Development of in-group favoritism in children’s third-party punishment of selfishness

Jillian J. Jordan*a,b,†, Katherine McAuliffea,b,c, and Felix Warnekena

*Department of Psychology and †Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138; and ‡Department of Psychology, Yale University, New Haven, CT 06511

Edited* by Susan T. Fiske, Princeton University, Princeton, NJ, and approved July 23, 2014 (received for review February 11, 2014)

When enforcing norms for cooperative behavior, human adults sometimes exhibit in-group bias. For example, third-party observers punish selfish behaviors committed by out-group members more harshly than similar behaviors committed by in-group members. Although evidence suggests that children begin to systematically punish selfish behavior around the age of 6 y, the development of in-group bias in their punishment remains unknown. Do children start off enforcing fairness norms impartially, or is norm enforcement biased from its emergence? How does bias change over development? Here, we created novel social groups in the laboratory and gave 6- and 8-year-olds the opportunity to engage in costly third-party punishment of selfish sharing behavior. We found that by age 6, punishment was already biased: Selfish resource allocations received more punishment when they were proposed by out-group members and when they disadvantaged in-group members. We also found that although costly punishment increased between ages 6 and 8, bias in punishment partially decreased. Although 8-y-olds also punished selfish out-group members more harshly, they were equally likely to punish on behalf of disadvantaged in-group and out-group members, perhaps reflecting efforts to enforce norms impartially. Taken together, our results suggest that norm enforcement is biased from its emergence, but that this bias can be partially overcome through developmental change.

cooperation | ontogeny | equality

Social norms—standards of behavior enforced through informal rewards and sanctions—are thought to play a key role in human cooperation (1–4). The enforcement of fairness norms can serve to maintain cooperative interactions by disincentivizing selfish behavior (2). Experiments show that third-party observers are often willing to incur personal costs (e.g., spend money) to enforce these norms by punishing one individual for behaving selfishly toward another (4–6). Thus, human adults are willing to make personal sacrifices to enforce fairness norms, even when they have not been directly affected by the norm violation.

A striking feature of norm-enforcement behavior is that it can be influenced by social group membership. Although moral codes often value impartiality, evidence suggests that punishment sometimes favors the in-group. For example, third parties are more likely to punish unfair behavior when it disadvantages an in-group member than an out-group member (7, 8). Additionally, third parties are more likely to punish out-group members than in-group members for unfair behavior that disadvantages an in-group member (7, 9). More generally, adults have shown in-group bias in their norm-enforcement behavior in a variety of contexts, and can be influenced by both the group membership of the selfish actor and the disadvantaged recipient (11–14). Such in-group bias reflects the role our intergroup psychology plays in shaping our standards of fairness and morality, perhaps because social norms are defined within groups, and in-group bias, cooperation, and norm enforcement may be mutually reinforcing processes (15–18). Therefore, social group identity provides critical context for understanding norm-enforcement behavior.

Whereas research with adults highlights the important relationship between costly norm enforcement and in-group bias, the developmental origins of this relationship are not known. However, understanding the developmental trajectory of this relationship can provide important insight into our underlying norm psychology, and recent advances in developmental psychology have provided us with tools to investigate these questions in young children (19). Concerning the development of fairness, experiments show that infants expect individuals to share resources equally (20), and that young children have an increasing tendency to create equal shares with others (21, 22). By at least 6 to 7 y of age, children’s fairness preferences are so strong that they are even willing to sacrifice resources to create equality (17, 23, 24). Concerning norm enforcement, evidence suggests that from a young age, antisocial behavior that is directed at a deserving target (i.e., punishment) can be evaluated positively: infants prefer puppets that hinder rather than help targets who have previously hindered others (25). Furthermore, experiments show that young children intervene against norm violations by spontaneously protesting when somebody violates a conventional or moral norm (26, 27) and by selectively punishing those who harm others (25). Finally, by 6 y of age, children begin to pay personal costs to engage in systematic third-party punishment of peers who allocate resources selfishly.

