).



Appendix 4

Table S1: Total number of observations contributed by each subject in Study 1.

Subject	Total *n* observations for this actor
201	99
202	59
203	32
204	13
205	3
206	1
301	124
302	89
303	30
306	5

Appendix 5

Table S2: Robustness analysis. Table provides the coefficients for logistic regression models of how switching from 0/0 to 0/1 changes animas behavior, dropping out each of the animals in the sample one at a time. One of the animals contributed no data to this condition (206), but for each of the other animals excluding their data does not change the valence of the coefficient, nor does it dramatically change the magnitude of the coefficient (which remains larger than it's standard error). This means that even though there is substantial variation in the amount of data each animal contributes to the sample, the effect is not being entirely driven by any one subject.

Subject excluded from analysis	*n* observations for this actor	*n* observations remaining in sample	coefficient	standard error
none		150	1.24	.62
201	35	115	.87	.75
202	22	128	1.22	.62
203	10	140	1.22	.62
204	4	146	1.42	.69
205	1	149	1.26	.62
206	0			
301	43	107	1.03	.65
302	26	124	1.98	.83
303	7	143	1.12	.63
306	2	148	1.12	.63
201 & 302	61	89	2.08	1.24

Appendix 6

Table S3: Models of the effect of Actor's Trial Number on actors' willingness to pull the handle and operate the apparatus in Study 1, represented in Figure 1B.

DV: Actor Pulled Handle	Model 1	Model 2	Model 3	Model 4	Model 5
	0 / 0	0/0(1)	0 / 1	1 / 0	1 / 1
	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Actor's Trial Number	-.079 (.047)	-.092 (.042)	.002 (.01)	.002 (.01)	.02 (.03)
Random Effect	1.01 (1.24)	.001 (.61)	.51 (.69)	.001 (.02)	.003 (.11)
Constant	-.91	-.44	-1.89	2.80	3.64

Appendix 7

Table S4: Total number of observations contributed by each subject in Study 2.

Subject	Total *n* observations for this actor
401	235
402	58
403	133
405	29
501	244
601	304
602	145
603	23

Appendix 8

Table S5: Models of the effect of Actor's Trial Number on actors' choices of the prosocial option in Study 2, represented in Figure 3B from the main text.

DV: Chose Prosocial Outcome (For Control 1/1 coded as '1')	Model 6	Model 7	Model 8	Model 9
	Rewards	Rewards (food balanced)	No Rewards	Control
	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Actor's Trial Number	< -.001 (.001)	.001 (.001)	< -.001 (.002)	.015 (.01)
Constant	.19	-.06	.31	3.16
Random Effect	<.001 (.02)	.001 (.02)	.004 (.09)	.004 (.13)

Appendix 9

Table S6: Models of the effect of Actor's Trial Number on actors' tendency to do nothing in the No Reward condition of Study 2, represented in Figure 3B in main text.

DV: Actor did nothing	Model 10
	No Rewards
	Coef. (SE)
Actor's Trial Number	.014 (.002)
Constant	-2.48
Random Effect	1.16 (.38)

Appendix 10

Table S7: Multi-level logistic regression models replicating the results displayed in Figure 1A, also controlling for non-independence in the data. Animals pulled the handle on every 1/3 trial.

DV: Actor Pulled Handle	Coef. (SE)
0 / 0 (1)	.34 (.67)
0 / 1	1.22 (.61)
1 / 0	5.96 (.77)
1 / 1	7.42 (1.16)
1 / 3	24.1 (n/a)
Random effect	.30 (.41)
Constant	-2.06

BIBLIOGRAPHY

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in: Petrov, BN and Csaki, F (Eds.), Second International Symposium of Information Theory. Akademiai Kiado, Budapest, Hungary, pp. 267–281.

Axelrod, R., Hamilton, W., 1981. The evolution of cooperation. Science 211, 1390–1396.

Barta, Z., McNamara, J.M., Huszár, D.B., Taborsky, M., 2011. Cooperation among non-relatives evolves by state-dependent generalized reciprocity. Proc. R. Soc. B Biol. Sci. 278, 843 – 848.

