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

Jillian J. Jordan[a,b,1], Katherine McAuliffe[a,b,c], and Felix Warneken[a]

[a]Department of Psychology and [c]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138; and [b]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, 10). 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.

[1]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.

[†]McAuliffe K, Jordan JJ, Warneken F, Biennial Meeting of the Society for Research in Child Development, April 18–20, 2013, Seattle, WA.

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, older children were more likely to preferentially allocate positive 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 may be exposed to social norms encouraging in-group loyalty and bias (43–45). Evidence suggests that children view disloyalty within a group as counter-normative, and expect groups to exclude disloyal individuals (46). Furthermore, whereas children generally view the use of stereotypes to exclude others as morally wrong (47), evidence suggests that over childhood they become more willing to tolerate exclusion to promote effective group functioning (48). Thus, some evidence suggests that in-group bias can increase over development.

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 most 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 selfish sharing behavior, but not fair behavior. In contrast, 5-y-olds did not systematically discriminate between selfishness and fairness, suggesting that punishment emerged at age 6. In the present study, we used this same paradigm and subject pool to measure in-group bias in 6- and 8-y-old's punishment. By testing 6-y-olds, we asked if punishment is biased from its emergence in development. By investigating differences between 6- and 8-y-olds, we asked how a potential in-group bias might change between these ages. We focused on these ages because of the evidence that the window between 6 and 8 y is important for the development of children's intergroup and moral psychologies, with changes in parochial altruism (17), explicit in-group bias (33, 50, 51), and conflicts between group-based and morality-based values (42).

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, Manipulation Check Results*).

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. S1*C*).

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. S1*D*). To measure the effect of group membership, we manipulated within subject whether the actor and recipient were in the subject's group, resulting in four conditions (actor in, recipient in;

[‡]McAuliffe K, Jordan JJ, Warneken F, Biennial Meeting of the Society for Research in Child Development, April 18–20, 2013, Seattle, WA.

actor in, recipient out; actor out, recipient in; actor out, recipient out). To reduce the effects of noise and gain more signal from our binary punishment measure, subjects received five trials (four selfish, one fair) per condition (*SI Text*, *Discussion of the Use of Repeated Trials*). We led subjects to believe that the actors and recipients were peer children who had played the game previously and would later receive allocations of candy that varied in size, depending on the subject's decisions (Fig. S1*E*). We then asked subjects two questions to test their beliefs that the other children were real and would really receive their candy. For more details, see *Materials and Methods*.

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 ($\beta = 2.31$, $P < 0.001$, odds ratio = 10.04) (Table S1, column 1). There was also a significant positive interaction between selfishness and age ($\beta = 2.04$, $P = 0.022$, 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 ($\beta = 1.44$, $P = 0.001$, odds ratio = 4.21) (Table S2, column 1) and 8-y-olds ($\beta = 3.50$, $P < 0.001$, 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.

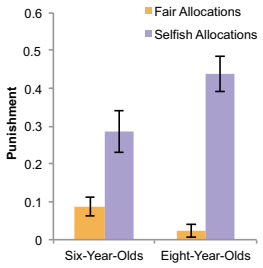

Fig. 1. Subjects ($n = 64$) pay costs to punish selfishness, but rarely punish fairness and become more systematic with age. We plot the proportion of trials in which 6-y-olds ($n = 32$) and 8-y-olds ($n = 32$) engaged in costly third-party punishment of fair and selfish allocations. Error bars reflect ± 1 SEM (clustered on subject to account for repeated observations).