In-group bias also has early developmental roots (28). Infants prefer to look at and interact with individuals who share their race, language, and preferences (29–31). Furthermore, although infants prefer puppets that help targets with similar food preferences

Significance

Humans are unique among animals in their willingness to cooperate with friends and strangers. Costly punishment of unfair behavior is thought to play a key role in promoting cooperation by deterring selfishness. Importantly, adults sometimes show in-group favoritism in their punishment. To our knowledge, our study is the first to document this bias in children. Furthermore, our results suggest that from its emergence in development, children’s costly punishment shows in-group favoritism, highlighting that group membership provides critical context for understanding the enforcement of fairness norms. However, 8-y-old children show attenuated bias relative to 6-y-olds, perhaps reflecting a motivation for impartially. Our findings thus demonstrate that in-group favoritism has an important influence on human fairness and morality, but can be partially overcome with age.

Author contributions: J.J.J., K.M., and F.W. designed research; J.J.J. performed research; J.J.J. analyzed data; and J.J.J., K.M., and F.W. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.
†To whom correspondence should be addressed. Email: jillian.jordan@yale.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1402280111/-/DCSupplemental.

to themselves, they prefer puppets that hinder dissimilar targets (32). By early childhood, children favor same-race peers (28, 33) and share preferentially with in-group members (34, 35). Strikingly, children’s in-group bias also extends to “minimal groups” (36), or arbitrary social groups created in the laboratory. Whereas evidence suggests that children more readily show in-group bias in the context of natural groups (i.e., race and sex) than minimal groups (37), children also show minimal in-group bias in some contexts (38, 39). These effects demonstrate that the concept of a group, even without previous experience or associations with the group, can be sufficient to create bias (36).

Thus, over development children gain a sense of fairness, a willingness to enforce fairness norms, and a tendency toward in-group bias. However, it is unknown how these processes interact to produce the observed in-group bias in adult’s third-party punishment of selfishness. Do young children start off enforcing norms impartially, only later becoming biased by their group identity, or is norm enforcement biased by group identity from its earliest emergence? How does in-group bias change over development? One possibility is that children start off holding standards of fairness that they enforce impartially on individuals across groups, only later becoming biased or “corrupted” by their developing in-group identity. This hypothesis leads to the prediction that children’s in-group bias in norm enforcement may increase with age. Alternatively, children may display in-group bias in their norm enforcement from early in development. This hypothesis leads to the prediction that in-group bias may remain stable or decrease with age. The question of whether punishment starts off as impartial or biased has important implications for questions about children’s “default” sense of fairness and morality (40).

Evidence from studies of first-party resource allocations supports the hypothesis that children start off behaving impartially and become more biased with age. In one study of 3- through 8-y-old children, children became more biased over development in a task measuring their willingness to share resources (17). Children did not begin favoring their own school group until they were 7 to 8 y old, suggesting that costly sharing became more biased over development. Furthermore, in another study of 6- and 8-y-olds, children were more likely to preferentially allocate resources to minimal in-group members and negative resources to minimal out-group members, suggesting that in-group bias increased with age (41).

One potential explanation for this effect is that as children get older, their identities within social groups develop (42) and they experience or associations with the group, can be sufficient to create bias (36).

However, other evidence suggests that in-group bias may emerge early and remain stable or decrease over development. The finding that infants prefer puppets that hinder dissimilar others but help similar others suggests that from a very young age, the evaluation of antisocial actors may be biased by whom they harm (32). Furthermore, studies using implicit association tasks (which measure implicit, automatic forms of bias) have found that implicit in-group bias develops early and stays constant through adulthood (28, 33, 49). A large body of research on explicit bias suggests that children increasingly come to inhibit this implicit bias over development. Around age 7, explicit in-group bias begins to decline on a variety of measures (28, 50, 51), likely reflecting that children gain exposure to cultural norms against certain forms of bias and discrimination (49, 52) and improve their ability to monitor their self-presentation to conform to such norms (53). As a result, older children are more likely to show reduced bias when antidiscrimination norms are salient, and when they are motivated to look good in the eyes of observers (35, 49, 54).