Benenson, J.F., Pascoe, J., Radmore, N., 2007. Children's altruistic behavior in the dictator game. Evol. Hum. Behav. 28, 168–175.

Birch, L.L., Billman, J., 1986. Preschool Children's Food Sharing with Friends and Acquaintances. Child Dev. 57, 387.

Blake, P.R., McAuliffe, K., 2011. "I had so much it didn't seem fair": Eight-year-olds reject two forms of inequity. Cognition 120, 215–224.

Blake, P.R., Rand, D.G., 2010. Currency value moderates equity preference among young children. Evol. Hum. Behav. 31, 210–218.

Bliege Bird, R., Bird, D.W., Smith, E.A., Kushnick, G.C., 2002. Risk and reciprocity in Meriam food sharing. Evol. Hum. Behav. 23, 297–321.

Boesch, C., 2007. What makes us human (Homo sapiens)? The challenge of cognitive cross-species comparison. J. Comp. Psychol. 121, 227–240.

Boesch, C., 2008. Taking development and ecology seriously when comparing cognition: Reply to Tomasello and Call (2008). J. Comp. Psychol. 122, 453–455.

Boesch, C., Boesch, H., 1989. Hunting behavior of wild chimpanzees in the Taï National Park. Am. J. Phys. Anthropol. 78, 547–573.

Boesch, C., Boesch, H., Vigilant, L., 2006. Cooperative hunting in chimpanzees: kinship or mutualism? Van Schaik Cp Kappeler Pm Eds Coop. Primates Humans 139–150.

Boysen, S.T., Berntson, G.G., Hannan, M.B., Cacioppo, J.T., 1996. Quantity-based interference and symbolic representations in chimpanzees (Pan troglodytes). J. Exp. Psychol.-Anim. Behav. Process. 22, 76–86.

Boysen, S.T., Berntson, G.G., Mukobi, K.L., 2001. Size matters: Impact of item size and quantity on array choice by chimpanzees (Pan troglodytes). J. Comp. Psychol. 115, 106–110.

Bräuer, J., Call, J., Tomasello, M., 2009. Are apes inequity averse? New data on the token-exchange paradigm. Am. J. Primatol. 71, 175–181.

Bräuer, J., Hanus, D., 2012. Fairness in Non-human Primates? Soc. Justice Res. 25, 256–276.

Brosnan, S., Silk, J., Henrich, J., Mareno, M., Lambeth, S., Schapiro, S., 2009. Chimpanzees (Pan troglodytes) do not develop contingent reciprocity in an experimental task. Anim. Cogn. 12, 587–597.

Brosnan, S.F., 2011. A Hypothesis of the Co-evolution of Cooperation and Responses to Inequity. Front. Neurosci. 5.

Brosnan, S.F., Talbot, C., Ahlgren, M., Lambeth, S.P., Schapiro, S.J., 2010. Mechanisms underlying responses to inequitable outcomes in chimpanzees, Pan troglodytes. Anim. Behav. 79, 1229–1237.

Brownell, C.A., Svetlova, M., Nichols, S., 2009. To Share or Not to Share: When Do Toddlers Respond to Another's Needs? Infancy 14, 117–130.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer Verlag, New York, NY.

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., Tomasello, M., 2011. Early social cognition in three cultural contexts. Monogr. Soc. Res. Child Dev. 76, 1–142.

Camerer, C., 2003. Behavioral game theory: Experiments in strategic interaction. Princeton University Press Princeton, NJ.

Chudek, M., Henrich, J., 2011. Culture–gene coevolution, norm-psychology and the emergence of human prosociality. Trends Cogn. Sci. 15, 218–226.

Cronin, K.A., 2012. Prosocial behaviour in animals: the influence of social relationships, communication and rewards. Anim. Behav. 84, 1085–1093.

Dahlman, S., Ljungqvist, P., Johannesson, M., 2007. Reciprocity in young children. Stockholm School of Economics.

De Waal, F.B.M., 1997. The chimpanzee’s service economy: Food for grooming. Evol. Hum. Behav. 18, 375–386.