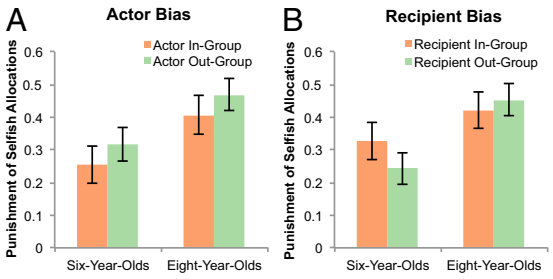

Fig. 2. Six-year-olds show two kinds of in-group bias, but 8-y-olds show only one. We plot the proportion of trials in which subjects ($n = 64$) punished selfish allocations. (*A*) Both 6-y-olds ($n = 32$) and 8-y-olds ($n = 32$) were more likely to punish out-group than in-group actors. (*B*) Six-year-olds were more likely to punish on behalf of in-group than out-group recipients, but 8-y-olds showed no recipient bias. Error bars reflect ± 1 SEM (clustered on subject to account for repeated observations).

We began by asking if actor group influenced costly punishment of selfishness. Fig. 2*A* 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 ($\beta = -0.280$, $P = 0.013$, 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 ($\beta = 0.053$, $P = 0.822$, 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. 2*B* 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 ($\beta = 0.123$, $P = 0.182$, odds ratio = 1.13) (Table S3, column 1). However, we did find a significant negative interaction between age and recipient group ($\beta = -0.555$, $P = 0.003$, 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 ($\beta = 0.426$, $P = 0.012$, 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 ($\beta = -0.129$, $P = 0.107$, 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

[§]McAuliffe K, Jordan JJ, Warneken F, Biennial Meeting of the Society for Research in Child Development, April 18–20, 2013, Seattle, WA.

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 these biases are strong enough to influence children's norm-enforcement behavior, even when this requires personally sacrificing resources and could violate potential impartiality concerns. Furthermore, we demonstrate that the influence of these early-emerging biases declines between 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 over this age range, 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 of a game), but not moral norms (which children view as universally applicable; 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 can show in-group bias (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 = 8.0–8.9 y, 16 females). Eleven additional children were excluded because of: experimenter error (eight children), 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. Over 10 trials, we presented subjects with 10 choices between an in- and out-group member (alternating between presenting the in-group member on the left or the right). In the first four trials, we described an action of positive valence (i.e., "he helped his parents clean up") and asked who had completed the action; in the next three trials, we gave the subject a sticker to drop in one of the two bags; in last three trials, we asked the subject who they liked better (methods adapted from ref. 39). For each subject, we randomized the pairing between individual bags and (*i*) trial type and (*ii*) blue or yellow team.

Punishment Game. ***Overview.*** In the third and final stage of the experiment, we measured third-party punishment. We used a modified version of the third-party punishment task used in previous work from our group.[¶] Each subject made 20 decisions to accept or reject a proposed allocation of candy between an actor and a recipient. We crossed manipulations of actor (A) and recipient (R) group membership by presenting each subject with four different sex-matched actor–recipient pairs (A in, R in; A in, R out; A out, R in; A out, R out). For each actor–recipient pair, subjects were presented with five allocations that the actor proposed, four of which were selfish (actor kept six and gave zero), and one of which was fair (actor kept three and gave three). If a subject chose to accept an allocation, it was enacted; if a subject chose to reject an allocation, the candy was thrown out. Subjects also received their own endowment of 33 Skittles and had to sacrifice one candy every time they rejected. Thus, rejection constituted costly punishment: it imposed a cost on both the subject and actor. Previous research using this method found that this cost (one Skittle per trial) was salient enough to deter rejection, relative to a cost-free control condition[¶] (for more discussion, see *SI Text, Discussion of Costly Punishment Method*).

We presented the fair allocation in a different position (first, second, third, or fourth trial) for each actor–recipient pair. We used Latin squares to counterbalance: (*i*) the order in which the four group conditions were presented; (*ii*) the order in which the four fair-allocation positions were presented; (*iii*) the order in which the actor and recipient names were presented; and (*iv*) the group condition that each actor and recipient name was paired with.

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 these children had formed pairs and done the first two jobs in the game. We then asked subjects to do the third.