Thus, over development children face the challenge of integrating their developing group-based values that may promote bias, and morality-based values that may promote impartiality (42). However, it is unknown how these developing values influence the enforcement of fairness norms. Here, we address this question by investigating the development of in-group bias in children’s costly third-party punishment of selfishness.

A recent study from our group investigated the developmental origins of third-party punishment in children using a resource-sharing paradigm. In this study, 6-y-olds paid costs to punish others but help similar others suggests that from a very young age, the evaluation of antisocial actors may be biased by whom they harm (32). Furthermore, studies using implicit association tasks (which measure implicit, automatic forms of bias) have found that implicit in-group bias develops early and stays constant through adulthood (28, 33, 49). A large body of research on explicit bias suggests that children increasingly come to inhibit this implicit bias over development. Around age 7, explicit in-group bias begins to decline on a variety of measures (28, 50, 51), likely reflecting that children gain exposure to cultural norms against certain forms of bias and discrimination (49, 52) and improve their ability to monitor their self-presentation to conform to such norms (53). As a result, older children are more likely to show reduced bias when antidiscrimination norms are salient, and when they are motivated to look good in the eyes of observers (35, 49, 54).

Thus, over development children face the challenge of integrating their developing group-based values that may promote bias, and morality-based values that may promote impartiality (42). However, it is unknown how these developing values influence the enforcement of fairness norms. Here, we address this question by investigating the development of in-group bias in children’s costly third-party punishment of selfishness.

We used the minimal group paradigm (36) by assigning subjects to a “blue” or “yellow” team. Because minimal group effects reflect the concept of groups, rather than effects of familiarity or previous experiences with specific groups, minimal group effects tend to be weaker than natural group effects. Thus, this method provides a conservative test: observed minimal group effects are likely to generalize to natural groups. After assigning subjects to groups, we confirmed with a manipulation check that subjects showed preferences for their in-group (adapted from ref. 39). This result demonstrated that our minimal group manipulation successfully induced in-group preferences and had the potential to influence punishment behavior (SI Text, Discussion of Costly Punishment Method and Fig. S1C).

Next, we measured punishment. In a series of trials, we demonstrated to subjects how an actor wanted to allocate candy between him- or herself and a recipient, using an experimental apparatus (Fig. S1 A and B). In each of these trials, the subject was a third-party “judge” who could use the apparatus handle to choose between accepting (enacting) the actor’s proposed allocation and rejecting (punishing) the allocation so that the candy would be thrown away. Subjects received their own endowment of candy and choosing to reject was costly: in each trial, subjects had to sacrifice one piece of candy if they chose to punish the actor (SI Text, Discussion of Costly Punishment Method and Fig. S1C).

In the majority of trials (80%), subjects were shown an allocation in which the actor selfishly shared no candy with the recipient (6 for actor, 0 for recipient). However, we varied allocations to hold subjects’ attention: in the remaining trials (20%), subjects were shown an allocation in which the actor was fair and shared equally (3 for actor, 3 for recipient) (Fig. S1D). To measure the effect of group membership, we manipulated within subjects whether the actor and recipient were in the subject’s group, resulting in four conditions (actor in, recipient in;
Results

Our primary analysis compared punishment of selfish trials across conditions and by age group. We used logistic regressions to predict the binary decision to punish an allocation (1 = punish/reject, 0 = accept/enact), and clustered SEs on subject to account for the nonindependence of decisions from the same subject. For full regression tables, see Tables S1–S4.

