Paying back those who harmed us but not those who helped us:

Direct negative reciprocity precedes direct positive reciprocity in early development

Nadia Chernyak
Department of Cognitive Sciences, University of California - Irvine

Kristin L. Leimgruber’
Department of Psychology, Franklin & Marshall College

Yarrow C. Dunham
Department of Psychology, Yale University

Jingshi Hu
Department of Psychological & Brain Sciences, Boston University

Peter R. Blake
Department of Psychological & Brain Sciences, Boston University

In press at Psychological Science

Acknowledgements. We would like to thank the Social Development and Learning Lab for assistance with data collection and coding; the Living Laboratory at the Museum of Science, Boston and participating preschools for use of space; Samuel Ronfard and Felix Warneken for providing helpful suggestions on earlier versions of this manuscript; Alexander Etz for fielding questions about statistical analyses. This work was funded by an Expanding the Science and Practice of Gratitude Postdoctoral Fellowship from the Greater Good Science Center and the John Templeton Foundation to NC.
Abstract

The principle of direct reciprocity, or paying back specific individuals, is assumed to be a critical component of everyday social exchange and a key mechanism for the evolution of cooperation. Young children know the norm of reciprocity, but it is unclear whether they follow the norm for both positive and negative direct reciprocity or whether reciprocity is initially generalized. Across 5 experiments (N=330), we show that children between 4 and 8 years of age engaged in negative direct reciprocity but generalized positive reciprocity, despite recalling benefactors. Children did not endorse the norm of positive direct reciprocity as applying to them until about 7 years of age (Study 4), but a short social norm training enhanced this behavior in younger children (Study 5). Results suggest that negative direct reciprocity develops early, whereas positive reciprocity only becomes targeted to specific others as children learn and adopt social norms.

**Keywords:** direct reciprocity; altruism; social groups; cognitive development; social norms

Abstract Word Count: 146 (max: 150)
Paying back those who harmed us but not those who helped us:

Direct negative reciprocity precedes direct positive reciprocity in early development.

Proverbs such as “You scratch my back, I scratch yours” and “tit for tat” are deeply embedded within our cultural landscape. These sayings refer to the principle of direct reciprocity, rewarding and punishing those who benefit and harm us, respectively. Direct reciprocity has been proposed as a key mechanism for the evolution of cooperation (Axelrod & Hamilton, 1981; Nowak, 2006; Trivers, 1971) and may be a universal component of human social group functioning (Fiske, 1992; Gouldner, 1960). Yet despite its theoretical and social importance, little is known about the developmental trajectory of direct reciprocity or the psychological mechanisms that support its emergence.

Existing studies on the development of reciprocal behavior have found evidence of generalized reciprocity in children (paying forward rewards and punishments to new people) but no clear evidence of direct reciprocity (paying back specific individuals). In particular, past experiments have given children a choice of actions (i.e., to reward or punish), but not a choice of whom to direct those actions towards (House, Henrich, Sarnecka, & Silk, 2013; Leimgruber et al., 2014; Paulus, 2016; Sebastián-Enesco & Warneken, 2015; Warneken & Tomasello, 2013). Prior work has also found that children as young as 3.5 years of age understand the principle of direct positive reciprocity and distribute rewards to third parties accordingly (Hamlin, Wynn, Bloom, & Mahajan, 2011; Olson & Spelke, 2008). However, whether children use this principle to guide their own behavior remains an open question. This is particularly relevant for resource allocation situations given that children know the norms for distribution well before they follow them (Blasi, 1983; Smith, Blake, & Harris, 2013).
Another consideration in the development of reciprocity is the potential role of social groups. By learning about the cooperative tendencies of groups, individuals can simplify the problem of finding good cooperation partners by identifying kinds of people that tend to be better or worse cooperators and directing reciprocal actions towards them (Rand & Nowak, 2013). We refer to this form of reciprocity as group-based reciprocity, i.e., a form of reciprocity in which the individual directs reciprocal actions (whether positive or negative) towards individuals who share the group membership of the prior benefactor or malefactor. Preschoolers are adept at tracking and reasoning about group membership and may find it easier to generalize to a group instead of tracking individuals and their actions (Dunham, Baron, & Carey, 2011; Misch, Over, & Carpenter, 2014; Rhodes, 2012). Thus, it is possible that group-based reciprocity emerges earlier in development than direct reciprocity, which is person-specific. To ensure that we could capture any such trend we incorporated an intergroup dimension into our study design.

A final consideration is that positive and negative forms of reciprocity may be supported by distinct cognitive mechanisms and therefore have different developmental trajectories. For example, recent research with adults has found that negative actions result in stronger reciprocal responses compared to positive actions (Gray, Ward, & Norton, 2014; Keysar, Converse, Wang, & Epley, 2008). This is due, in part, to the fact that negative actions evoke strong negative affect compared to control conditions whereas positive actions do not necessarily increase positive affect (Gray, Ward, & Norton, 2014). Existing research on reciprocity in children, however, has not examined positive and negative reciprocity separately. Instead, studies have presented children with a forced choice between a positive and negative actor, thus making it unclear whether they are rewarding the positive actor or punishing the negative one (Dunfield &
Kuhlmeier, 2010; House et al., 2013; Robbins & Rochat, 2011; Vaish, Carpenter, & Tomasello, 2010).