Next, we introduced subjects to the first actor–recipient pair. Actors and recipients were represented with paper bags, which were similar to the bags used in the previous stage of the experiment. We explained that the actor had divided Skittles between him- or herself and the recipient, and demonstrated the actor's proposed allocations on index cards. In a counterbalanced order, we showed subjects one fair card (corresponding to three Skittles on each side) and one selfish card (corresponding to six Skittles on the actor's side). Subjects practiced pulling the handle in both the green (accept) and red (reject) directions, and we explained that accepted 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 to take their own Skittles 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 to 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,

[¶]McAuliffe K, Jordan JJ, Warneken F, Biennial Meeting of the Society for Research in Child Development, April 18–20, 2013, Seattle, WA.

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 recoded for reliability. Disagreements between coders were rare (less than 3% of trials) and resolved by rewatching 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.

ACKNOWLEDGMENTS. We thank members of the F.W. laboratory for help with data collection and processing; and three anonymous reviewers, C. Baker, A. Bear, J. Cone, Y. Dunham, M. Hoffman, D. Rand, J. Phillips, and A. Zheutlin for helpful feedback. This study was supported in part by the Science of Generosity Initiative of the John Templeton Foundation, the Harvard University Mind/Brain/Behavior Interfaculty Initiative, the Harvard University BLISS Fellowship Program, and the Herchel Smith Harvard Undergraduate Science Research Program.