We found that 90.6% of subjects expressed belief that the actors and recipients were real on at least one question, and 53.1% expressed belief on both questions. We included all subjects in our analyses, but results were robust to excluding incredulous subjects. For more details, see SI Text, Analysis of Incredulous Subjects and Table S5.

We first confirmed that subjects enforced fairness norms by punishing selfishness more than fairness. We found that subjects paid a cost to punish 36.1% of selfish allocations, but only 5.5% of fair allocations. Fig. 1 shows the probability of punishing by age and allocation type. A regression predicting punishment as a function of allocation type (1 = selfish, 0 = fair), controlling for age and sex, showed a significant positive effect of selfishness (β = 2.31, P < 0.001, odds ratio = 10.04) (Table S1, column 1). There was also a significant positive interaction between selfishness and age (β = 2.04, P = 0.022, odds ratio = 7.72) (Table S1, column 2), indicating that 8-y-olds were more sensitive to selfishness than 6-y-olds. However, the effect of selfishness was significant both within 6-y-olds (β = 1.44, P = 0.001, odds ratio = 4.21) (Table S2, column 1) and 8-y-olds (β = 3.50, P < 0.001, odds ratio = 33.09) (Table S2, column 2). Thus, costly norm enforcement was present by age 6, and increased with age.

We next investigated the development of in-group bias in punishment of selfishness. In these analyses, we specifically focused on selfish trials, as fair trials were only included to hold attention. Preliminary analyses revealed no significant interactions between sex and group membership on punishment, or between actor and recipient group membership on punishment (SI Text, Analysis of Sex Interactions and Analysis of Interactions Between Actor and Recipient Group Membership, and Tables S6–S9). Thus, in our analyses, we collapsed across sex, and separately evaluated the effects of actor and recipient groups.

We began by asking if actor group influenced costly punishment of selfishness. Fig. 2A shows the probability of punishing by age and actor group. A regression predicting punishment as a function of actor group (1 = in-group, 0 = out-group), controlling for recipient group, age, and sex, showed a significant negative effect of actor group (β = −0.280, P = 0.013, odds ratio = 0.76) (Table S3, column 1), indicating that subjects were more likely to punish selfish out-group actors than selfish in-group actors. We also found no significant interaction between age and actor group (β = 0.053, P = 0.822, odds ratio = 1.05) (Table S3, column 2), indicating that 6- and 8-y-olds showed comparable actor bias. Thus, actor bias was present by age 6 y, and did not change between the ages of 6 and 8 y.

We next asked if recipient group influenced costly punishment of selfishness. Fig. 2B shows the probability of punishing selfishness by age and recipient group. A regression predicting punishment as a function of recipient group (1 = in-group, 0 = out-group), controlling for actor group, age, and sex, showed no significant effect of recipient group (β = 0.123, P = 0.182, odds ratio = 1.13) (Table S3, column 1). However, we did find a significant negative interaction between age and recipient group (β = −0.555, P = 0.003, odds ratio = 0.57) (Table S3, column 3). When we split our analyses by age group, we found a significant positive effect of recipient group within 6-y-olds (β = 0.426, P = 0.012, odds ratio = 1.53) (Table S4, column 1), indicating that 6-y-olds were more likely to punish selfishness that harmed in-group members than out-group members. In contrast, we found no significant effect of recipient group within 8-y-olds (β = −0.129, P = 0.107, odds ratio = 0.88) (Table S4, column 2), indicating that 8-y-olds punished selfishness equally, regardless of whether an in-group or an out-group member was harmed. Thus, recipient bias was present by age 6 y, but declined between the ages of 6 and 8 y.

Discussion

Our results show that from the earliest age at which children are known to systematically punish selfish sharing behavior, their norm enforcement favors the in-group over the out-group. Six-year-olds were more likely to sacrifice their own resources to punish selfish behavior when the selfish actor was an out-group member and when the disadvantaged recipient was an in-group member. Notably, these effects were observed when using minimal groups with which children had no prior experience or associations. Previous research using our same paradigm found that third-party children first began to systematically pay costs to punish selfishness at age 6; thus, our results suggest that norm enforcement is biased from its emergence. Our results have
implications for broader questions concerning children’s “default” morality, suggesting that punishment of unfair behavior starts off favoring the in-group over the out-group (40).