Dreman, S.B., Greenbaum, C.W., 1973. Altruism or Reciprocity: Sharing Behavior in Israeli Kindergarten Children. Child Dev. 44, 61–68.

Dunfield, K., Kuhlmeier, V.A., O’Connell, L., Kelley, E., 2011. Examining the Diversity of Prosocial Behavior: Helping, Sharing, and Comforting in Infancy. Infancy 16, 227–247.

Eisenberg, N., Fabes, R.A., Spinrad, T.L., 2006. Prosocial Development, in: Handbook of Child Psychology, Social, Emotional, and Personality Development. John Wiley & Sons, pp. 646–718.

Fabes, R.A., Eisenberg, N., 1998. Meta-Analysis of age and sex differences in children’s and adolescents’ prosocial behavior (Working Paper). Arizona State University.

Fehr, E., Bernhard, H., Rockenbach, B., 2008. Egalitarianism in young children. Nature 454, 1079–1083.

Fishbein, H.D., Kaminski, N.K., 1985. Children’s reciprocal altruism in a competitive game. Br. J. Dev. Psychol. 3, 393–398.

Fletcher, G.E., 2008. Attending to the outcome of others: disadvantageous inequity aversion in male capuchin monkeys (Cebus apella). Am. J. Primatol. 70, 901–905.

Fujisawa, K.K., Kutsukake, N., Hasegawa, T., 2008. Reciprocity of prosocial behavior in Japanese preschool children. Int. J. Behav. Dev. 32, 89 –97.

Gilby, I.C., 2006. Meat sharing among the Gombe chimpanzees: harassment and reciprocal exchange. Anim. Behav. 71, 953–963.

Greenberg, J.R., Hamann, K., Warneken, F., Tomasello, M., 2010. Chimpanzee helping in collaborative and noncollaborative contexts. Anim. Behav. 80, 873–880.

Gummerum, M., Hanoch, Y., Keller, M., 2008. When Child Development Meets Economic Game Theory: An Interdisciplinary Approach to Investigating Social Development. Hum. Dev. 51, 235–261.

Gummerum, M., Hanoch, Y., Keller, M., Parsons, K., Hummel, A., 2010. Preschoolers’ allocations in the dictator game: The role of moral emotions. J. Econ. Psychol. 31, 25–34.

Gurven, M., 2004. To Give and to Give Not: The Behavioral Ecology of Human Food Transfers. Behav. Brain Sci. 27, 543–559.

Gurven, M., 2006. The evolution of contingent cooperation. Curr. Anthr. 47, 185–192.

Gurven, M., Hill, K., Kaplan, H., 2002. From Forest to Reservation: Transitions in Food-Sharing Behavior among the Ache of Paraguay. J. Anthr. Res. 58, 93–120.

Gurven, M., Hill, K., Kaplan, H., Hurtado, A., Lyles, R., 2000. Food transfers among Hiwi foragers of Venezuela: tests of reciprocity. Hum. Ecol. 28, 171–218.

Hamilton, W.D., 1964. The genetical evolution of social behaviour. I. J. Theor. Biol. 7, 1–16.

Harbaugh, W.T., Krause, K., 2000. Children's altruism in public good and dictator experiments. Econ. Inq. 38, 95–109.

Harbaugh, W.T., Krause, K., Liday, S.G., 2003. Bargaining by Children (Working Paper). University of Oregon, Department of Economics, Eugene, Oregon.

Hay, D.F., Castle, J., Davies, L., Demetriou, H., Stimson, C.A., 1999. Prosocial Action in Very Early Childhood. J. Child Psychol. Psychiatry 40, 905–916.

Henrich, J., 2004. Animal behaviour (communication arising): Inequity aversion in capuchins? Nature 428, 139.

Henrich, J., 2006. SOCIAL SCIENCE: Enhanced: Cooperation, Punishment, and the Evolution of Human Institutions. Science 312, 60–61.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N.S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F.W., Patton, J.Q., Tracer, D., 2005. In Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies. Behav. Brain Sci. 28, 795–815.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., Ziker, J., 2010a. Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. Science 327, 1480–1484.