Documenting the developmental trajectories of both positive and negative direct reciprocity is critical to the broader project of understanding the ontogeny and evolution of cooperation, as well as to properly contextualize prominent findings in this area. These distinctions become even more important in light of recent debates concerning whether direct reciprocity is a later-emerging ability (Warneken, 2018) whose function is to stabilize cooperation, or whether, in contrast, direct reciprocity is an early-developing ability whose function is to reward good agents in a targeted manner in order to encourage future positive interactions (Dunfield & Kuhlmeier, 2010).

In the current study, we investigated the early emergence of direct reciprocity in a novel experimental paradigm. Children were the recipients of a positive or negative action from one of four agents. In different trials, the benefactors or malefactors were either in-group members, outgroup members, or not in any group. Children could then reciprocate the same action (with a different resource) to any of the four agents. Critically, this design allows us to distinguish between children’s engagement in *direct reciprocity* (paying back the specific prior actor), *group-based reciprocity* (i.e., paying back members of the same group that previously acted), or *generalized reciprocity* (paying forward the prior action indiscriminately). Because direct reciprocity requires tracking specific individuals (Olson & Spelke, 2008; Trivers, 1971), we tested children’s memories for who performed the prior actions. This allowed us to determine whether any lack of direct reciprocity could be explained by a failure to encode the individuals involved in each transaction.

**Study 1: Do preschoolers engage in direct reciprocity?**
In the first study, we asked whether a) children track the specific identities of individuals who benefit and harm them and b) are motivated to reciprocate in kind. We investigated both positive and negative reciprocity.

**Methods**

We designed a computer game in which children interacted with four computerized confederates (represented via animal avatars) described as other children who were playing the game remotely. During each trial of the game (see Figure 1), children started with an *initial phase* in which they encountered benefactors (people who gave the children a benefit while incurring a personal cost; Giving Condition) or malefactors (Stealing Condition; people who took benefits away from the children). During the second, *reciprocating phase*, children had the opportunity to engage in reciprocity towards their benefactors and malefactors by either receiving a *new* object that they could then give to one of the other participants (Giving Condition) or watching others receive new objects, one of which that they could then steal (Stealing Condition). In both conditions, children were given a choice of four agents to reciprocate towards, and our main analyses focused on whether children directly reciprocated towards the prior benefactor or malefactor (direct reciprocity). A secondary analysis focused on group-based reciprocity: whether children reciprocated towards social groups that had previously benefitted or harmed them rather than specific individuals.

In Study 1, we were interested in children’s capacity to engage in reciprocity as opposed to their motivation to engage in it. We thus used a non-voluntary design in which children were directed during reciprocating phases to give or steal and thus had no choice of action but only a choice of recipient. For the group-based trials, we used a minimal group paradigm (Dunham, 2018; Rhodes, Leslie, Saunders, Dunham, & Cimpian, 2017) and focused on 4-5-year-olds based
on prior research showing a capacity for generalized reciprocity in this age range (Leimgruber et al., 2014). Computerized software (Inquisit 5, 2016) allowed us to fully control the interaction children had as well as ensure that our stimuli did not incidentally contain any gender, racial, or socioeconomic cues.
Participants

Eighty total children were tested at parks, in the laboratory, at local school centers, or a local children’s museum (Mean age = 4.88 years, Range = 3.65-5.93, 38 females, 42 males). At each site, parents were first approached to ask if they were interested in having their child participate. In this first study, we focused on the ages of 4-5 since those are the earliest ages during which intergroup biases are clearly present. Seven additional children were tested but excluded due to an equipment/software malfunction (n = 2), being outside the targeted age range (n = 4), or insufficient command of English (n = 1). For all studies, we set an a priori minimum of N = 40 per condition in order to double recommended minimum standards (Simmons, Nelson, & Simonsohn, 2011; but see also 2018 for an updated view). To further increase power, when feasible, we used a within-subjects design with multiple trials and treated age as a continuous variable. We tested until we achieved this minimum or until the lead researcher was no longer available (Studies 4 and 5). All participants were drawn from a mid-sized US city (Boston, MA). Demographics on individual participants were not collected but evaluations of the testing sites and studies using the same sites have found that the majority (60-70% of participants) were predominantly of European-American background, mid-to-high socioeconomic status, and generally reflected the demographics of the surrounding area (Blake & McAuliffe, 2011).

Procedure

All children were told that they would be playing a computerized game with “other kids” who were at “other schools/parks/museums”. To ensure that children believed they had anonymity and were not interacting with peers they had previously seen, children were asked two questions (“Do you know who the other kids are?” and “Do they know who you are?”) and
were provided feedback to correct their responses if necessary. Once children answered these questions correctly, they were shown four animal characters (a bear, monkey, dog, and cat; presented in the middle of the screen in random order) and asked to choose an avatar that was then used to represent the child for the remainder of the game. The next sequence of events differed for the positive versus negative reciprocity conditions and is depicted in Figure 1. Between each trial (and before the first warm-up trial) a clock was shown to simulate a delay in which the confederate children were ostensibly choosing their avatars.