We also found that although norm enforcement increased between ages 6 and 8 y, with 8-y-olds showing more sensitivity to selfishness than 6-y-olds, in-group bias in norm enforcement did not increase: in fact, it partially declined. Six-year-olds and 8-y-olds were equally sensitive to the actor’s group, with both ages punishing selfishness more harshly when the actor was an out-group member. However, 6-y-olds were more influenced by the recipient’s group than 8-y-olds, with 8-y-olds punishing selfishness equally regardless of the recipient’s group. Thus, favoritism declined partially between 6 and 8 y, as children appeared to transition from a biased understanding of fairness norms (selfishness is wrong when it harms us) to a more impartial perspective (selfishness is generally wrong, regardless of whom it harms).

This result is consistent with a body of evidence that many children show declining explicit in-group bias over midchildhood as they are exposed to norms against discrimination (28, 50, 51). Our results thus suggest that between ages 6 and 8 y, children’s norm-enforcement behavior may become increasingly under control of explicit reasoning rather than implicit biases, and that older children may inhibit bias in their punishment. In contrast, our results are inconsistent with theories suggesting that punishment should become more biased with age as children adopt group norms supporting loyalty and bias. Although in-group bias appears to increase over childhood when children share resources with another individual in a first-party context (17, 41), we did not observe this pattern among 6- and 8-y-olds in our study of third-party punishment, perhaps suggesting that norms for group loyalty are less operative in this context.

To our knowledge, our study is the first to demonstrate in-group bias in children’s third-party punishment and to investigate the developmental trajectory of this bias. Previous research on infants has demonstrated that from a very young age, evaluations of an antisocial actor are biased by whether the evaluator is similar to the victim (32). However, to our knowledge, our work is the first to examine these early-emerging biases in the domain of fairness and sharing, and to ask how they influence costly third-party punishment of selfishness. Our results demonstrate that in-group biases are equally strong at ages 6 and 8 y, suggesting that they may be partially overcome over development.

Previous research has also investigated in-group bias in 7- and 11-y-olds’ punishment, but the authors simultaneously manipulated the group of both the actor and the recipient (i.e., both were in-group members or both were out-group members) and found no overall effect of group membership (14). One potential explanation for this null result is that these two manipulations can have opposite effects: Our 6-y-old subjects were more likely to punish out-group actors and to punish on behalf of in-group recipients. Thus, in-group actors may have decreased punishment, but in-group recipients may have increased punishment, resulting in no overall effect of group membership. Accordingly, to our knowledge, our results both document bias in children’s punishment behavior for the first time, and highlight the importance of separately manipulating actor and recipient group membership when investigating their influence on punishment.

Interestingly, we found that different forms of in-group bias showed different developmental trajectories in our study. Whereas bias on the basis of recipient group declined between ages 6 and 8 y, bias on the basis of actor group did not. This result suggests that children may come to punish fairness violations equally regardless of whom they harm, while at the same time remaining more lenient in their responses to selfish behavior that comes from in-group members.

Furthermore, although we found that recipient bias declined partially between ages 6 and 8 y, the tendency toward recipient bias in punishment does not completely disappear over development. Unlike our older subjects, third-party adults show evidence of punishing more on behalf of in-group recipients than out-group recipients (7, 8), suggesting that this bias can persist into adulthood. An interesting question is why adults in these studies did not inhibit this bias. One possibility is that the studies did not make salient the possibility for bias or norms against bias. Although our study used a group induction and a within-subjects design, emphasizing the potential for bias, these studies used preexisting groups and between-subjects designs. Thus, although we found that punishment became less biased between ages 6 and 8 y, this may reflect effortful inhibition of a persisting tendency toward in-group favoritism. Another possibility is that bias in punishment may continue to change in important ways over development into adulthood. Future research should investigate developmental changes over a wider range of ages.