Henrich, J., Heine, S.J., Norenzayan, A., 2010b. The weirdest people in the world? Behav. Brain Sci. 33, 61–83.

Henrich, J., Henrich, N., 2007. Why Humans Cooperate: A Cultural and Evolutionary Explanation: A Cultural and Evolutionary Explanation. Oxford University Press, USA, New York.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Henric, N., Lesorogol, C., Marlowe, F., Tracer, D., Ziker, J., 2006. Costly Punishment Across Human Societies. Science 312, 1767–1770.

Herrmann, B., Thoni, C., Gachter, S., 2008. Antisocial Punishment Across Societies. Science 319, 1362–1367.

Heyes, C., 2012. Simple minds: a qualified defence of associative learning. Philos. Trans. R. Soc. B Biol. Sci. 367, 2695–2703.

Horner, V., Carter, J.D., Suchak, M., Waal, F.B.M. de, 2011. Spontaneous prosocial choice by chimpanzees. Proc. Natl. Acad. Sci. 108, 13847–13851.

House, B., Henrich, J., Sarnecka, B., Silk, J.B., 2013. The development of contingent reciprocity in children. Evol. Hum. Behav. 34, 86–93.

House, B.R., Henrich, J., Brosnan, S.F., Silk, J.B., 2012. The ontogeny of human prosociality: behavioral experiments with children aged 3 to 8. Evol. Hum. Behav. 33, 291–308.

Jaeggi, A.V., De Groot, E., Stevens, J.M.G., Van Schaik, C.P., 2013. Mechanisms of reciprocity in primates: testing for short-term contingency of grooming and food sharing in bonobos and chimpanzees. Evol. Hum. Behav. 34, 69–77.

Jensen, K., Call, J., Tomasello, M., 2007. Chimpanzees Are Rational Maximizers in an Ultimatum Game. Science 318, 107–109.

Jensen, K., Hare, B., Call, J., Tomasello, M., 2006. What's in it for me? Self-regard precludes altruism and spite in chimpanzees. Proc. R. Soc. B Biol. Sci. 273, 1013–1021.

Lancy, D.F., Grove, M.A., 2011. Getting Noticed. Hum. Nat. 22, 281–302.

Langergraber, K.E., Mitani, J.C., Vigilant, L., 2007. The limited impact of kinship on cooperation in wild chimpanzees. Proc. Natl. Acad. Sci. 104, 7786 –7790.

Leman, P.J., Keller, M., Takezawa, M., Gummerum, M., 2009. Children's and Adolescents' Decisions about Sharing Money with Others. Soc. Dev. 18, 711–727.

Levitt, M.J., Weber, R.A., Clark, M.C., McDonnell, P., 1985. Reciprocity of exchange in toddler sharing behavior. Dev. Psychol. 21, 122–123.

Lucas, M.M., Wagner, L., Chow, C., 2008. Fair Game: The Intuitive Economics of Resource Exchange in Four-Year Olds. J. Soc. Evol. Cult. Psychol. 2.

Lunn, D., Jackson, C., Best, N., Spiegelhalter, D.J., Thomas, A., 2012. The BUGS Book: A Practical Introduction to Bayesian Analysis. Chapman & Hall.

Madsen, M.C., Lancy, D.F., 1981. Cooperative and Competitive Behavior Experiments Related to Ethnic Identity and Urbanization in Papua New Guinea. J. Cross-Cult. Psychol. 12, 389–408.

Marlowe, F.W., Berbesque, J.C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J.C., Ensminger, J., Gurven, M., Gwako, E., Henrich, J., Henrich, N., Lesorogol, C., McElreath, R., Tracer, D., 2008. More "altruistic" punishment in larger societies. Proc. R. Soc. B Biol. Sci. 275, 587–592.

McElreath, R., 2013. glmer2stan: Rstan models defined by glmer formulas. University of California, Davis.

McElreath, R., Bell, A.V., Efferson, C., Lubell, M., Richerson, P.J., Waring, T., 2008. Beyond existence and aiming outside the laboratory: estimating frequency-dependent and pay-off-biased social learning strategies. Philos. Trans. R. Soc. B Biol. Sci. 363, 3515–3528.