Children then completed four trials (full details provided in Supplementary Materials), two of which occurred within a group context (referred to as the in-group trial and the out-group trial) and two outside of a group context (referred to as the initial trial and perceptual similarity control). After completing the initial trial (no group markers), children were assigned to either a green group or blue group (following procedures used in Rhodes et al., 2017). They then received the in-group and out-group trials (order counterbalanced) in which the benefactor or malefactor was an in-group member (same group as the child) or out-group member, respectively. Following these trials, group assignment was removed (children were told there would be no more groups and all group-based markers were taken away), and children completed a final trial (perceptual similarity control) in which there were 2 pairs of 2 perceptually similar characters. We included this trial in order to test for the possibility that any group-based reciprocity was driven by perceptual similarity between characters. Children completed all of the trials for one of two conditions, described below:

**Giving Condition.** In the giving condition, participants were shown that the game had allocated a sticker to each of the confederate children, but not the participant child (see Figure 1, panel A). Children were then told that each of the confederates had the choice of either keeping
the sticker given to him/her or giving their sticker to the child. After a 3-second delay meant to simulate a decision-making process, one randomly selected character (referred to hereafter as the benefactor) gave a sticker to the participant child’s avatar on screen. The experimenter reiterated that one person had chosen to give the sticker to the child and placed 1 sticker into the child’s box.

All stickers were then cleared off the screen, and children proceeded to the reciprocating phase, in which they were told they received a new sticker from the game that was different from the one they had just received from the benefactor. The experimenter held up a new sticker to make this clear. The child was then told that s/he could choose which of the confederates to give the new sticker to. We note that this giving behavior was not voluntary (nor costly) to the child in the sense that all children had to give the sticker to one of the four confederates and were not allowed to keep the sticker for themselves. After making their choices, children viewed the sticker being moved towards their chosen recipient on the screen. After each trial, children answered two memory check questions (described at the end).

Stealing Condition. Children saw a similar sequence presented in a stealing context: there were four computerized confederates on screen and the game had allocated a sticker only to the child (their avatar). Next, children were told that each of the confederates could either let the child keep the sticker or steal it for themselves. After a 3-second delay meant to simulate a decision-making process, one randomly selected character (referred to hereafter as the malefactor) stole the sticker from the participant child.

All stickers were then cleared off the screen, and children proceeded to the reciprocating phase, in which they were told that all the confederates now received new stickers but the
participant child did not. The child then had to choose which of the confederates to steal a sticker from. Similar to the Giving condition, stealing was not voluntary and was not costly to the child.

**Memory Check Questions:** After each trial, children were asked (1) to recall their benefactor/malefactor (“Who gave you a sticker?”/ “Who took a sticker from you?”) and (2) to recall the child’s beneficiary/maleficiary (“Whom did you give a sticker to?”/ “Whom did you take a sticker from?”). We included these questions to test whether a lack of direct reciprocity could be attributed to having forgotten who the benefactor was. Due to a program glitch, a small portion of memory check questions were asked but not recorded by the software ($n = 24$ of $640$ total memory check questions asked in Study 1; $n = 56$ of $680$ in Study 2).

**Analytic approach**

All analyses were performed in R. Initial analyses for all studies showed no gender effects, and thus data were collapsed across these variables. For all models, we ran binary logistic mixed effects models using Age (coded continuously in years), Trial Type, and Condition (when applicable) as fixed effects predictors, Subject ID as a random effect, and the target dependent variable (e.g., Engaged in Direct Reciprocity) as the binomial response. We also tested for potential interactions by running fuller models including interactions of Trial Type, Age, and Condition (when applicable). Unless otherwise noted, the interactions were not significant (given the number of predictors in these fuller models, alpha was set at .01) and so parsimonious models are discussed in the main manuscript (with alpha set at .05).

Secondary analyses focused on analyzing whether children displayed direct reciprocity or group-based reciprocity at above chance levels. For direct reciprocity analyses, we compared children’s choices of benefactor/malefactor to chance levels (i.e., did children select the benefactor at above the chance level of 25%?). For group-based reciprocity, we compared
children’s choices of the benefactor’s/malefactor’s in-group to the chance level of 50%. For the latter set of analyses, we focused only on trials in which group-based (or perceptual-based) reciprocity was possible (in-group trial, out-group trial, and perceptual similarity control). All data, stimuli, and analysis code are available at: https://osf.io/4r9zs/.

Results

We first analyzed what impacts the likelihood of engaging in direct reciprocity (coded as 1 if children engaged in direct reciprocity, and 0 otherwise) across all trials (see Figure 2A). There was a significant effect of Condition, with children being more likely to engage in direct reciprocity in the Stealing Condition than the Giving Condition, $\chi^2 (1) = 8.206, p = 0.004$, no effect of Age ($p = 0.630$), and an effect of Trial Type, $\chi^2 (3) = 8.565, p = 0.036$, which was driven by children generally being less likely to reciprocate, positively or negatively, towards out-group members. This particular effect of trial type did not replicate in subsequent studies, so we do not offer a rich interpretation here (See Supplementary Materials, Figure S2 for means

Figure 2. Rates of direct reciprocity in (A) Study 1, (B) Study 2, (C) Study 3, and (D) Study 5. Dashed lines indicate chance comparisons. For Studies 2 and 3, proportions are out of the subset of children who opted into giving or stealing.
Follow-up binomial tests comparing rates of direct reciprocity to chance (chance levels were defined as 25% since there were four agents to choose from; see Figure 1) showed that children displayed negative direct reciprocity at above-chance levels ($p < .001$), but showed no evidence of positive direct reciprocity (i.e., were at chance levels; $p = 0.157$).