The finding that selfish out-group members were punished more harshly than selfish in-group members is not predicted by group norm maintenance theories that people should punish selfish in-group members more harshly to maintain cooperative norms within their groups (7, 11). The finding is also inconsistent with “black-sheep effect” theories in social psychology, which posit that deviant in-group members should be judged more harshly than their out-group counterparts, because they threaten to damage the group’s norms or reputation (55). An interesting question is why these theories were not operative in our study. Notably, there are many contexts in which people do not show the black-sheep effect, but instead show straightforward in-group bias (56). The black-sheep effect is most likely to occur when “group-based motivational concerns” (57) are activated [for example, when one strongly identifies with the group, perceives a threat to its reputation, or believes that its members are seen as similar to each other (58–62)] and when a group-specific norm has been violated (63). This latter effect is consistent with evidence that children are more likely to protest when in-group members violate conventional norms (which children view as specific to groups or contexts; e.g., rules against hitting) (64–66). Our results suggest that in this way, children may view fairness norms more like moral rules than social conventions.

Additionally, our results are consistent with evidence that adults punish selfish out-group members more harshly than their in-group counterparts. Adult studies demonstrating harsher third-party punishment of selfish in-group members have simultaneously manipulated the group of the selfish actor and disadvantaged recipient (11, 14). Studies that have separately manipulated actor and recipient group membership and found an effect of actor group on adult’s third-party punishment have consistently found harsher punishment of out-group actors (7, 9, 10). Similarly, out-group members receive more second-party punishment for selfish behavior (12, 13), and in the context of legal punishment, laboratory and field studies suggest that racial out-group members are judged more harshly (67, 68). Thus, there appear to be many contexts in which selfish behavior does not trigger black-sheep or group norm maintenance effects.

Another interesting open question is if the effects of group membership operated by influencing judgments or behavior. In other words, would an experiment measuring moral judgments have produced the same results? In our study, punishment of selfish behavior required subjects both to judge the actor’s allocation as bad and to act on this judgment by punishing the actor. Thus, one possibility is that the effects of group membership on
punishment, or the developmental decline in bias, primarily reflected effects on subjects’ judgments. That is, perhaps unequal allocations were judged as worse when they came from out-group members and harmed in-group members, and this later effect declined between ages 6 and 8 y. This possibility is congruent with studies showing that adults perceive selfish behavior from an out-group member as more hostile or aggressive (12). Alternatively, subjects may have judged selfish allocations equally regardless of group membership, but shown differences in their punishment responses to those judgments; for example, if it were more satisfying to impose costs on selfish actors who were out-group members and harmed in-group members. Future research should attempt to differentiate between these possibilities.

Finally, we note that our design involved a minimal group induction that emphasized the contrast between the two groups, perhaps fostering a sense of intergroup competition that strengthened our effects (69). An interesting open question is the extent to which group membership influences children’s punishment when more than two groups are involved.

Fairness norms play a key role in human cooperation, and adults sometimes express in-group bias when enforcing these norms. Here, we have investigated the development of in-group bias in children’s costly punishment of selfishness. Our results suggest that the development of in-group bias does not “corrupt” norm-enforcement behavior over childhood. Rather, norm enforcement appears to be biased by our intergroup psychology from its emergence, with in-group favoritism in part declining between ages 6 and 8 y. These results have implications for theories of human fairness and morality, building on a body of research suggesting that adult punishment may develop (9, 11–14) and more broadly, that in-group bias, cooperation, and norm enforcement are interacting processes (15–18) that children must integrate over development (42). Our finding that costly punishment is biased from its earliest emergence in development suggests that our intergroup psychology has an important influence on the way we think about and respond to selfishness. However, our finding that punishment becomes less biased between the ages of 6 and 8 y suggests that children can partially overcome their in-group favoritism. Taken together, these results shed light on the psychology of norm enforcement in both children and adults.