Melis, A.P., Hare, B., Tomasello, M., 2006. Chimpanzees Recruit the Best Collaborators. Science 311, 1297–1300.

Melis, A.P., Hare, B., Tomasello, M., 2008. Do chimpanzees reciprocate received favours? Anim. Behav. 76, 951–962.

Melis, A.P., Semmann, D., 2010. How is human cooperation different? Philos. Trans. R. Soc. B Biol. Sci. 365, 2663–2674.

Melis, A.P., Warneken, F., Jensen, K., Schneider, A.C., Call, J., Tomasello, M., 2011. Chimpanzees help conspecifics obtain food and non-food items. Proc. R. Soc. B Biol. Sci. 278, 1405.

Mitani, J.C., 2006. Reciprocal exchange in chimpanzees and other primates, in: Van Schaik, C.P and Kappeler P.M. (Eds.), Cooperation in Primates and Humans. Springer Verlag, Heidelberg, pp. 107–119.

Mitani, J.C., 2009. Male chimpanzees form enduring and equitable social bonds. Anim. Behav. 77, 633–640.

Mitani, J.C., Watts, D.P., 2001. Why do chimpanzees hunt and share meat? Anim. Behav. 61, 915–924.

Moore, C., 2009. Fairness in Children's Resource Allocation Depends on the Recipient. Psychol. Sci. 20, 944–948.

Nishida, T., Turner, L.A., 1996. Food transfer between mother and infant chimpanzees of the Mahale Mountains National Park, Tanzania. Int. J. Primatol. 17, 947–968.

Normann, H.T., Wallace, B., 2004. The Impact of the Termination Rule in Cooperation Experiments. Royal Holloway, University of London: Discussion Papers in Economics.

Price, S.A., Brosnan, S.F., 2012. To Each According to his Need? Variability in the Responses to Inequity in Non-Human Primates. Soc. Justice Res. 25, 140–169.

Proctor, D., Williamson, R.A., Waal, F.B.M. de, Brosnan, S.F., 2013. Chimpanzees play the ultimatum game. Proc. Natl. Acad. Sci. 110, 2070–2075.

R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria.

Richerson, P.J., Boyd, R., 2005. Not by genes alone: how culture transformed human evolution. University of Chicago Press.

Rochat, P., Dias, M.D.G., Liping, G., Broesch, T., Passos-Ferreira, C., Winning, A., Berg, B., 2009. Fairness in Distributive Justice by 3- and 5-Year-Olds Across Seven Cultures. J. Cross-Cult. Psychol. 40, 416–442.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. Am. Econ. Rev. 81, 1068–1095.

Sally, D., Hill, E., 2006. The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. J. Econ. Psychol. 27, 73–97.

Schino, G., Aureli, F., 2009. Reciprocal Altruism in Primates: Partner Choice, Cognition, and Emotions, in: Advances in the Study of Behavior. Academic Press, pp. 45–69.

Selten, R., Stoecker, R., 1986. End behavior in sequences of finite Prisoner's Dilemma supergames: A learning theory approach. J. Econ. Behav. Organ. 7, 47–70.

Siegler, R.S., 2004. U-Shaped Interest in U-Shaped Development-and What It Means. J. Cogn. Dev. 5, 1–10.

Silk, J.B., 2003. Cooperation without counting, in: Hammerstein, P. (Ed.), Genetic and Cultural Evolution of Cooperation. The MIT Press, Cambridge, MA, pp. 37–54.

Silk, J.B., Brosnan, S.F., Vonk, J., Henrich, J., Povinelli, D.J., Richardson, A.S., Lambeth, S.P., Mascaro, J., Schapiro, S.J., 2005. Chimpanzees are indifferent to the welfare of unrelated group members. Nature 437, 1357–1359.

Silk, J.B., House, B.R., 2011. Colloquium Paper: Evolutionary foundations of human prosocial sentiments. Proc. Natl. Acad. Sci. 108, 10910–10917.

Silk, J.B., House, B.R., 2012. The Phylogeny and Ontogeny of Prosocial Behavior. Vonk J Shackelford T Eds Oxf. Handb. Comp. Evol. Psychol.