We next considered the possibility that children simply misremembered who the benefactor was. An analysis of the memory check questions revealed that children did in fact remember who the benefactor was: children correctly recalled the benefactor on 127 of 156 (81%) trials, which well-exceeded chance levels of 25%, binomial $p < .001$. Children were similarly aware of their own actions: they correctly recalled the beneficiary on 151 of 157 total trials 96%, binomial $p < .001$. Recall of the malefactor was similarly high: children correctly recalled who stole from them on 124 of 152 (82%) trials and also correctly recalled whom they stole from on 143 of 152 (81.5%) trials, both binomial $p$'s < .001. There were no condition differences on rates of recall (see Supplementary Materials Figure S3), and the effects of condition on direct reciprocity held even when considering only the subset of children that correctly answered all memory check questions. Thus, the lack of positive direct reciprocity could not be attributed to young children’s memory errors.

We also tested for positive group-based reciprocity during the trials that group information was available. For this analysis, we focused on whether children repaid the group that had given to them (in-group; out-group; and perceptual similarity control). Binomial tests (chance levels defined as 50%) showed no evidence of positive group-based reciprocity: children gave to the same group that had benefited them on 63 of 120 (53%) trials, binomial $p = 0.648$. 

across trials).
Thus, children showed no evidence of positive reciprocity even at the group level and even when benefactors were in-group members. These results were striking considering the particularly low bar we had set for children to display reciprocity or group-based reciprocity at all: children were instructed to give but in a non-costly manner. That is, children did not have to give up any of their own resources to reciprocate. Moreover, the reciprocating phase took place \textit{immediately} following the receiving phase (though it entailed a new object), limiting the possibility that children experienced a decay in feelings of obligation or anger.

\textbf{Study 2: Do children engage in voluntary positive and negative direct reciprocity?}

In Study 2, we extended the age range (ages 4-8) to investigate potential developmental changes and made giving and stealing voluntary.

\textbf{Methods}

In Study 2, children in the Giving condition had the option of keeping a sticker or giving it to one of the actors. In the Stealing condition, children had the option to take a sticker away from one of the actors or do nothing. We note that while giving was costly in this design, stealing involved a gain. However, our primary question was whether children willingly would engage in direct reciprocity if they now had a choice to opt in.

We reasoned that though making the action voluntary would likely lower overall rates of engaging in the action, children who opted into giving might do so because they were particularly motivated by direct reciprocity (giving back to benefactors). Moreover, in Study 1, children were told that benefactors had given up their own stickers voluntarily. However, because children’s own actions were non-voluntary, it was possible that children may have inferred that the benefactors’ actions were as well (given that everyone was described as playing the same game).
Participants

A total of 85 4-8-year-olds (Mean age: 5.93 years; Range = 4.03 - 8.92; 42 female; 43 male) from a local children’s museum. Fifteen additional children were tested but excluded due to an equipment/software malfunction (n = 8), being outside the targeted age range (n = 2), repeated failure to respond to instructions or manipulation check questions (n = 1), failure to accept the group assignment (n = 1), and failure to complete the task (n = 3).

Procedure

The procedure mimicked Study 1, except that children were randomly assigned to condition and could opt out of giving or stealing. Thus, as in Study 1, they were faced with four characters on each trial, but told they could press on their own avatar to keep the sticker (giving condition) or “steal from no one” (stealing condition).

Results

We first analyzed factors that impact children’s likelihood of engaging in reciprocity at all (i.e., likelihood of giving or stealing). There was a significant effect of Condition, with children being more likely to steal in the Stealing Condition than give in the Giving Condition, $\chi^2(1) = 29.104, p < .001$, and a significant effect of Age, $\chi^2(1) = 5.545, p = .019$, with older children being more likely to engage in the target behavior, and no effect of Trial Type ($p = 0.689$). See Figure 3.

The more critical question concerned the rates of direct reciprocity. For this analysis, we restricted ourselves to trials during which children engaged in any giving or stealing (53% in the Giving Condition and 98% in the Stealing condition). Once again, there was a significant effect of Condition, with children being more likely to engage in direct reciprocity in the Stealing Condition than the Giving Condition, $\chi^2(1) = 7.988, p = .005$ and no effect of Trial Type or Age.
Follow-up binomial tests showed that as in Study 1, children showed no evidence of positive direct reciprocity (see Figure 2B): of the 99 trials on which children gave, they selected the benefactor on 25 of them (25%; binomial $p = 1.0$). In contrast, children stole stickers from the actor who stole from them on 71 of the 149 trials (48%), which exceeded chance levels (binomial $p < .001$).