1. Boyd R, Richerson PJ (1992) Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethol Sociobiol* 13(3):171–195.
2. Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415(6868):137–140.
3. Fowler JH (2005) Altruistic punishment and the origin of cooperation. *Proc Natl Acad Sci USA* 102(19):7047–7049.
4. Henrich J, et al. (2006) Costly punishment across human societies. *Science* 312(5781): 1767–1770.
5. Fehr E, Fischbacher U (2004) Third-party punishment and social norms. *Evol Hum Behav* 25(2):63–87.
6. Nikiforakis N, Mitchell H (2014) Mixing the carrots with the sticks: Third party punishment and reward. *Exp Econ* 17(1):1–23.
7. Bernhard H, Fischbacher U, Fehr E (2006) Parochial altruism in humans. *Nature* 442(7105):912–915.
8. Götte L, Huffman D, Meier S (2006) The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *Am Econ Rev* 96(2):212–216.
9. Schiller B, Baumgartner T, Knoch D (2014) Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evol Hum Behav* 35(3):169–175.
10. Baumgartner T, Götte L, Gügler R, Fehr E (2012) The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Hum Brain Mapp* 33(6):1452–1469.
11. Shinada M, Yamagishi T, Ohmura Y (2004) False friends are worse than bitter enemies: "Altruistic" punishment of in-group members. *Evol Hum Behav* 25(6):379–393.
12. Kubota JT, Li J, Bar-David E, Banaji MR, Phelps EA (2013) The price of racial bias: Intergroup negotiations in the ultimatum game. *Psychol Sci* 24(12):2498–2504.
13. Mussweiler T, Ockenfels A (2013) Similarity increases altruistic punishment in humans. *Proc Natl Acad Sci USA* 110(48):19318–19323.
14. Gummerum M, Takezawa M, Keller M (2009) The influence of social category and reciprocity on adults' and children's altruistic behavior. *Evol Psychol* 7(2):295–316.
15. Choi JK, Bowles S (2007) The coevolution of parochial altruism and war. *Science* 318(5850):636–640.
16. Greene J (2013) *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them* (Penguin, New York).
17. Fehr E, Bernhard H, Rockenbach B (2008) Egalitarianism in young children. *Nature* 454(7208):1079–1083.
18. Boyd R, Gintis H, Bowles S, Richerson PJ (2003) The evolution of altruistic punishment. *Proc Natl Acad Sci USA* 100(6):3531–3535.
19. Gummerum M, Hanoch Y, Keller M (2008) When child development meets economic game theory: An interdisciplinary approach to investigating social development. *Hum Dev* 51(4):235–261.
20. Schmidt MF, Sommerville JA (2011) Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS ONE* 6(10):e23223.
21. Hamann K, Warneken F, Greenberg JR, Tomasello M (2011) Collaboration encourages equal sharing in children but not in chimpanzees. *Nature* 476(7360):328–331.
22. Warneken F, Lohse K, Melis AP, Tomasello M (2011) Young children share the spoils after collaboration. *Psychol Sci* 22(2):267–273.
23. Blake PR, McAuliffe K (2011) "I had so much it didn't seem fair": Eight-year-olds reject two forms of inequity. *Cognition* 120(2):215–224.
24. Shaw A, Olson KR (2012) Children discard a resource to avoid inequity. *J Exp Psychol Gen* 141(2):382–395.
25. Hamlin JK, Wynn K, Bloom P, Mahajan N (2011) How infants and toddlers react to antisocial others. *Proc Natl Acad Sci USA* 108(50):19931–19936.
26. Rakoczy H, Warneken F, Tomasello M (2008) The sources of normativity: Young children's awareness of the normative structure of games. *Dev Psychol* 44(3):875–881.
27. Vaish A, Missana M, Tomasello M (2011) Three-year-old children intervene in third-party moral transgressions. *Br J Dev Psychol* 29(Pt 1):124–130.
28. Dunham Y, Baron AS, Banaji MR (2008) The development of implicit intergroup cognition. *Trends Cogn Sci* 12(7):248–253.
29. Kelly DJ, et al. (2005) Three-month-olds, but not newborns, prefer own-race faces. *Dev Sci* 8(6):F31–F36.
30. Kinzler KD, Dupoux E, Spelke ES (2007) The native language of social cognition. *Proc Natl Acad Sci USA* 104(30):12577–12580.
31. Mahajan N, Wynn K (2012) Origins of "us" versus "them": Prelinguistic infants prefer similar others. *Cognition* 124(2):227–233.
32. Hamlin JK, Mahajan N, Liberman Z, Wynn K (2013) Not like me = bad: Infants prefer those who harm dissimilar others. *Psychol Sci* 24(4):589–594.
33. Baron AS, Banaji MR (2006) The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychol Sci* 17(1):53–58.
34. Zinser O, Bailey RC, Edgar RM (1976) Racial recipients, social distance, and sharing behavior in children. *Soc Behav Personal* 4(1):65–74.
35. Monteiro MB, de França DX, Rodrigues R (2009) The development of intergroup bias in childhood: How social norms can shape children's racial behaviours. *Int J Psychol* 44(1):29–39.
36. Tajfel H, Billig MG, Bundy RP, Flament C (1971) Social categorization and intergroup behaviour. *Eur J Soc Psychol* 1(2):149–178.
37. Bigler RS (1995) The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Dev* 66(4):1072–1087.