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, e59510.

Sommerville, J.A., Schmidt, M.F.H., Yun, J., Burns, M., 2013. The Development of Fairness Expectations and Prosocial Behavior in the Second Year of Life. Infancy 18, 40–66.

Stan Development Team, 2013. Stan: A C++ Library for Probability and Sampling.

Staub, E., Sherk, L., 1970. Need for approval, children's sharing behavior, and reciprocity in sharing. Child Dev. 41, 243–252.

Sutter, M., Kocher, M.G., 2007. Trust and trustworthiness across different age groups. Games Econ. Behav. 59, 364–382.

Svetlova, M., Nichols, S.R., Brownell, C.A., 2010. Toddlers' Prosocial Behavior: From Instrumental to Empathic to Altruistic Helping. Child Dev. 81, 1814–1827.

Takezawa, M., Gummerum, M., Keller, M., 2006. A stage for the rational tail of the emotional dog: Roles of moral reasoning in group decision making. J. Econ. Psychol. 27, 117–139.

Thompson, C., Barresi, J., Moore, C., 1997. The development of future-oriented prudence and altruism in preschoolers. Cogn. Dev. 12, 199–212.

Thompson, J.L., Nelson, A.J., 2011. Middle Childhood and Modern Human Origins. Hum. Nat. 22, 249–280.

Tomasello, M., Call, J., 2008. Assessing the Validity of Ape-Human Comparisons: A Reply to Boesch (2007). J. Comp. Psychol. 122, 449–452.

Trivers, R.L., 1971. The Evolution of Reciprocal Altruism. Q. Rev. Biol. 46, 35.

Vaish, A., Carpenter, M., Tomasello, M., 2009. Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. Dev. Psychol. 45, 534–543.

Vonk, J., Brosnan, S.F., Silk, J.B., Henrich, J., Richardson, A.S., Lambeth, S.P., Schapiro, S.J., Povinelli, D.J., 2008. Chimpanzees do not take advantage of very low cost opportunities to deliver food to unrelated group members. Anim. Behav. 75, 1757–1770.

Warneken, F., 2013. Young children proactively remedy unnoticed accidents. Cognition 126, 101–108.

Warneken, F., Hare, B., Melis, A.P., Hanus, D., Tomasello, M., 2007. Spontaneous Altruism by Chimpanzees and Young Children. Plos Biol. 5, e184.

Warneken, F., Tomasello, M., 2009. Varieties of altruism in children and chimpanzees. Trends Cogn. Sci. 13, 397–402.

Watts, D.P., 1998. Coalitionary mate guarding by male chimpanzees at Ngogo, Kibale National Park, Uganda. Behav. Ecol. Sociobiol. 44, 43–55.

Watts, D.P., Mitani, J.C., 2001. Boundary Patrols and Intergroup Encounters in Wild Chimpanzees. Behaviour 138, 299–327.

Yamamoto, S., Humle, T., Tanaka, M., 2009. Chimpanzees Help Each Other upon Request. Plos One 4, e7416.

Yamamoto, S., Humle, T., Tanaka, M., 2012. Chimpanzees' flexible targeted helping based on an understanding of conspecifics' goals. Proc. Natl. Acad. Sci. 109, 3588–3592.

Yamamoto, S., Takimoto, A., 2012. Empathy and fairness: Psychological mechanisms for eliciting and maintaining prosociality and cooperation in primates. Soc. Justice Res. 25, 233–255.

Yamamoto, S., Tanaka, M., 2009a. Do Chimpanzees (Pan troglodytes) Spontaneously Take Turns in a Reciprocal Cooperation Task? J. Comp. Psychol. 123, 242–249.

Yamamoto, S., Tanaka, M., 2009b. How did altruism and reciprocity evolve in humans? Perspectives from experiments on chimpanzees (Pan troglodytes). Interact. Stud. 10, 150–182.

Yamamoto, S., Tanaka, M., 2010. The influence of kin relationship and reciprocal context on chimpanzees' other-regarding preferences. Anim. Behav. 79, 595–602.