As in Study 1, rates of recalling benefactors and malefactors were very high: children correctly identified the benefactor on 153 of 174 trials (88%) and the beneficiary on 156 of 175 trials (89%) in the giving condition; rates of correctly identifying the malefactor (117 of 137 trials; 85%) and the maleficiary (129 of 138; 93%) were also high (all binomial $p$’s < .001), and there were no condition differences. Once again, there were no condition effects on rates of recall and the condition effects held even when considering only children who answered both memory check questions correctly.

As in Study 1, there was also no evidence of positive group-based reciprocity: children did not give to the benefactor’s in group at above chance levels (binomial $p = 0.207$). See also Figure S4.

The fact that children showed no evidence of positive direct reciprocity is even more striking when considering that these analyses were conducted only on trials during which children opted to engage in voluntary giving behaviors. With age, children were increasingly willing to sacrifice a sticker to reciprocate the positive action but did so in a generalized manner. Positive direct reciprocity did not appear even though children recalled who had given to them. By contrast, children readily stole when told that they could and engaged in negative direct reciprocity at above chance levels even at the youngest ages tested.
Study 3: Do children engage in positive direct reciprocity towards non-minimal groups?

Thus far, we have found that children tracked which actors engaged in positive and negative actions but only engaged in negative direct reciprocity. When children did engage in positive reciprocity, it was in generalized form – there was no evidence of either direct or group-based positive reciprocity. However, it is possible that children would be more likely to engage in either positive direct or group-based reciprocity when group markers are more salient and mimic the groups that children encounter in their daily lives.

Figure 3. Rates of children choosing to engage in target behaviors (opting to give or steal) in Study 2. Red dots identify children who opted to give or steal but did not target the benefactor/malefactor; blue dots identify children who opted to give or steal and did target the benefactor/malefactor; gray dots identify children who opted to keep the sticker or not steal. For ease of presentation, data points are jittered along the Y-axis.
Methods

Participants

A total of 40 4-8-year-olds (Mean age: 5.95 years; Range = 4.37 - 8.06; 16 female; 24 male) were recruited from a local children’s museum or park. One additional child was tested but excluded due to being from a country other than the US (and thus we could not be sure that the US-based social group marking would apply).

Procedure

In Study 3, we focused only on positive reciprocity and added trials with two “real” (non-minimal) groups. We chose gender, because children are known to essentialize on the basis of gender (Taylor, Rhodes, & Gelman, 2009) and nationality, because it implies repeated interaction and shared characteristics such as common language and experiences, preferences for which emerge early in development (Kinzler & Dautel, 2012). We included one minimal group, food preference, as a control (see Supplementary Materials Figure S1 for example stimuli). Aside from now having six trials (Gender In-Group and Out-Group, Nationality In-Group and Out-Group, and Food In-Group and Out-Group), all other aspects of the design were the same as the giving Condition of Study 2 – children could opt out of giving and giving was costly.

The sequence of events mimicked Studies 1 and 2 with the following modifications: all computerized confederates within each trial were perceptually identical (e.g., 4 pandas on 1 trial, 4 sheep on the next trial). Although confederates in prior studies were randomly chosen and assigned, we nonetheless wanted to ensure that rates of positive reciprocity were not artificially deflated because children had preferences for specific animals across trials. Given this modification, we did not include a perceptually similar control (since all characters were
perceptually similar) or an initial trial in order to keep the protocol a reasonable length for the youngest children. Trials are described in detail in the Supplementary Materials, Full Methods.

**Results**

We first looked at the rate of opting into giving, regardless of the recipient, as opposed to keeping the sticker for themselves. An initial analysis showed that children chose to give on 133 of the 240 total trials (55%), similar to rates found in Study 2. There were no effects of Age, Block Type (Gender, Nationality, Food Preference), or In Group Trial Type (In Group vs. Out Group) on likelihood of opting into giving behavior (all $p$’s $>$ .25).

We next analyzed whether children engaged in direct reciprocity on trials in which they gave the sticker away. Overall, children failed to engage in positive direct reciprocity at above-chance levels, binomial $p = 0.422$ (Figure 2C). A model predicting direct reciprocity using Age, Block Type, and In Group Trial as fixed effects and Subject ID as a random effect showed that children were more likely to engage in direct reciprocity when the benefactor was an in-group member (comparing In Group Trial to Out Group Trial, $\chi^2(1) = 5.447, p = .020$), and no other significant effects (all $p$’s $>$ 0.25); however, direct reciprocity did not exceed chance levels in any of the individual trials (all $p$’s $>$ .05). More generally, across all trials children showed an in-group bias: children gave to in-group members on 88 of 133 trials in which children opted to give, binomial $p < .001$. Thus, children were more likely to give to in-group members in both In Group and Out Group trials, but once again failed to engage in positive direct reciprocity at above-chance levels even on those trials.

The lack of positive direct reciprocity again could not be attributed to memory errors: children correctly recalled the benefactor on 195 of 240 trials (81%) and the beneficiary on 210 of 239 trials (88%), binomial $p$’s $<$ .001.
Group-based reciprocity also did not exceed chance levels: children gave back to the same group that had benefited them on 71 of 133 trials on which they had opted to give at all (53%), which did not exceed chance levels of 50%, binomial \( p = 0.488 \). Thus, even when groups were of the form that children encounter in their day-to-day lives, children failed to engage in positive direct or group-based reciprocity.