38. Vaughan GM, Tajfel H, Williams J (1981) Bias in reward allocation in an intergroup and an interpersonal context. *Soc Psychol Q* 44(1):37–42.
39. Dunham Y, Baron AS, Carey S (2011) Consequences of "minimal" group affiliations in children. *Child Dev* 82(3):793–811.
40. Bloom P (2013) *Just Babies: The Origins of Good and Evil* (Random House, New York).
41. Buttelmann D, Böhm R (2014) The ontogeny of the motivation that underlies in-group bias. *Psychol Sci* 25(4):921–927.
42. Rutland A, Killen M, Abrams D (2010) A new social-cognitive developmental perspective on prejudice the interplay between morality and group identity. *Perspect Psychol Sci* 5(3):279–291.
43. Abrams D, Rutland A, Pelletier J, Ferrell JM (2009) Children's group nous: Understanding and applying peer exclusion within and between groups. *Child Dev* 80(1): 224–243.
44. Abrams D, Rutland A, Cameron L, Ferrell J (2007) Older but wilier: In-group accountability and the development of subjective group dynamics. *Dev Psychol* 43(1):134–148.
45. Rutland A (1999) The development of national prejudice, in-group favouritism and self-stereotypes in British children. *Br J Soc Psychol* 38(1):55–70.
46. Abrams D, Rutland A, Cameron L (2003) The development of subjective group dynamics: Children's judgments of normative and deviant in-group and out-group individuals. *Child Dev* 74(6):1840–1856.
47. Killen M, Lee-Kim J, McGlothlin H, Stangor C (2002) How children and adolescents evaluate gender and racial exclusion. *Monogr Soc Res Child Dev* 67(4):i–vii, 1–119.
48. Killen M, Stangor C (2001) Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child Dev* 72(1):174–186.
49. Rutland A, Cameron L, Milne A, McGeorge P (2005) Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Dev* 76(2):451–466.
50. Aboud F (1988) *Children and Prejudice* (Blackwell, New York, NY).
51. Raabe T, Beelmann A (2011) Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta-analysis of age differences. *Child Dev* 82(6):1715–1737.
52. Killen M, Pisacane K, Lee-Kim J, Ardila-Rey A (2001) Fairness or stereotypes? Young children's priorities when evaluating group exclusion and inclusion. *Dev Psychol* 37(5): 587–596.
53. Aloise-Young PA (1993) The development of self-presentation: Self-promotion in 6- to 10-year-old children. *Soc Cogn* 11(2):201–222.
54. Olson KR, Dweck CS, Spelke ES, Banaji MR (2011) Children's responses to group-based inequalities: Perpetuation and rectification. *Soc Cogn* 29(3):270–287.
55. Marques JM, Yzerbyt VY, Leyens JP (1988) The "black sheep effect": Extremity of judgments towards ingroup members as a function of group identification. *Eur J Soc Psychol* 18(1):1–16.
56. Linville PW, Jones EE (1980) Polarized appraisals of out-group members. *J Pers Soc Psychol* 38(5):689–703.
57. Reese G, Steffens MC, Jonas KJ (2013) When black sheep make us think: Information processing and devaluation of in-and outgroup norm deviants. *Soc Cogn* 31(4):482–503.
58. Branscombe NR, Wann DL, Noel JG, Coleman J (1993) In-group or out-group extremity: Importance of the threatened social identity. *Pers Soc Psychol Bull* 19(4):381–388.
59. Lewis AC, Sherman SJ (2010) Perceived entitativity and the black-sheep effect: When will we denigrate negative ingroup members? *J Soc Psychol* 150(2):211–225.
60. Castano E, Paladino MP, Coull A, Yzerbyt VY (2002) Protecting the ingroup stereotype: Ingroup identification and the management of deviant ingroup members. *Br J Soc Psychol* 41(Pt 3):365–385.
61. Coull A, Yzerbyt VY, Castano E, Paladino M-P, Leemans V (2001) Protecting the ingroup: Motivated allocation of cognitive resources in the presence of threatening ingroup members. *Group Process Intergroup Relat* 4(4):327–339.
62. Eidelman S, Biernat M (2003) Derogating black sheep: Individual or group protection? *J Exp Soc Psychol* 39(6):602–609.
63. Marques JM (1990) The black sheep effect: Outgroup homogeneity in social comparison settings. *Social Identity Theory: Constructive and Critical Advances*, eds Abrams D, Hogg MA (Harvester Wheatsheaf, London) pp 131–151.
64. Schmidt MF, Rakoczy H, Tomasello M (2012) Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition* 124(3):325–333.
65. Smetana JG (1981) Preschool children's conceptions of moral and social rules. *Child Dev* 52(4):1333–1336.
66. Turiel E (1983) *The Development of Social Knowledge: Morality and Convention* (Cambridge Univ Press, Cambridge, UK).
67. Sommers SR, Ellsworth PC (2000) Race in the courtroom: Perceptions of guilt and dispositional attributions. *Pers Soc Psychol Bull* 26(11):1367–1379.
68. Blair IV, Judd CM, Chapleau KM (2004) The influence of Afrocentric facial features in criminal sentencing. *Psychol Sci* 15(10):674–679.
69. Hartstone M, Augoustinos M (1995) The minimal group paradigm: Categorization into two versus three groups. *Eur J Soc Psychol* 25(2):179–193.