Materials and Methods

Participants. We tested n = 32 6–y-olds (mean = 6.3 y, range = 6.0–6.9 y, 16 females) and n = 32 8–y-olds (mean = 8.4 y, range = 8.0–9.9 y, 16 females). Eleven additional subjects were excluded because of: experimenter error (one child), apparatus malfunction (one child), refusal to participate (one child), or parental interference (one child). Children were recruited from the Harvard Laboratory for Developmental Studies database, as in previous work from our group. This study was approved by the Harvard University Institutional Review Board, F18470–117, and written parental consent was provided for all subjects.

Group Induction. In the first stage of the experiment, we assigned subjects to the “blue” or “yellow” team, based on their color preference. We then asked subjects to wear a team-colored party hat and to draw with a team-colored marker.

Manipulation Check. In the second stage of the experiment, we measured in-group bias to assess that our group induction successfully induced bias and to increase the salience of our minimal groups. We presented subjects with 10 pairs of in-group and out-group members, and assessed in-group preferences. We explained that a group of children of the subject’s age and sex had previously also been divided into a blue and yellow team. We then presented subjects with representations of sex-matched in-group and out-group members. In-group and out-group members were represented as paper bags with names, drawings of faces, and blue or yellow party hats.

Costly Punishment Method

Procedure. We began the third-party punishment phase by introducing subjects to Skittles, the candy reward being used, and the experimental apparatus. We demonstrated that Skittles could be distributed across the apparatus, which had a handle that could be moved in the green (accept) and red (reject) directions. We explained that the previously described children from the classroom had played a three-person game with the apparatus yesterday. We explained that each subject acted as the actor, and recipient name was paired with an endowment of 33 Skittles. Before proceeding, we explained that only fair allocations would go to the actor and recipient, but rejected allocations would be thrown out. We explained that the experimenter would later return the bags to the actors and recipients to keep. We also had subjects decorate a bag for their own Skittle home in at the end of the game.

We presented subjects with their own endowment of 33 Skittles, placed in a green box. We also introduced subjects to a red box. We instructed subjects to take a Skittle out of the green box before making each decision. To accept an allocation, subjects were to return the Skittle to the green box and then pull the handle in the green direction. To reject an allocation, subjects were to move the Skittle into the red box and then pull the handle in the red direction. Then, at the end of the game, subjects could take home all Skittles left in the green box, but skittles in the red box would be thrown away. Thus, rejection required sacrificing Skittles.

Before starting the game, we asked subjects comprehension questions to ensure that they understood: (i) that the actor and recipient had not yet taken Skittles home; (ii) what to do before accepting and rejecting allocations; (iii) who would get to take home the Skittles in the actor and recipient’s bags; (iv) what team the actor and recipient were each on, and if these were the subject’s team; and (v) what would happen to the Skittles in each box after the game. All subjects answered all questions either spontaneously correctly or correctly after additional questioning, except that one subject was not asked the first question because of experimenter error.

In each trial, we demonstrated the allocation and reminded subjects to move a Skittle to the green or red box before deciding. We transferred accepted allocations to the actor and recipient bags; rejected allocations disappeared under the apparatus. After each set of five trials, we introduced subjects to the next actor-recipient pair, repeating the relevant comprehension questions. After all trials, a second experimenter entered the room and asked a series of questions.

---

including whether the child believed that the other kids (i) were real and (ii) would really come in to collect their Skittles. All sessions were videotaped and recorded. Disagreements between coders were rare (less than 3% of trials) and resolved by revwatching the video (for more details, see SI Text, Reliability Coding Protocol). One subject received only 19 of 20 trials; this subject’s data were included in analyses and this trial was treated as a missing data point.