**Study 4: Do children understand the norm of positive direct reciprocity?**

Thus far, in three studies, children rarely displayed positive direct reciprocity. This is a strange result given that, by about 3.5 years of age, children understand the norm of positive direct reciprocity when applied to third parties (Olson & Spelke, 2008). However, recent work has also shown that children in the age range tested can recognize resource distribution norms for others but not use those norms to guide their own behavior (a phenomenon referred to as the knowledge-behavior gap; see Smith et al., 2013). More specifically, children may recognize the norm but not know that it applies to themselves. To assess this possibility, we investigated whether children recognized that the norms of positive direct reciprocity applied to their own interactions with others.

**Methods**

Using the same basic task, we asked children which person they should give to after receiving a sticker from one of the four avatars (Self Norm) and also which person should give to them after the child had given first (Other Norm). Children did not incur a cost in either trial. This design allowed us to assess whether children would identify the appropriate benefactor and the beneficiary for direct reciprocity to occur. In order to avoid confounding norms about direct reciprocity with norms about in-group giving or about giving in general, we focused on a context without social groups.
Participants

A total of 51 4-8-year-olds (Mean age: 6.67 years; Range = 4.19 – 9.33; 31 female; 20 male) from a local park. No children were excluded.

Procedure

Pre-Test Questions. All introductory procedures were identical to all previous studies, with the following modification: Prior to being introduced to the computer game and making an avatar selection, children were tested to ensure that they understood the word “should”. Since our main analyses hinged on children understanding this word, we included two explanatory pre-test questions. Children were told that this would be a game in which they would talk about what they should do. They were then told that should can be defined as things that “you have to do”. They were then asked two pre-test questions in which children’s desires were pitted against a social norm: “If your mom says to go to bed, should you stay up or should you go to bed?” and “If your teacher says you should do homework, should you play outside or should you do your homework?”. All except 2 children answered both questions correctly.

Trials. Children were then shown the Initial Trial stimuli of Studies 1 and 2, and completed two randomly ordered trials in which they answered questions about the norms of giving: a Self-Norm Trial, and an Other-Norm Trial. In the Self-Norm Trial, children were randomly allocated a sticker by one of the four benefactors. During the reciprocating phase, however, rather than being tasked with donating a new sticker, children were simply shown the new sticker and asked which of the 4 confederates they should give it to. In the Other-Norm Trial, the order of sequence was reversed. Children were first given a sticker that they could not keep and directed to give it to one of the four computerized confederates. Stickers were then cleared off the screen, and children were told that the game had allocated a new sticker to each of
the 4 confederates, but not to the child. Children were then asked which of the confederates *should* give to the child. The ordering of the trials was randomly chosen by the experimental software. After answering each norm question in each trial, children were asked to recall the benefactor/beneficiary in the same manner as in Studies 1-3.

**Results**

A binomial regression model including Age and Trial Type (Self-, Other-Norm) as fixed effects, Subject ID as a random effect, and likelihood of identifying the prior benefactor/beneficiary (coded as 1 = yes and 0 = no) as the dependent variable revealed a significant effect of Age, $\chi^2(1) = 6.015, p = .014$, with older children being more likely to say that both they and other children should engage in positive direct reciprocity, and no effect of Trial Type ($p = 0.270$).

As can be seen in Figure 4, children did not choose the benefactor or the beneficiary at above chance levels until around age 7. These results once again could not be explained by errors in memory encoding: children recalled both the benefactor and beneficiary (answered both questions correctly) on 98 of 102 (96%) total trials, binomial $p < 0.001$.

*Figure 4.* Rates of children endorsing direct reciprocity in Study 4. Dashed lines indicate chance responding. For ease of presentation, data points are jittered along the Y-axis.
These results show that younger children may fail to show positive direct reciprocity because they do not apply the norm to themselves. By about 7 years of age, children recognize that the positive norm applies to their own interactions with others, but children at this age still do not enact the norm behaviorally. Moreover, the two sides of the norm – “I should reciprocate with others” and “they should reciprocate with me” – emerge simultaneously in development. Given that children in this age range already known the norm and apply it to others (Olson & Spelke, 2008), it is possible that reminding them of the norm may induce them to apply it in their own interactions, closing the gap between knowledge and behavior.

**Study 5: Can positive direct reciprocity be shaped by normative input?**

In Study 5, we employed a teaching intervention to facilitate the mapping between the norm and children’s behavior in the reciprocity task.

**Methods**

To test whether a story intervention would enhance positive direct reciprocity, we created stories that mirrored the behavioral task and prompted children to think about what the characters did and why. Recent work suggests that children of this age readily learn descriptive norms (Roberts, Gelman, & Ho, 2016; Tworek & Cimpian, 2016) and adjust their own behavior accordingly (Blake, Corbit, Callaghan, & Warneken, 2016; McAuliffe, Raihani, & Dunham, 2017). Storybooks and testimony have been shown to be a relatively quick and powerful way to shape children’s behavior (Larsen, Lee, & Ganea, 2017; Steinbeis & Over, 2017).

Children heard either a Reciprocity Story or a Control Story. In the Reciprocity Story, one character (the benefactor) helped another character (the beneficiary) one day and this
kindness was reciprocated by the beneficiary the next day. In the Control Story, children heard about the same two characters engaging in the same surface prosocial behaviors outside of the context of reciprocation (i.e., one benefactor donates twice to the beneficiary). Children were then asked what happened, how the characters felt and why they acted the way they did. Following the experimental manipulation, children were given a non-voluntary version of the reciprocity task.

**Participants**

A total of 74 4-8-year-olds (*Mean age*: 6.42 years; *Range* = 4.03 - 8.89; 45 female; 29 male) were recruited from a local children’s museum or park. An additional 6 children were tested but excluded due to being from a country other than the US (so we could not be sure that the Country group-based markings would apply; *n* = 3), equipment/software malfunction (*n* = 2), or failure to complete the task (*n* = 1).

**Procedure**

Children were randomly assigned to either a *Reciprocity Condition*, in which they were told a short story about two children engaging in reciprocal altruism, or a *Control Condition*, in which they heard a language-matched story but one in which people did not engage in any reciprocity. The script of both stories began as follows:

“Today, before we start the game, let me tell you a story about my friend, Bob. One day Bob felt left out because he didn’t have any stickers to play with at school. One of the other kids, Tom, saw that Bob didn’t have any stickers and gave Bob one of his stickers so Bob would have one.” At this point, children were asked two manipulation check questions to ensure they understood the sequence of events (“So how did Bob feel? and What did Tom do?”) and given corrective feedback if necessary.
In the *Reciprocity Group*, children were then told that Bob engaged in reciprocal altruism the next day: “But the next day, Bob saw that Tom felt left out and didn’t have a sticker. Bob thought, ‘I remember Tom gave me a sticker yesterday so I should do the same for him today.’ So, Bob gave Tom one of his stickers because he remembered that Tom did the same for him yesterday.”

In the *Control Group*, children heard that Tom engaged in the same exact behavior towards Bob and simply repeated his own behavior: “But the next day, Tom saw Bob felt left out again and didn’t have a sticker. Tom thought, ‘I remember when I gave Bob a sticker yesterday, so I should do the same for him today.’ So, Tom gave one of his stickers to Bob because he remembered that he did the same for him yesterday.”

Children were asked three questions to prompt story processing (“So how did Bob/Tom feel the next day?” “And what did Bob/Tom do?”) including one explanation question (“And why do you think Tom did that?”) in order to facilitate processing of the story events. No feedback was provided. Children were then given the reciprocity task from Study 3 with groups for Gender, Nationality and Food Preference. Giving in this task was non-voluntary to determine whether children had adopted the norm of positive direct reciprocity.

**Results**

The reciprocity story effectively raised rates of positive direct reciprocity relative to the control (Figure 2D). An analysis of positive direct reciprocity showed that there were main effects of Age, $\chi^2(1) = 5.646$, $p = .017$, with older children being more likely to engage in direct reciprocity than younger children, and Condition, $\chi^2(1) = 4.012$, $p = 0.045$, with children being more likely to engage in direct reciprocity in the Reciprocity than the Control condition, but no
interaction of Age and Condition. There was no effect of Block Type (Gender, Nationality, Food Preference) or In Group Trial Type (In-Group or Out-Group Trials; both $p$’s > 0.25).

Follow-up binomial tests (comparing each condition to 25%) showed that children elected to give to the benefactor in both conditions: Control Condition, 78 of 234 trials (33%), $p = .004$; Reciprocity Condition, 92 of 210 trials (44%), $p < .001$. These results suggest that both conditions enhanced rates of positive direct reciprocity but that the effect was stronger in the Reciprocity Condition.

As in all studies, rates of recalling the benefactor and beneficiary were high in both conditions: children correctly recalled the benefactor on 352 of 444 total trials (79%; binomial $p < .001$) and correctly recalled the beneficiary on 413 of 444 total trials (93%; binomial $p < .001$), and results continued to hold even when considering only the subsample of children who answered both memory check questions correctly (Supplementary Materials, Figure S3).

**Overall Analyses**

Combining the data from the Giving Conditions of Studies 1-3, and the Control Condition of Study 5, we found no evidence of positive direct reciprocity: of the 625 trials on which children had a chance to display positive reciprocity, children gave back to the benefactor on 175 of them (28%), which was no different than chance levels of 25%, binomial $p = 0.087$. Given the marginal effect, and to make sure that we were in fact, uncovering a null effect rather than failing to find a small effect, we ran a Bayesian analysis which allowed us to estimate the degree to which these data supported direct reciprocity versus the null of no direct reciprocity. The result revealed a Bayes factor of 5.151 in favor of the null (95% Credible Interval = 25% - 32%), suggesting that these data were over 5 times more likely to be observed under the null hypothesis of no positive direct reciprocity than the alternative.
In contrast, of the 309 trials on which children could display negative direct reciprocity (Studies 1 and 2), children did so on 135 of them (40%), reflecting a significant deviation from chance, binomial $p < .001$. A Bayesian analysis yielded a Bayes factor of $8.307 \times 10^9$ of the data being observed under the alternative over the null (95% Credible Interval = 38% - 49%), providing strong evidence in favor of negative direct reciprocity.

Further, across all studies that used groups, we found no evidence of positive group-based reciprocity: children displayed group-based reciprocity on 301 of 563 total giving trials (54%), binomial $p = 0.109$, Bayes factor = 4.917 favoring the null over the alternative hypothesis (95% Credible Interval = 49% - 58%). Children did show evidence of negative group-based reciprocity and stole from the same groups that had stolen from them a total of 179 of 271 (66%) trials, binomial $p < 0.001$, Bayes factor = 107,261.162 favoring the alternative over the null (95% Credible Interval = 60% - 71%). However, we note that this result is largely driven by greater negative direct reciprocity overall: a follow-up analysis looking at only the subset of trials on which participants did not steal from the malefactor ($n = 174$) showed that children were no more likely to steal from the group member that had stolen from them ($n = 62$) than would be expected by chance (i.e., they did so 35.6% of the time; binomial $p > 0.25$; Bayes factor = 8.430 favoring the null (95% Credible Interval = 29% - 43%).

Finally, mimicking a wealth of prior work on early-developing in-group bias (Dunham et al., 2011), we found evidence of a general positive in-group bias. Across the 864 in-group and out-group trials (Studies 1, 2, 3 and 5), children gave to the in-group on 487 of these (56%) binomial $p < .001$, Bayes factor = 47.393, favoring the alternative over the null (95% Credible Interval = 53% - 60%). Thus, children were more likely to give to the in-group overall but this
tendency was not moderated by whether a prior benefit came from an in-group versus out-group member.

**Discussion**

The idea of “paying back” is deeply embedded in adult social cognition, proposed to be critical to group functioning, and is generally assumed to be evolutionarily ancient and early-developing. We began this line of work in an honest attempt to investigate the developmental progression of early-developing positive direct reciprocity. However, across multiple studies we found no evidence of its early emergence. Children rarely engaged in positive direct reciprocity across a wide range of contexts: voluntary and forced reciprocity, within and outside of social groups, and within real and minimal groups. Children also did not use group membership as a simplified form of targeting reciprocal exchanges. In contrast to these repeated failures to exhibit targeted forms of positive reciprocity, children at the youngest age sampled in our task readily engaged in both voluntary and compulsory negative direct reciprocity suggesting that preschoolers can and do engage in some targeted reciprocal actions towards others.

The distinct developmental trajectories for negative and positive direct reciprocity align with evidence of a dissociation between positive and negative reciprocity in adults (Egloff, Richter, & Schmukle, 2013; Gray et al., 2014; Keysar et al., 2008). The current studies extend this research and offer further insight into how separate psychological mechanisms for positive and negative direct reciprocity emerge during childhood.

The fact that young children show only generalized positive reciprocity aligns with theoretical proposals for indiscriminate altruism which becomes more targeted and selective with age (Warneken, 2018). By contrast, negative reciprocity was direct for even the youngest ages tested. Across the age range tested (4 to 8 years of age) children were highly motivated to correct
an injustice against themselves and targeted the malefactor at well above chance levels. This relatively early emergence is in line with proposals for an adaptive cognitive system for revenge that protects oneself from exploitation (McCullough, Kurzban, & Tabak, 2013).

Although it is possible that negative direct reciprocity is learned from others to some extent, the protracted development of positive direct reciprocity suggests a longer learning process for positive direct reciprocity. The process of connecting knowledge about norms of reciprocity with children’s own behavior may take several years. Similar to studies on children’s giving behavior, children as young as 3.5 years recognize the norm of positive direct reciprocity and apply that norm to other people (Olson & Spelke, 2008). However, despite this early knowledge, until about 7 years of age children do not seem to know that they should return a favor to the same person who benefitted them previously, and do not even expect positive reciprocity from a person they had just rewarded. Surprisingly, even after children recognize that they should repay a benefactor, they still do not actually reciprocate when they have the chance. This link between knowledge and behavior can be induced by a brief story describing positive reciprocity between peers.

We propose that early in ontogeny humans prioritize negative reciprocity over positive reciprocity. One distinct possibility for the asymmetry concerns the expectations that young children have surrounding positive exchanges – prior work finds that even infants expect others to exhibit positive behaviors, and are relatively more surprised by antisocial actions (Hamlin, Wynn, & Bloom, 2010; Vaish, Grossmann, & Woodward, 2008). If this is the case, children may view negative actions as more diagnostic of individuals’ underlying dispositional traits. An encounter with a likely “bad” person may thus cause greater emotional arousal and drive the
higher rate of reciprocation that we see in our study. Future work should directly address these cognitive and emotional mechanisms.

To conclude, although school-aged children are capable of engaging in direct reciprocity, they are more likely to reciprocate negative actions than positive actions. This early propensity for punishing those who have harmed you suggests a developmental priority for preventing future victimization. In contrast, children remember positive actors but do not feel obligated to pay them back. On a more optimistic note, positive direct reciprocity is easily learned through a relatively brief narrative about reciprocation. The developmental asymmetry of positive and negative direct reciprocity suggests that our propensities for “an eye for an eye” and our propensities for “you scratch my back, I scratch yours” are underpinned by distinct psychological mechanisms.

References


Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I should but I won’t: Why young children endorse norms of fair sharing but do not follow them. *PloS One, 8*(3), e59510.


